Project no. CIT5-CT-2005-028647

Project acronym: UPP

Project title: "Understanding Privatisation Policy: Political Economy and Welfare Effects"

Instrument: SPECIFIC TARGETED RESEARCH PROJECT
Thematic Priority:7- "Citizens and governance in a Knowledge-based society"

Deliverable: D.6.4

"Databases on welfare measures for consumers and employees"

Measuring Service Quality: the opinion of Europeans about Utilities

P. A. Ferrari, S. Salini [1]

Due date of deliverable: 31 July 2007
Actual submission date: 28 September 2007

Start date of project: 01/02/2006 Duration: 2 years

Organisation name of lead contractor for this deliverable:
University of Milan (UMIL)

Revision: draft 1

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

[1] Department of Economics, Business and Statistics. University of Milan. Via Conservatorio 7. 20122 Milan Italy. Corresponding Authors: pieralda.ferrari@unimi.it, silvia.salini@unimi.it

Measuring Service Quality: the opinion of Europeans about Utilities

P. A. Ferrari, S. Salini [2]

Abstract

This paper provides a comparative analysis of statistical methods to evaluate the consumer perception about the quality of Services of General Interest. The evaluation of the service quality perceived by users is usually based on Customer Satisfaction Survey data and an *ex-post* evaluation is then performed. Another approach, consisting in evaluating Consumers preferences, supplies an *ex-ante* information on Service Quality. Here, the ex-post approach is considered, two non-standard techniques - the Rasch Model and the Nonlinear Principal Component Analysis - are presented and the potential of both methods is discussed. These methods are applied on the Eurobarometer Survey data to assess the consumer satisfaction among European countries and in different years.

Keywords: Service Quality, Eurobarometer, Non Linear Principal Component Analysis, Rasch Analysis, Conjoint analysis

JEL: C33, C35, C43, L94, L95, L96

Acknowledgements

This paper has been supported under the VI Framework Programme (project title: “Understanding Privatisation Policy: Political Economy and Welfare Effects”, Project no. CIT5-CT-2005-028647) of the European Union. The authors are grateful to Giancarlo Manzi for competent research assistance.

[2] Department of Economics, Business and Statistics. University of Milan. Via Conservatorio 7. 20122 Milan Italy. Corresponding Authors: pieralda.ferrari@unimi.it, silvia.salini@unimi.it

1. Introduction

A Service of General Interest’s quality can be considered from different points of view and from various angles. In this paper the Consumer’s point of view is analyzed. This is not an easy task, especially when the context in question is complex as the European one. In fact, the size and characteristics of these services present the first problem. Services of General Interest are managed and supplied differently in each European State. Furthermore, in each State conditions of management change with time. An example of this is provided by the European privatization process, implemented in a different way in each State (Fiorio and Florio, 2007). A second important problem is how to measure the perceived quality of a Service. The classic way to do this is by using Customer Satisfaction Surveys. In particular, Eurobarometer Surveys conducted recently (Eurobarometer 2002, 2004, 2006), have included questions concerning Services of General Interest. By using these surveys and applying advanced statistical methods, it is possible to evaluate Consumer satisfaction regarding different aspects of the Service (accessibility, price, etc) as well as extract an evaluation of Service Quality. The analysis of Customer Satisfaction Survey data is always performed *ex-post* and gives useful information to both legislators and service providers who require decision support in order to improve Service Quality.

An other way to analyze Service Quality is to evaluate the preferences of the Consumer in order to understand which aspects a Consumer is ready to surrender for the benefit of some other aspects. Some ad hoc surveys in which the data structure is obviously different to that in Customer Satisfaction Surveys are required but it is possible to estimate the usefulness of the attributes of a Service by considering Consumer preferences. In this way, the legislator or supplier knows *ex-ante* upon which attributes of the Service to focus his attention in order to improve the Consumer’s opinion of the overall quality. For some applications of preferences analysis techniques, such as Conjoint Analysis, in the Service Quality context see for example Marcucci et. al. (2007), Barone et. al. (2004).

In this paper we deal with Service Quality according to the first approach, Customer Satisfaction Survey data structure and methods for its statistical analysis. In particular we shall discuss non-standard methods of Customer Satisfaction, which are useful in order to assess the customer satisfaction and to draw comparisons between different years and Countries.

The paper is curried out as follows. Section 2 is dedicated to Service Quality and Customer Satisfaction, a critical description of Customer Satisfaction analysis is briefly presented and Rasch Model (RM) and Nonlinear Principal Component Analysis (NLPCA) are introduced. In Section 3 RM and NLPCA methods are applied to Eurobarometer data for years 2000, 2002, 2004, according to (Fiorio et al. 2007), The services examined are Fixed Telephone, Electricity, Gas and Water Supply Services. Finally some conclusions are given.

2. Measuring Service Quality: Customer Satisfaction Approach

In order to measure the quality of a Service by users', two different approaches can be followed: *ex-post* based on Customer Satisfaction Survey or *ex-ante* based on Customer Preferences Survey. In this section two method for the first approach are detailed.

It is generally accepted that Customer Satisfaction, like every subjective attitude, is a complex concept that can not be directly observed but should instead be measured using other observed variables which are connected to different aspects and to the level of satisfaction itself.

In order to have a knowledge of Customer Satisfaction, survey questionnaires are used, in which respondents are asked to declare their degree of satisfaction with regard to different aspects of the service or product. Hence, statistical analysis of data from these surveys is carried out and measures of each aspect or/and of overall satisfaction are obtained. Nonetheless this data is rather troublesome to handle for many different reasons related to the specific and subjective nature of the observed variables.

First of all, the relevance and/or weighting of the manifest variables that contribute to determining the level of satisfaction are unknown. In addition, these variables often have an ordinal measurement scale which needs to be suitably dealt with. Therefore, the level of satisfaction is generally dependent on both expectation and individual features of respondents as well as on contextual variables (Fiorio et al., 2006). Furthermore, surveys contain subjective questions like "What do you think of the quality of X services that you use?" or "Would you say that price you pay for the X is fair or unfair?". This leads not to objective but subjective variables instead that express what people think or what people say. Measurement errors may then emerge from the subjective nature of the variables and a cognitive dissonance can affect data with undesired consequences on the effectiveness of the results, as has already been addressed (see for example Bertrand et al, 2001).

With the objective to solve, or at least control, some of these problems, many different methods to assess Customer Satisfaction have been proposed (for an interesting review see Zanella, 2001). Two main approaches can be identified. The first one uses statistical models to estimate the relationship between the latent variable and the manifest variable and involves, among others, the structured equation models by applying Partial Least Squares (PLS, presented for example in Tenenhaus et al., 2005) or LInear Structured RELationship (LISREL, see for example Joreskog, 1994). The second approach does not assume any model, but uses instead descriptive analysis by adopting dimensionality reduction methods, such as for example: Factor Analysis (FA) or Principal Component Analysis (PCA).

All of the above methods assume that the observed variables' categories are numerical, so that as a minimum, a Likert scale (Likert, 1932) is required. But the manifest variables are usually categorical ones even if they are measured at an ordinal level scale, so that the (ordinal) categories need to be coded in points to carry out the analysis. Adopting this method might be a good practice to make the scale *less subjective,* but it does not solve the problem of ordinal data. In fact the numerical labels indicate the rating of categories but not their values. Consequently, the resulting distance between subsequent numerical labels cannot be assumed to be actual. In addition many of these methods postulate linear relationships between the variables, this assumption might be not realistic.

In this paper two different approaches are proposed to assess Customer Satisfaction. These approaches are able to take into account the order of categories without establishing an *a priori* difference between them and can pick up on nonlinear relationships. They are: the Rasch Model (RM) and the Nonlinear Principal Components Analysis (NLPCA). The first approach assumes a model entirely known except for the values of parameters which have to be estimated. The second method is connected to an algorithmic procedure instead, no data generating process is assumed but the best representation of data is searched. In addition these two methods can be used in a complementary way (see Ferrari et al., 2005). To be precise RM is a good tool for calibrating the questionnaire properly whilst NLPCA can be used subsequently to quantify the variables' categories and weights of manifest variables in order to set up a synthetic indicator of the level of satisfaction available for further analysis.

Both these methods allow an overall analysis of the problem and constitute a preliminary study with the objective of emphasising the main features of satisfaction and detecting hotspots for the satisfaction itself. For a more profound analysis we need to exam in detail the single situations. Some suggestions for this will be given in the last paragraph.

2.1. Rasch Model

Rasch models (RM) are used for analysing data returned from assessments performed to measure things such as abilities, attitudes, and personality traits. RM are particularly used in psychometrics but due to their general applicability, they are being used increasingly in other areas, including the health profession and market research. In a recent paper the use of RM was extended to quality and satisfaction measurement (De Battisti et al. 2005). The mathematical theory underlying Rasch models is in some respects the same as item response theory (IRT) (Hambleton, 1991). However, Rasch models have a specific measurement property that provides a criterion for successful measurement. This formal property distinguishes Rasch models from other models used to model people responses to items or questions. Application of the models provides diagnostic information regarding how well the criterion is met. Application of the models also provides information about how well items or questions in assessments work in measuring the ability or a latent trait.

The Rasch Model was introduced in the 60s in order to evaluate ability tests (Rasch 1960/1980). These tests are based on a set of items and the assessment of a tested subject's ability depends on two factors: his relative *ability* and the item's intrinsic *difficulty*. Through the RM the two factors are measured by the parameters θ_i referred to the subject i and β_j referred to the item j. Their relationship is expressed by the difference ($\theta_i - \beta_j$). In a deterministic sense a positive difference means that the subjects' abilities are superior to the item's difficulty and therefore an exact response will always been given. In a probabilistic perspective, as in the RM, this is not true since a subject that can intrinsically give a right answer $(\theta_i > \beta_j)$, can instead, for negative circumstances, give a wrong response as, on the contrary, a subject with lacking abilities can accidentally give a right answer[3].

---

[3] In applying the Rasch model, item parameters are often scaled first. This part of the process of scaling is often referred to as item *calibration*. In educational tests, the smaller the proportion of correct responses, the higher the difficulty of an item and hence the higher the item's scale location. Once item locations are scaled, the person locations are measured on the scale.

In Rasch dichotomous model, the probability of a correct answer $x_{ij}=1$ by the subject i of ability θ_i when he meets the item j of difficulty β_j is:

$$P\left\{x_{ij}=1\middle|\theta_i,\beta_j\right\}=\exp\left(\theta_i-\beta_j\right)\Big/\left(1+\exp\left(\theta_i-\beta_j\right)\right) \qquad (1)$$

while the probability of a wrong answer $x_{ij}=0$ is: $P\left\{x_{ij}=0\middle|\theta_i,\beta_j\right\}=1\Big/\left(1+\exp\left(\theta_i-\beta_j\right)\right)$.

In the dichotomous model data are collected in the *raw score matrix*, with n rows and J columns, whose values are equal to 0 or 1. The sum of each row $r_i=\sum_{j=1}^{J}x_{ij}$ represents the total score of the subject i for all the items, while the sum of each column $s_j=\sum_{i=1}^{n}x_{ij}$ represents the score given by all the subjects to the item j. These scores are given according to a metric that, being nonlinear, produces some conceptual distortion when you wish to compare the row's and column's totals. Then, it is necessary to change these scores according to a metric that is founded on the conceptual distances between subjects and items. The transformation takes place through the logit:

$$\log\frac{p_{ij}}{1-p_{ij}} \qquad (2)$$

where p_{ij} is the probability associated to $x_{ij}=1$ and (1- p_{ij}) is the probability associated to $x_{ij}=0$. It is possible to define the parameters θ_i e β_j in the same measurement unit of an interval scale; consequently even the difference $\theta_i-\beta_j$ is gauged according to the same measurement unit.

The Rasch dichotomous model has been extended to the case of more than two ordered categories. The innovation of this approach is to assume that between every category and the next one there is a parameter threshold that "qualifies" the item position and specializes the β_j as a function of the difficulty presented by every answer category. So, the answer to every threshold k of the item j depends on the value $\beta_j+\tau_k$, where the second term represents the k-th threshold of the item j. The thresholds are ordered ($\tau_{k-1}<\tau_k$), because they reflect the category order. Different politomous models have been proposed, thus briefly described.

i) The *Rating Scale Model* (RSM), presented by David Andrich (Andrich, 1978a). A fundamental condition of the RSM, and also its limitation, is the equality of the threshold values for all the items; that is, even if the distance between a threshold and another one can differ, the pattern of these distances is constant for all the items.

ii) The *Partial Credit Model* (PCM), proposed by Masters (Masters, 1982). In this model the "difficulty" levels differ item by item and the subject receives a partial credit (score for each item) equivalent to the relative level of difficulty of the completed performance. The thresholds can differ freely in the same item or from an item to another one.

We will consider model *ii*) in the version known as *Extended Logistic Model* (ELM), proposed by Andrich (1988b). The ELM gives the probability that the subject i responds to item j through the answer x_{ij} by the following equation:

$$P(X=x_{ij})=\exp\left[\kappa_{jx}+x_{ij}(\theta_i-\beta_j)\right]\Big/\sum_{k=0}^{m}\exp\left[\kappa_{jk}+k(\theta_i-\beta_j)\right]$$

where X is the random variable which describes the answer of the subject i to item j; x_{ij} = 0,1,…,m is the number of ordered overtaken thresholds; κ_{jx} are the coefficients of each category x for each item j and they can be estimated considering that: $\kappa_{j0} = \kappa_{jm} = 0$ (the first and the last parameters are equal to zero) and that: $\kappa_{jx} = -\sum_{k=1}^{x} \tau_{jk}$ (the category coefficients are defined in terms of thresholds); τ_{jk} is the k-th ordered threshold of item j.

The defining property of Rasch models is their formal or mathematical embodiment of the principle of invariant comparison. RM embodies this principle due to the fact that their formal structure permits algebraic separation of the person and item parameters, in the sense that the person parameter can be eliminated during the process of statistical estimation of item parameters. This result is achieved through the use of *conditional maximum likelihood* estimation, in which the response space is partitioned according to person total scores. The consequence is that the raw score for an item or person is the sufficient statistic for the item or person parameter. That is to say, the person total score contains all information available within the specified context about the individual, and the item total score contains all information with respect to item, with regard to the relevant latent trait.

The RM requires a specific structure in the response data, namely a probabilistic Guttman[4] structure. In the RM, the Guttman response pattern is the most probable response pattern for a person when items are ordered from least difficult to most difficult.

As mentioned the Rasch model is a *model* in the sense that it represents the structure which data should exhibit in order to obtain measurements from the data; i.e. it provides a criterion for successful measurement. It is therefore a model in the sense of an ideal or standard. The perspective or paradigm underpinning the Rasch model is distinctly different from the perspective underpinning statistical modelling. Models are most often used with the intention of describing a set of data. Parameters are modified and accepted or rejected based on how well they fit the data. In contrast, when the Rasch model is employed, the objective is to obtain data which fits the model (Andrich, 2004). The rationale for this perspective is that the Rasch model embodies requirements which must be met in order to obtain measurement, in the sense that measurement is generally understood in the physical sciences.

Nevertheless one has to expect a data divergence from the model expectations. Various techniques have been developed in order to control the congruency between data and model.

As previously mentioned, the model has been employed in the context of Customer Satisfaction in recent years. The two factors become: the subject's (Customer's) satisfaction and the item's quality. The interpretation of satisfaction and quality parameters changes when compared to the interpretation of ability and difficulty parameters. In particular, high values for item parameters, which originally indicated high difficulties, now indicate low quality. On the other hand, the reading of subject parameters remains direct: originally high values

[4] A Guttman scale is a psychological instrument developed using the scaling technique developed by Louis Guttman in 1944 called Guttman scaling or scalogram analysis. A primary purpose of the Guttman scaling is to ensure that the instrument measures only a single trait (a property called unidimensionality, a single dimension underlies responses to the scale). Guttman's insight was that for unidimensional scales, those who agree with a more extreme test item will also agree with all less extreme items that preceded it.

indicating very skilful persons, now indicate very satisfied persons. Through calculation, a ranking of the items with regard to their quality can be obtained from the coefficients β_j. The model presents a second series of parameters, θ_i, that express each person's satisfaction; this parameter can be considered as a Customer Satisfaction Index.

2.2. Non Linear Principal Components Analysis

In order to achieve a suitable measurement of satisfaction now, we assume the hypothesis that the measure of Customer satisfaction can be given by a synthetic indicator that reduces the multiple items indicating the different aspects into a univariate variable. In other words, we assume that observations on the set of categorical variables for each respondent can be mapped onto a single real number, that expresses her/his level of satisfaction.

To reach this goal an appropriate tool presents itself in the form of NLPCA. This analysis belongs to the so-called "Gifi system" and was proposed by the Data Theory Group of the University of Leiden in 1990 and developed in the years following (see Gifi, 1990 and Michailidis and De Leeuw,1998). It is a dimensionality reduction method which is capable of handling nominal, ordinal and numerical variables, all at the same time, according to their measurement level.

The adoption of NLPCA for data analysis of satisfaction seems particularly suitable because it allows for the synthesis of observed variables in a reduced space, preserving measurement levels of qualitative ordinal data without assuming an *a priori* difference between subsequent categories. The latent dimension is derived as a linear combination of the observed variables characterized by an optimal quantification of their ordinal categories and of their weights in the construction of the linear combination.

In NLPCA each of the m columns of the $n \times m$ data matrix (each column is an variable and each row is an object) is monotonically transformed in such a way that a reduced number p of new continuous variables (components) optimally fits the transformed data. Here variables are the items of the questionnaire, objects are respondents and only one continuous component, the level of satisfaction, is needed. Then p=1 and NLPCA can be formalized as follows.

Let $\mathbf{c}_j$ be the k_j-dimensional vector containing the ordinal categories of the jth variable, j=1,2,…,m, $\mathbf{H}$ the $n \times m$ matrix containing the observations of the m variables on the n objects, $\mathbf{h}_j$ the jth column of the matrix $\mathbf{H}$, $\mathbf{G}_j$ the $n \times k_j$ indicator matrix such that $\mathbf{G}_j\mathbf{c}_j = \mathbf{h}_j$. The target of NLPCA is to find the vector $\mathbf{x}_{[n\times 1]}$ of object scores (here interpreted as respondents' satisfaction measures) that minimizes the following loss function:

$$\sigma^2\left(\mathbf{x},\mathbf{q}_j,\beta_j\right)=\frac{1}{m}\sum_{j=1}^{m}\left(\mathbf{x}-\mathbf{G}_j\mathbf{q}_j\beta_j\right)^T\left(\mathbf{x}-\mathbf{G}_j\mathbf{q}_j\beta_j\right) \qquad (3)$$

where $\mathbf{q}_j$ (j=1,.... m) is the k_j vector that contains optimal category quantifications for variable j and β_j is a scalar of *component loading* for variable j.

In order to avoid trivial solutions, identification constraints are required. Usually, object scores are standardized, so the following conditions are imposed:

$$\mathbf{x}^T\mathbf{x} = n$$
$$\mathbf{u}_n^T\mathbf{x} = 0 \tag{4}$$

with $\mathbf{u}_n^T$ vector of ones of order *n*. Given the need to preserve the order of categories, the further condition $\mathbf{q}_j \in C_j$,being C_j the convex cone of vectors with non-decreasing elements, is imposed [5].

The optimal solution is derived by means of an iterative algorithm called *Alternating Least Squares* (ALS), conveniently adapted to this case to assure the above condition regarding to the order of quantifications (Michailidis, 1998). From the conditions (4) it follows that $\mathbf{t}^T\mathbf{t} = n$ and $\mathbf{u}_n^T\mathbf{t} = 0$ so that the transformed variables are also standardized.

The one dimensional solution yields the following object scores:

$$\mathbf{x} = \frac{1}{m}\sum_j \mathbf{G}_j\mathbf{q}_j\beta_j = \frac{1}{m}\sum_j \mathbf{t}_j\beta_j \tag{5}$$

where $\mathbf{t}_j$ is the $n \times 1$ vector of the transformed variable j . Therefore, given the standardisation conditions, it also follows that component loadings β_j are correlations between object scores and optimally quantified variables and represent the weights of the manifest variables on the common indicator.

By formula (5), a quantitative value, obtained as weighted mean of transformed variables t_j with loadings β_j as weights, is assigned to each respondent. The value x_i of the *ith* respondent is used as measure of her/his level of satisfaction.

Before using the one-dimensional solution here described as a feasible indicator of satisfaction, it is appropriate to evaluate its validity. For that the following conditions must be verified:

a) the first eigenvalue of the NLPCA solution is effectively much larger than the others and the solution itself fits well with the data;

b) the signs of component loadings are coherent ;

c) the solution is stable.

With regard to point a), the first eigenvalue of the NLPCA constitutes a measure of goodness of fit of the procedure. In fact, the goodness of an indicator depends on the minimization of the sum of the squared distances between the obtained scores and the data. In order to evaluate the goodness of the procedure it is thus possible to use the first eigenvalue λ_1 of the correlation matrix of transformed variables or, better, a percentage ratio between λ_1 and *m* (the number of variables in the dataset), known as the percentage of total variance accounted for the first dimension: the larger the ratio, the better the synthesis. Alternatively, Cronbach's α (Cronbach, 1951) can be determined. This coefficient, introduced as a tool for assessing the reliability of scales, is strictly connected with λ_1 (Heiser and Meulman, 1994) by the

[5] When some missing data are present a *nxn* binary diagonal matrix **M,** with entries 1 if the observation *i* is present for variable *j* and 0 otherwise, is introduced in the loss function (3) (see. Michailidis and De Leeuw, 1998).

following: $\alpha = \frac{m(\lambda_1 - 1)}{(m - 1)\lambda_1}$, the nearer that α is to 1 the better the solution.

Regarding point b), since the vector of object scores $\mathbf{x}$ is built as a simple linear combination of quantified categories, each x_i has to fulfil the mathematical conditions for the index to be valid, so that the higher the rank of observed variables, the higher the value of the satisfaction indicator, and this has to be true for each variable. This requires that the weights of combinations have the same sign, specifically the positive one.

Finally, for point c), it is also important to evaluate the stability, with regards to sample changes, of the produced outputs (i.e. eigenvalues, component loadings, category quantifications and average scores). Resampling methods are useful for evaluating the stability of a statistical output. In particular, amongst the various resampling methods, the bootstrap serves the purpose well (Efron, 1979).

In the NLPCA context, the bootstrap method can be used to check the stability of all the outputs. Here we are especially interested in verifying the stability of component loadings and quantifications, used for setting up the indicator, and of the average scores, needed for comparison. An algorithm which consists of bootstrapping samples *with* replacement from the entire data set or from some subsets, according to different objectives, will be adopted and bootstrap percentile Confidence Intervals (CI) produced (see for details Ferrari et al., 2007).

3. Application to Eurobarometer data

In this section after a brief description of the structure of Eurobarometer surveys, the RM and NLPCA methods are applied in order to analyse the level of satisfaction amongst European citizens for some Services of General Interest, their main characteristics and the advantages deriving from their application are discussed.

The results will be point out in according to the different objectives of the analysis. The first one is connected with item interpretation, the second one deals with subjects or respondents, and perceives the aim to measure the global level of satisfaction of the respondents. Finally other potential uses of the two techniques will be highlighted and in particular we will discuss the item calibration problem and further analysis of satisfaction.

3.1. Eurobarometer data

Eurobarometer public opinion surveys (henceforth, EB) have been conducted on behalf of the Directorate-General for Education and Culture of the European Commission each Spring and Autumn since the Autumn of 1973. They have included Greece since the Autumn of 1980, Portugal and Spain since the Autumn of 1985, the former German Democratic Republic since the Autumn of 1990 and Austria, Finland and Sweden from the Spring of 1995 onwards.

An identical set of questions is asked to a representative sample of the population aged fifteen years and over in each Member State. In each household, the respondent is drawn at random. All interviews are face-to-face in people's homes and in the appropriate national language. A detailed analysis of the Eurobarometer data can be found on the official Eurobarometer Web site.[6] The questions concern various aspects, including support and benefit for EU membership, support for an EU constitution, satisfaction with EU democracy and the single currency, general outlook on life and so on.

The regular sample in standard Eurobarometer surveys is 1000 people per country except Luxembourg (600) and the United Kingdom (1000 in Great Britain and 300 in Northern Ireland). In order to monitor the integration of the five new Länder into unified Germany and the European Union, 2000 persons have been sampled in Germany since the Eurobarometer 34: 1000 in East Germany and 1000 in West Germany.

In each of the 15 Member States, the survey is carried out by national institutes associated with the “INRA (Europe) European Coordination Office”. This network of institutes was selected by tender. All institutes are members of the “European Society for Opinion and Marketing Research” (ESOMAR) and comply with its standards.

Each survey comes with a set of weights obtained, using marginal and intercellular weighting, carried out on the basis of the population description provided by EUROSTAT in the Regional Statistics Yearbook (data for 1997 or 1996).

In the years 2000, 2002 and 2004 the Eurobarometer surveys included some questions relating to Services of General Interest (henceforth, SGI). The SGI considered are mobile telephone services, fixed telephone services, electricity supply services, gas supply services, water supply services, postal services, transport services within towns/cities and rail services between towns/cities. The criteria used to analyse these services are *accessibility*, the *price* of the services, the *quality*[7] of the services, the clarity of the information aimed at EU Consumers, how fair the terms and conditions of the contracts applied to the services are, Consumer complaints and how they are handled and Customer Service.

3.2. Prelimary Analysis

In this application, in accordance with Fiorio at al. (2007) we consider four Services: fixed telephone, electricity supply, gas supply, water supply and for each Service we examine three aspects: *accessibility*, *price* and *quality*.

A preliminary descriptive analysis of this data can be curried out distinguishing by year, by country, by service, by aspects of service. For example in Special Eurobarometer 219 Wave 62.1, for each aspect a distribution plot distinct by service are produced, in Figure 1 the aspect *quality* is shown: the fist two and the last two categories are collapsed.

[6] http://europa.eu.int/comm/public_opinion/

[7] With *quality* we refer to the question ‘*Overall, what do you think of XXX service that you use?*’ presents in Eurobarometer Survey.

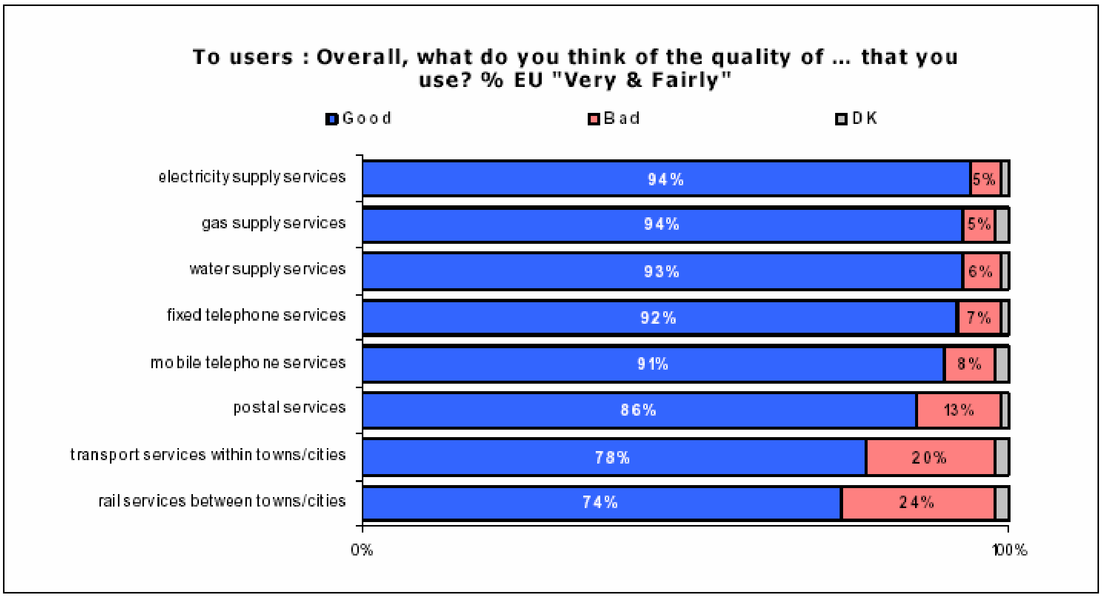


Figure 1. Special Eurobarometer 219 Wave 62.1 - TNS Opinion & Social – pag. 51

A comparision among years is done, but a single aspect is always considered and the categories are again collapsed. See Figure 2.

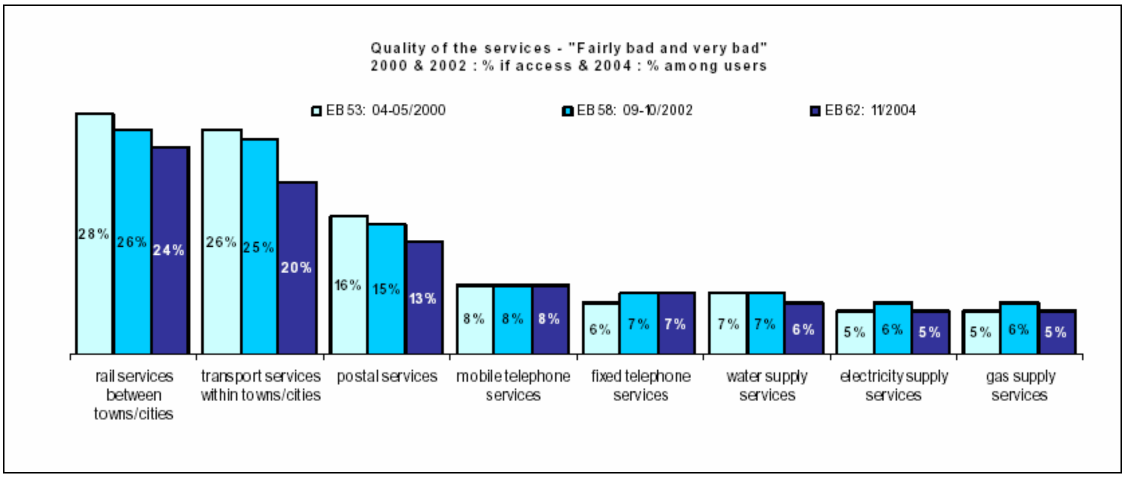


Figure 2. Special Eurobarometer 219 Wave 62.1 - TNS Opinion & Social - pag. 52

A comparison among Country is also realized, with a different graphic for each service and for each aspects of service. See for example Figure 3.

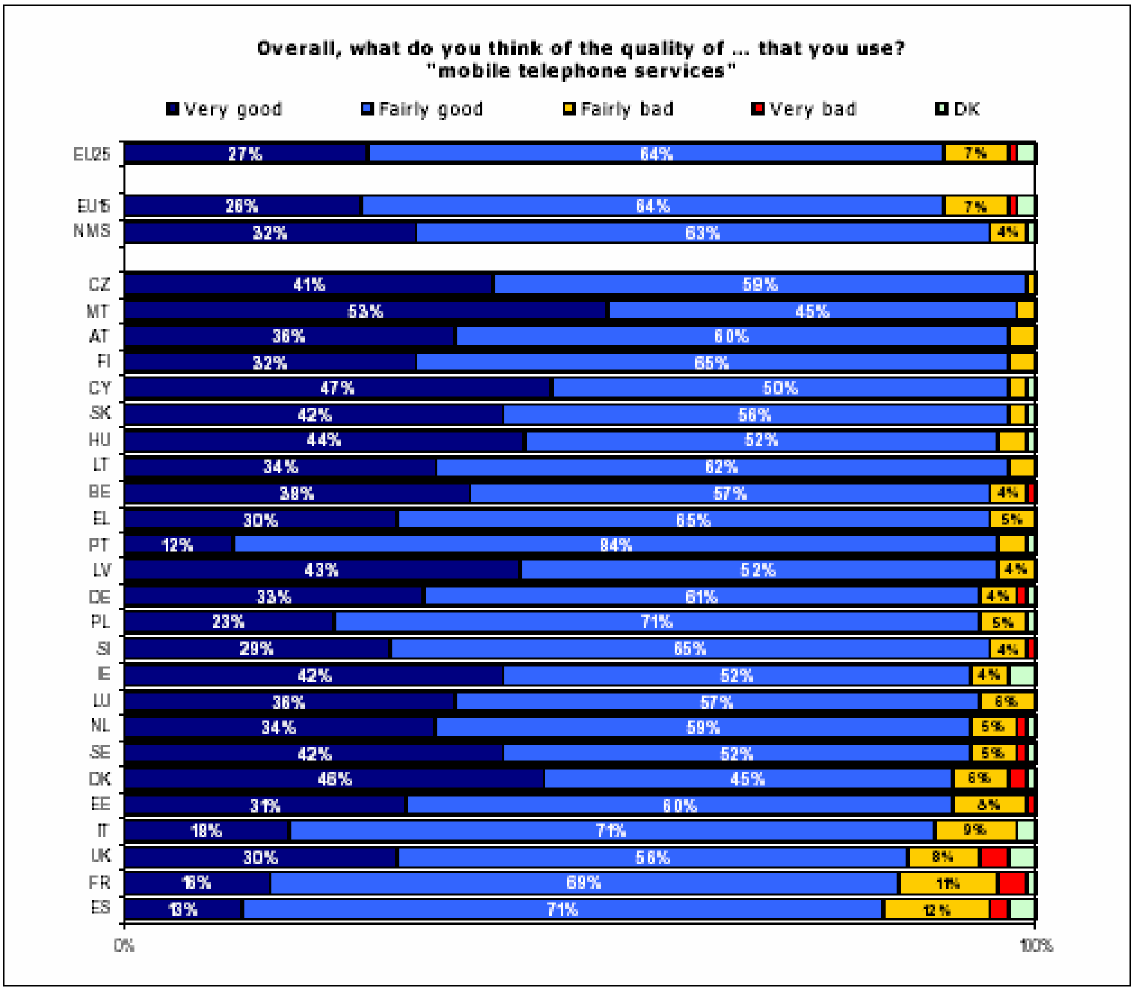


Figure 3. Special Eurobarometer 219 Wave 62.1 - TNS Opinion & Social - pag 54

This approach leads an appropriate and detailed analysis of the information contained in the data and produces a lot of useful statistical tables and graphics but misses a global and synthetic view of the situation and makes less effective a comparative analysis. Moreover it does not take into account the different role of services and aspects of service in the overall satisfaction. Finally all the categories are not always considered and that lead some information missing.

The two methods described above RM and NLPCA allow us for an overall analysis of the complex problem, constitute a preliminary study with the objective of emphasising the main features of satisfaction, detecting hotspots and permiting to establish the role played by the different services and aspects of service in consumer's satisfaction.

In particular the RM allows a ranking of items to be obtained, from one with the best quality to one with the worst, an overall consumer satisfaction measure and some indications as to how to calibrate the questionnaire. Whilst the NLPCA allows for suitable quantification of the categories and weights for observed variables to be obtained and for a satisfaction indicator to be defined. Provided that the vector of the quantifications and the vector of the weights are stable, they can be used to establish a common tool of measurement for use when comparing the level of satisfaction in subsets differentiated by factors which can influence the level of satisfaction. To show the potential of the approach here proposed the data related to three years (2000, 2002, 2004) is pooled and the analysis is carried out on the entire data set in order to obtain a comparison between countries and years. Hence the final data set is structured in the following way: the rows (near 47.000) represent the respondents belonging to different Countries (near 15.000 for each year in question), the columns refer to items (12 dimensions): the *accessibility* of the fixed telephone service (SGIaccT), the *accessibility* of the electricity supply service (SGIaccE), the *accessibility* of the gas supply service (SGIaccG), the *accessibility* of the water supply (SGIaccW), the *price* of the fixed telephone service (SGIpriT), the *price* of the electricity supply service (SGIpriE), the *price* of the gas supply service (SGIpriG), the *price* of the water supply (SGIaccW), the *quality* of the fixed telephone service (SGIquaT), the *quality* of the electricity supply service (SGIquaE), the *quality* of the gas supply service (SGIquaG), the *quality* of the water supply (SGIquaW). For the statistical analysis and to facilitate interpretation, all the item categories are ordered according to the same polarity, in particular the higher the category the higher the level of satisfaction. So that we have three levels for *accessibility* (not accessible, difficult to access, easy to access), three levels for *price* (excessive, unfair, fair) and four levels for *quality* (very bad, fairly bad, fairly good, very good).

It is important to say before the presentation of the results that Rasch diagnostic indices frankly show that the structure of the data considered does not completely satisfy the Rasch statement, in fact the overall χ^2 is 4572 with 84 degrees of freedom and the p-value is 0.000, so the null hypothesis is rejected[8]. The assumption that there is a one dimensional trait that summarizes individual Services and individual aspects of a Service might not be realistic or else some items may not be well defined, in which case the questionnaire probably needs a calibration analysis. More detail about this will be presented in the paragraph 2.4.5. Fortunately the value of Pearson Separation Index[9] is 0.876, so the further analyses that consider Rasch parameters are justified.

On the other hand the data structure makes the NLPCA particularly suitable for the analysis. In fact the first eigenvalue is by far the highest one, the Cronbach Index assume a high value (0.815), the sign of component loadings are coherent and the stability of the component loadings and quantifications are very good. So all the conditions required are fulfilled and the common indicator can be calculated. This is then used to measure the satisfaction level and allows for comparison of different situations.

3.3. Items analysis

With item analysis we mean a study that evidences which items (service or aspects of service) are more important for the Consumer (NLPCA) and which of these are perceived by the Consumer to be of a high or low quality (RM).

---

[8] The item-trait test-of-fit examines the consistency of every item parameters across the subject measures: data are combined across all items in order to give an overall test-of-fit. This shows the overall agreement for all items across different subjects. The observed answer distribution is compared to the expected answer distribution, calculated with the logistic function, by means of the χ^2 criterion. We examine the χ^2 probability (*p-value*) for the whole item set; there is not a well-defined lower limit defining a good fit (minimum acceptability level); a reference level might be 5%. The null hypothesis is that there is no interaction between responses to the items and locations of the subjects along the trait.

[9] The Separation Index is the Rasch reliability estimate, computed as the ratio (true/(true+error)) variance whose estimates come from the model. A value of 1 indicates a lack of error variance, and thus full reliability. This index is usually very close to the classic Cronbach α coefficient computed on raw scores. The power of test-of-fit, based on the Separation Reliability of 0.876, is good.

By sorting the item parameters estimated by the RM[10], it is possible to obtain a ranking of items from the one with the best quality score to the one with the least, according to the interpretation of the scale given in the previous paragraph. In our case we can observe that the item with the best quality score is *accessibility of water supply* and the item with the least quality score is *price of fixed telephone service* as it is shown in Table 1.

Items	Location Parameters	Thresholds		
SGIaccW	-1.899	-3.074	-0.725	
SGIaccE	-1.804	-2.866	-0.742	
SGIaccT	-1.473	-2.062	-0.884	
SGIaccG	-1.179	-1.85	-0.509	
SGIquaE	0.277	-1.667	-0.673	3.171
SGIquaG	0.443	-1.214	-0.669	3.213
SGIquaW	0.587	-0.915	-0.503	3.18
SGIquaT	0.68	-0.921	-0.411	3.372
SGIpriG	0.929	0.156	1.703	
SGIpriE	1.045	0.237	1.853	
SGIpriW	1.052	0.403	1.7	
SGIpriT	1.342	0.546	2.137	

Table 1: Items sorted by Rasch Item Location Parameter ordered from the best to the worst quality.

Table 1 shows also the non-centred thresholds, as we expect for the choice of *Extended Logistic Model* the thresholds are not the same for every item and the distance between the thresholds is not constant. In order to facilitate interpretation of the scale of the item parameter in Figures 4 and 5 the Category Probability Curves are plotted. The Pearson location is shown on the horizontal axis and the probability related to each response category on the vertical axis. Figure 41 shows the Category Probability Curves related to the item with the best effective quality (smaller value of location item parameter). We can observe that, apart from person location and therefore apart from satisfaction level, the bigger categories of response are more probable. On the contrary, Figure 5 shows the Category Probability Curves related to the item with the worst effective quality (bigger value of location item parameter). We can observe that, apart from person location and therefore apart from satisfaction level, the smaller categories of response are more probable.

[10] We use the Politomous Rasch Model, in particular the *Extended Logistic Model* (see paragraph 2.2), available in the computer program RUMM (Rasch Unidimensional Measurement Models) by Andrich, Sheridan, Lyne and Luo (2000). It makes scale-free customer measures and sample-free item quality (Andrich 1988; Wright and Masters 1982). Items are calibrated from bad to good and customer measures are aligned, on the same scale, from lower to higher.

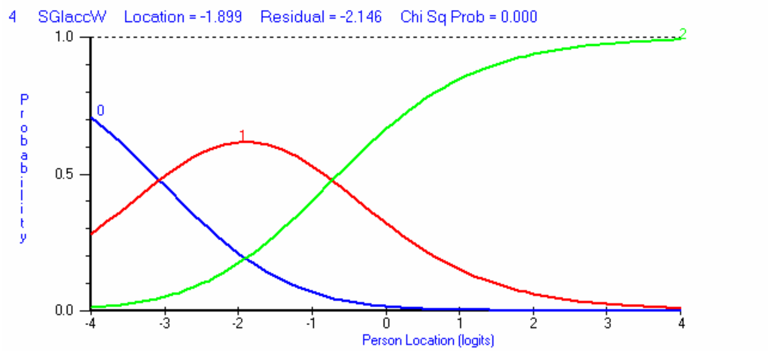


Figure 4. Category Probability Curves for the item *accessibility of water supply*

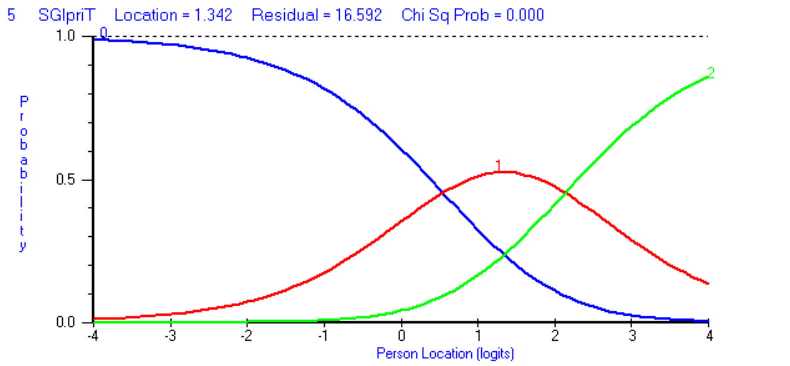


Figure 5. Category Probability Curves for the item *price of fixed telephone service*

Similar analysis is carried out by NLPCA. Component loadings and quantifications of categories are determined. The component loadings, reported in Table 2, assume the meaning of weights of the manifest variables in defining the indicator of satisfaction. So a low weight means a small relevance of the corresponding item in determining the level of satisfaction and detects an incoherent behaviour of the item itself. Table 2 highlights that *accessibility of gas supply* has a low loading, and then a further analysis that investigates the reasons is required (see Fiorio et. Al., 2007). The obtained quantifications are reported in Figure 6 (on the vertical axis) in conjunction with the ordered categories (on the horizontal axis). It can be notice that the hypothesis of equal distance between categories does not hold and for some items a few categories are redundant in a way, because they load to the same level of satisfaction. That happens for example with category 1 (very bad) and 2 (fairly bad) of *quality gas supply* and of *quality water supply*. It is interesting to note that for the same items the thresholds estimated by the RM are also very close (see Table 1).

Items	accT	accE	accG	accW	priT	priE	priG	priW	quaT	quaE	quaG	quaW
Component Loadings	0.45	0.663	0.188	0.61	0.474	0.563	0.478	0.521	0.653	0.747	0.603	0.721

Table 2. Component Loadings

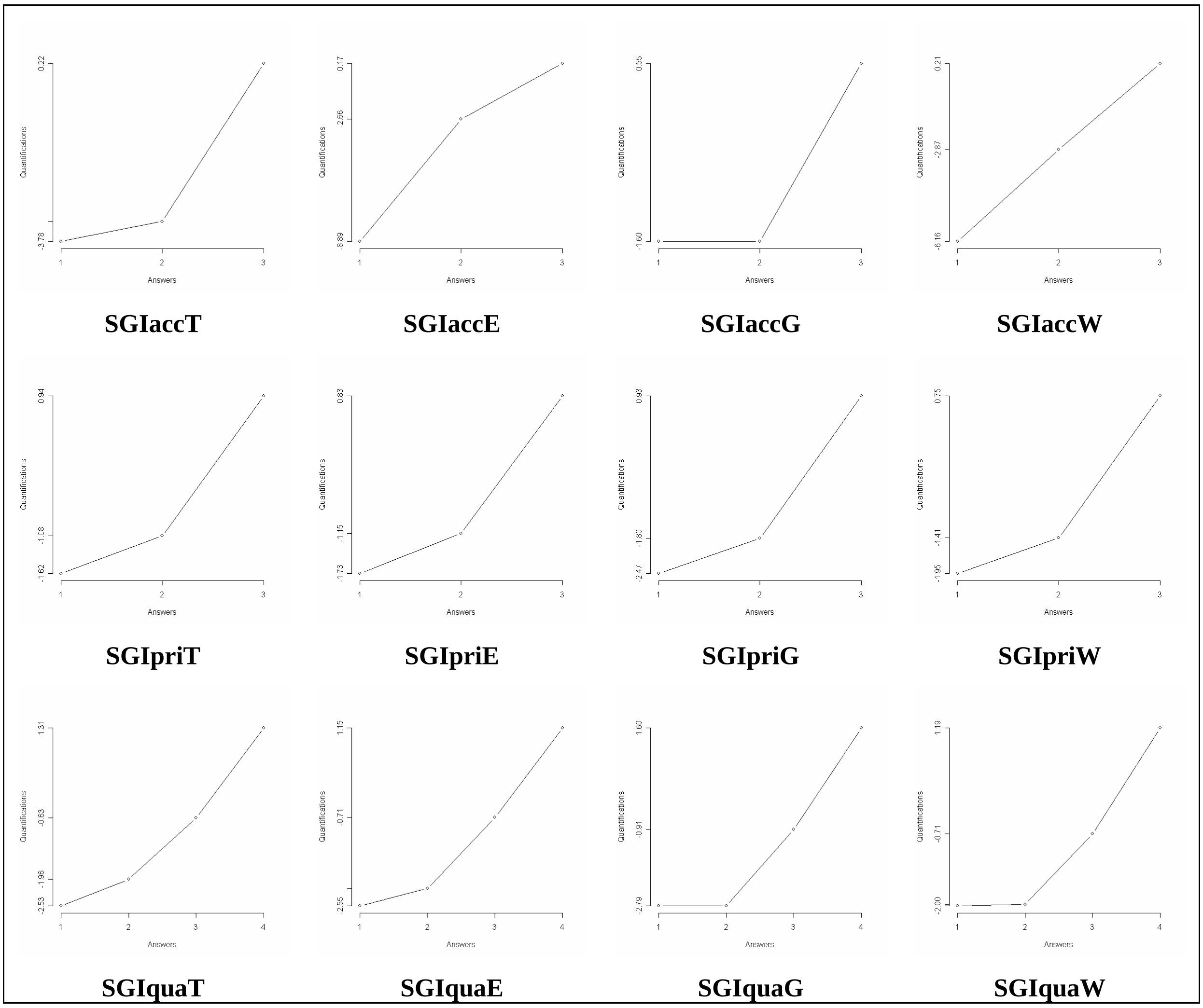


Figure 6. Quantifications of item categories

3.4. Subjects analysis

With subject analysis we mean a study that evidences the level of satisfaction of respondents.

By the θ_i, i=1,…,n, Rasch coefficients related to the persons we can obtain a ranking of the subjects from the most satisfied to the least. Before this, in general, an analysis of residuals is suggested[11]. As mentioned the Rasch person parameter can be interpreted as a global customer satisfaction index that summarises all the services and all the aspects of a service. Figure 7 shows a comparison of satisfaction by *Country*. Italy and Portugal have the lowest level, Ireland and Denmark the highest.

[11] If the mean and the standard deviation (SD) of subject satisfaction overlap the mean and the SD of the item quality, the targeting of the scale is good. The subject average satisfaction (3.009) is greater than the item mean quality (0) and the subject's SD (1.743) is greater than the item's SD (1.22). Therefore, the targeting of the scale seems good. When data perfectly "fit" the model the subject residuals are expected to have zero mean and SD close to 1. In our case the subject residual means are quite good, -0.692 and the subjects residual SD is not so bad (1.305).

Similar scores are obtained by NLPCA. They are sketched in Figure 8 where for each European country the average score in conjunction with its bootstrap percentile confidence interval are reported. The CI is obtained by bootstrap percentile method; the average scores by European countries are ordered from the highest to the lowest.

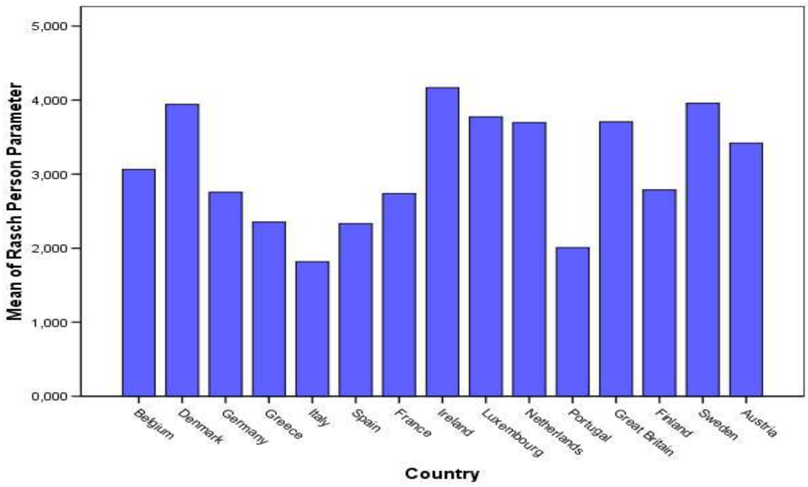


Figure 7. Mean Customer Satisfaction Index by Country.

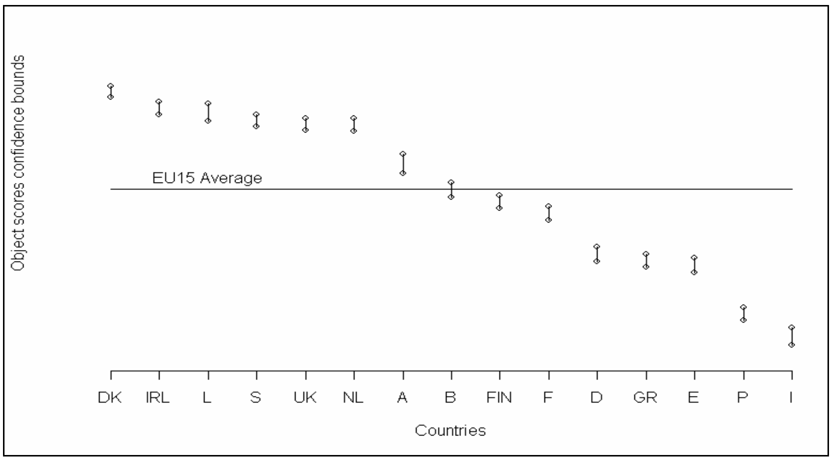
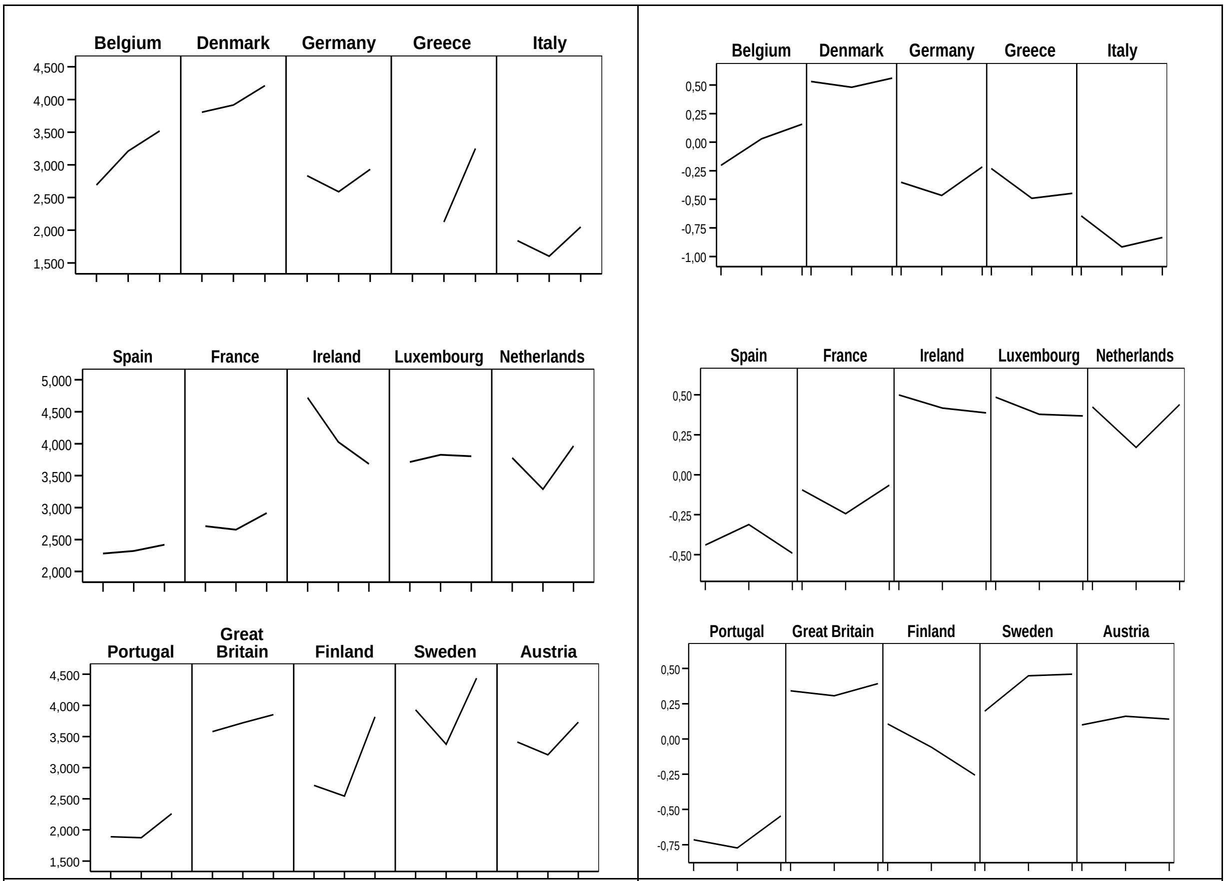


Figure 8. 95% bootstrap CI of Average Satisfaction in of European Countries

In order to analyse in major detail the level of satisfaction, we consider the average satisfaction distinguished by year and country. Figure 9 shows the results for both methods: RM and NLPCA. First of all we can see, for example, that Ireland has a high average overall satisfaction but it is decreasing over the years, whilst Portugal has a low average overall satisfaction but seems to show improvement in 2004. An important remark is that for most countries the two methods show a similar trend, but that does not happen for example with Greece and Finland, pointing them out as hotspots which need specific investigation.



a)Mean Rasch Person Parameter | b)Mean Objective Scores

Figure 9. Average Satisfaction by Country and by Year (2000, 2002, 2004)

3.5. Further Analysis

As mentioned the RM model is frequently used in order to calibrate the questionnaire (De Battisti et. al 2006). In Figure 10 the frequency distribution of raw scores is shown. The distribution is asymmetric in the right. The majority of subjects show a raw score in the 27-36 range respectively the first and the third quartile, with a minimum of 1 and a maximum of 40. This is the first indication that probably some items present one or more redundant answers, as proven in the comments regarding the qualifications of NLPCA.

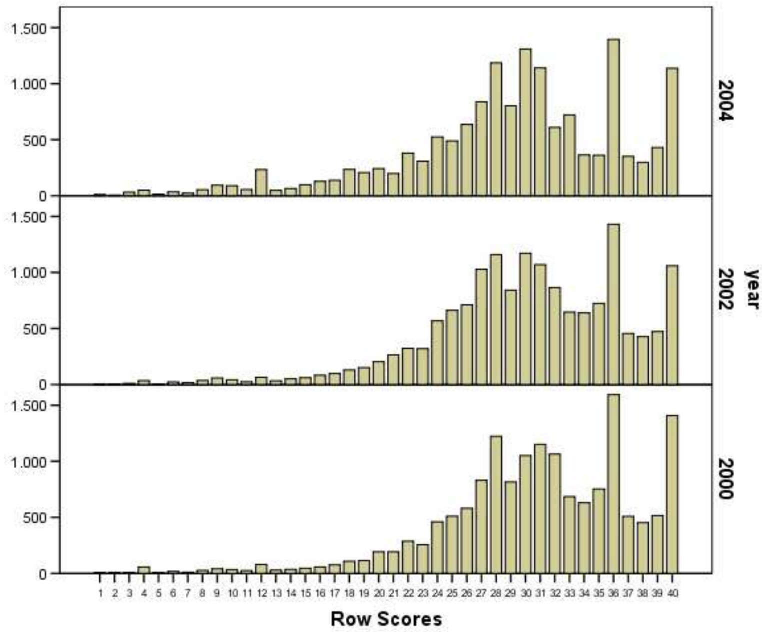


Figure 10. Frequency distribution of raw scores

Figure 11 shows the "Rasch ruler" (also called the "Item map") obtained for analysed data. Items and customers share the same linear measurement units (logits, left column). Subjects are represented on the left of the line with X symbol, thresholds (before and after the dot symbol respectively) are on the right of the line item. The range of items does not entirely match the range of satisfaction scores. There are many subjects at the upper end of the scale but there are no subjects at the lower end. Thus, it does not seem that the item quality is appropriately targeted to subjects satisfaction. Furthermore item thresholds are not well spanned and spaced throughout the continuum. This can be taken as an indicator of low accuracy. With the "same" increase in the satisfaction level there is not the "same" increase in the total raw score. In particular it can be noted that the items at the bottom of the map, in which there is no subject are 1,2,3,4 i.e the item related to the *accessibility*, Moreover, the thresholds at the bottom of the map have low values. So in a calibration intention *accessibility* should be formulated in a different way or separated from the other aspects and the measurement scale would have to be rescored.

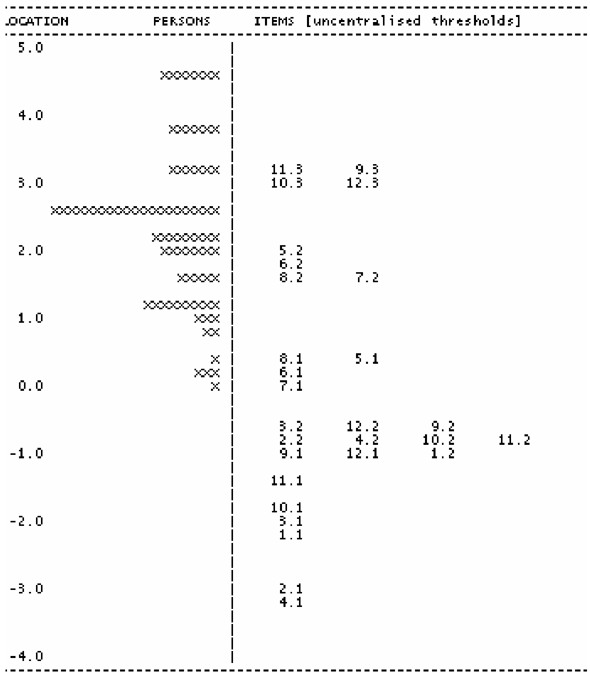


Figure 11. Item Map Overall Model (the map considers also the thresholds)

We will now consider a further contribution of importance supplied by NLPCA. Once the optimal quantification and weight are determined and validated, one can consider the *n* x *m* matrix of transformed variables corresponding to *n* x *m* matrix of the original variables and use these values and relative weights for further statistical analysis, for example for comparing different services or different aspects.

We recalled that the level of satisfaction depends on individual and contextual variables. That makes the comparison among countries difficult because of the presence of these effects. Nonetheless the component loadings and quantifications obtained with NLPCA allow this comparison. Table 3 shows the average overall satisfaction of each country with relative rank (column 2) and the average satisfaction of each service with relative ranks (*between* countries and *within* country). So we can see that Denmark has the highest overall level of satisfaction (rank 1) but not for its gas supply (rank 5). A similar situation is present in Sweden and Finland. An opposite situation is observed in the Netherlands (rank 6 versus rank 1). In Italy dissatisfaction is spread across all services.

The comparison of services within a country seams particularly interesting because in this case the effects of individual and contextual variables is avoidable and a direct comparison is meaningful. Table 3 stresses, once more, the worst situation of the gas supply in Greece. A similar situation is also present in Sweden where this dissatisfaction is more evident in light of the high level of satisfaction for the other services. Figure 12 makes evident this finding and the peculiar case of gas supply service.

Finally, in Table 4 the weighted averages of each aspect are reported in conjunction with their ranks. First of all we can note the non-relevance of *access*: all values are quite near to zero. *Price* and *quality* are instead more important, see Figure 13. On this point, it is evident that in the countries with low or high satisfaction levels, *quality* rather than the *price* makes the difference. That seems to suggest that an analysis of the preferences of users amongst a set of options might give useful indications for the improvement of the service from the viewpoint of its users. For this issue the *ex-ante* approach mentioned in the Introduction and an *ad-hoc* data structure could be useful. Our future researches will be devoted to this application.

Country	Overall Satisfaction	Fixed Telephone	Electricity Supply	Gas Supply	Water Supply
Denmark	0.180 (1)	0.170 (2) [3]	0.226 (1) [1]	0.054 (5) [4]	0.209 (2) [2]
Ireland	0.157 (2)	0.119 (5) [3]	0.194 (2) [1]	0.115 (3) [4]	0.184 (3) [2]
Luxembourg	0.146 (3)	0.138 (4) [3]	0.168 (3) [1]	0.105 (4) [4]	0.162 (4) [2]
Sweden	0.131 (4)	0.174 (1) [2]	0.109 (5) [3]	-0.218 (12) [4]	0.238 (1) [1]
Great Britain	0.125 (5)	0.125 (3) [3]	0.141 (4) [1]	0.129 (2) [2]	0.105 (7) [4]
Netherlands	0.125 (6)	0.108 (6) [3]	0.106 (6) [4]	0.146 (1) [1]	0.141 (5) [2]
Austria	0.075 (7)	0.043 (7) [4]	0.075 (7) [2]	0.047 (6) [3]	0.125 (6) [1]
Belgium	0.006 (8)	0.002 (10) [3]	-0.003 (9) [4]	0.025 (7) [1]	0.003 (9) [2]
Finland	-0.011 (9)	0.006 (8) [2]	0.000 (8) [3]	-0.205 (11) [4]	0.030 (8) [1]
France	-0.037 (10)	-0.022 (11) [3]	-0.018 (10) [2]	0.004 (8) [1]	-0.103 (11) [4]
Germany	-0.079 (11)	0.006 (8) [1]	-0.096 (11) [3]	-0.086 (9) [2]	-0.142 (13) [4]
Greece	-0.126 (12)	-0.109 (12) [2]	-0.165 (13) [3]	-0.263 (15) [4]	-0.063 (10) [1]
Spain	-0.130 (13)	-0.152 (13) [4]	-0.143 (12) [3]	-0.116 (10) [2]	-0.109 (12) [1]
Portugal	-0.216 (14)	-0,206 (14) [1]	-0.214 (14) [2]	-0.220 (13) [3]	-0.220 (14) [3]
Italy	-0.245 (15)	-0.239 (15) [2]	-0.269 (15) [4]	-0.223 (14) [1]	-0.248 (15) [3]

Table 3. Averages satisfaction: overall and for each service with countries ranks (.) and within country ranks [.]

Country	Access	Prices	Quality
Denmark	0.024 (3) [3]	0.161 (3) [2]	0.390 (1) [1]
Ireland	0.006 (5) [3]	0.148 (5) [2]	0.346 (3) [1]
Luxembourg	0.025 (2) [3]	0.184 (2) [2]	0.246 (5) [1]
Sweden	-0.036 (8) [3]	0.125 (6) [2]	0.358 (2) [1]
Great Britain	0.041 (1) [3]	0.210 (1) [1]	0.135 (7) [2]
Netherlands	0.024 (3) [3]	0.150 (4) [2]	0.205 (6) [1]
Austria	-0.070 (12) [3]	0.051 (7) [2]	0.271 (4) [1]
Belgium	-0.037 (9) [3]	-0.030 (10) [2]	0.089 (8) [1]
Finland	-0.091 (15) [3]	0.018 (8) [2]	0.065 (9) [1]
France	0.002 (6) [1]	-0.074 (11) [3]	-0.043 (10) [2]
Germany	-0.101 (14) [2]	-0.010 (9) [1]	-0.122 (11) [3]
Greece	-0.060 (11) [1]	-0.121 (13) [2]	-0.215 (12) [3]
Spain	-0.033 (7) [1]	-0.083 (12) [2]	-0.281 (13) [3]
Portugal	-0.057 (10) [1]	-0.160 (14) [2]	-0.450 (15) [3]
Italy	-0.085 (13) [1]	-0.266 (15) [2]	-0.399 (14) [3]

Table 4. Averages satisfaction: for each aspect with countries ranks (.) and within country ranks [.]

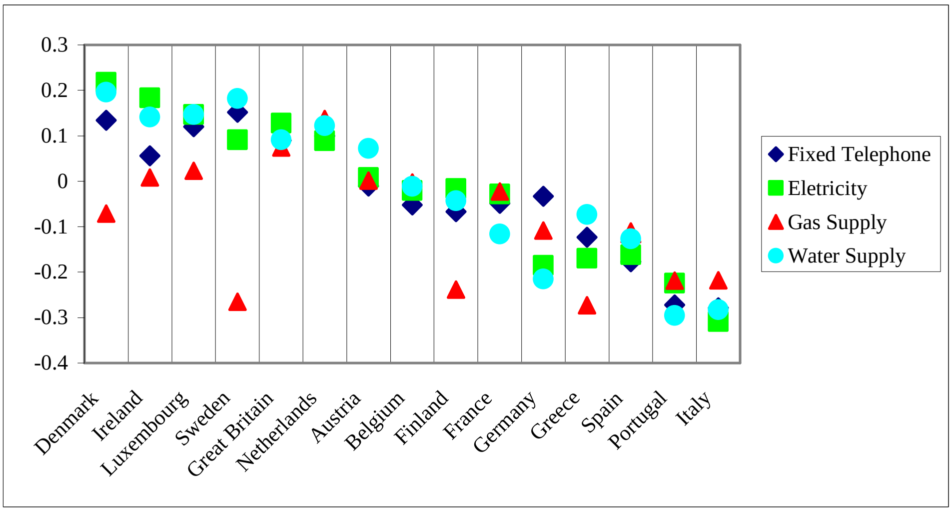


Figure 12. Averages satisfaction for each service

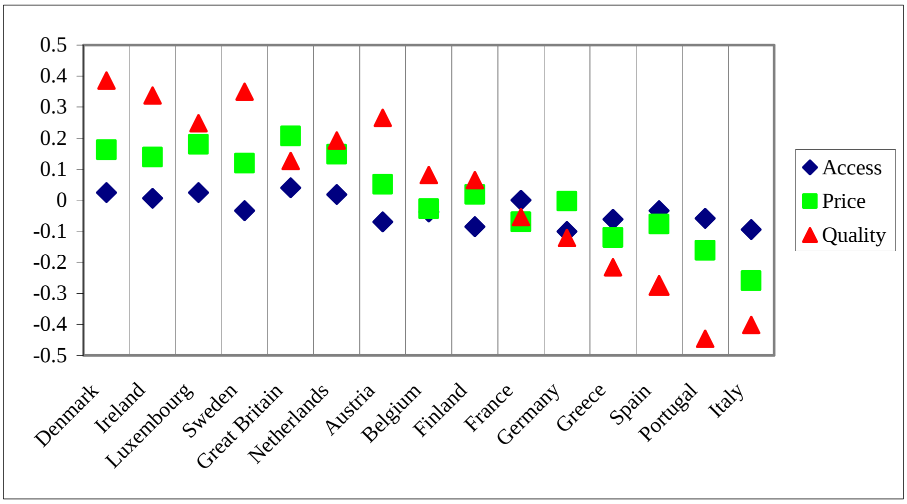


Figure 13. Averages satisfaction for each aspect of service

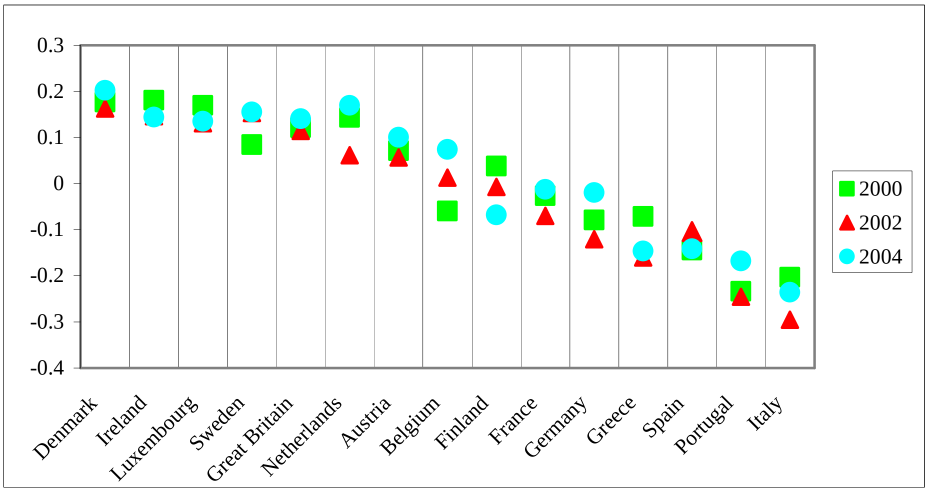


Figure 14. Averages overall satisfaction for each year

In Figure 14 and in Figure 15 the satisfaction indexes are plotted year by year. More specifically in Figure 14 the average overall satisfaction is reported while in Figure 15 the average satisfaction is disaggregated by service and by aspect of services. It seems evident that some countries do not have the same performance in the three years. It is interesting the opposite trend (Figure 14) of Belgium where satisfaction is increasing and Finland where at the contrary satisfaction level is decreasing. In Figure 15 it is one more marked the strange behaviour of gas supply with respect to the other services, therefore the disaggregated view make evident the role of the *Price* in the improvement of satisfaction level in Belgium, while the worsening of satisfaction level in Finland seems generalized.

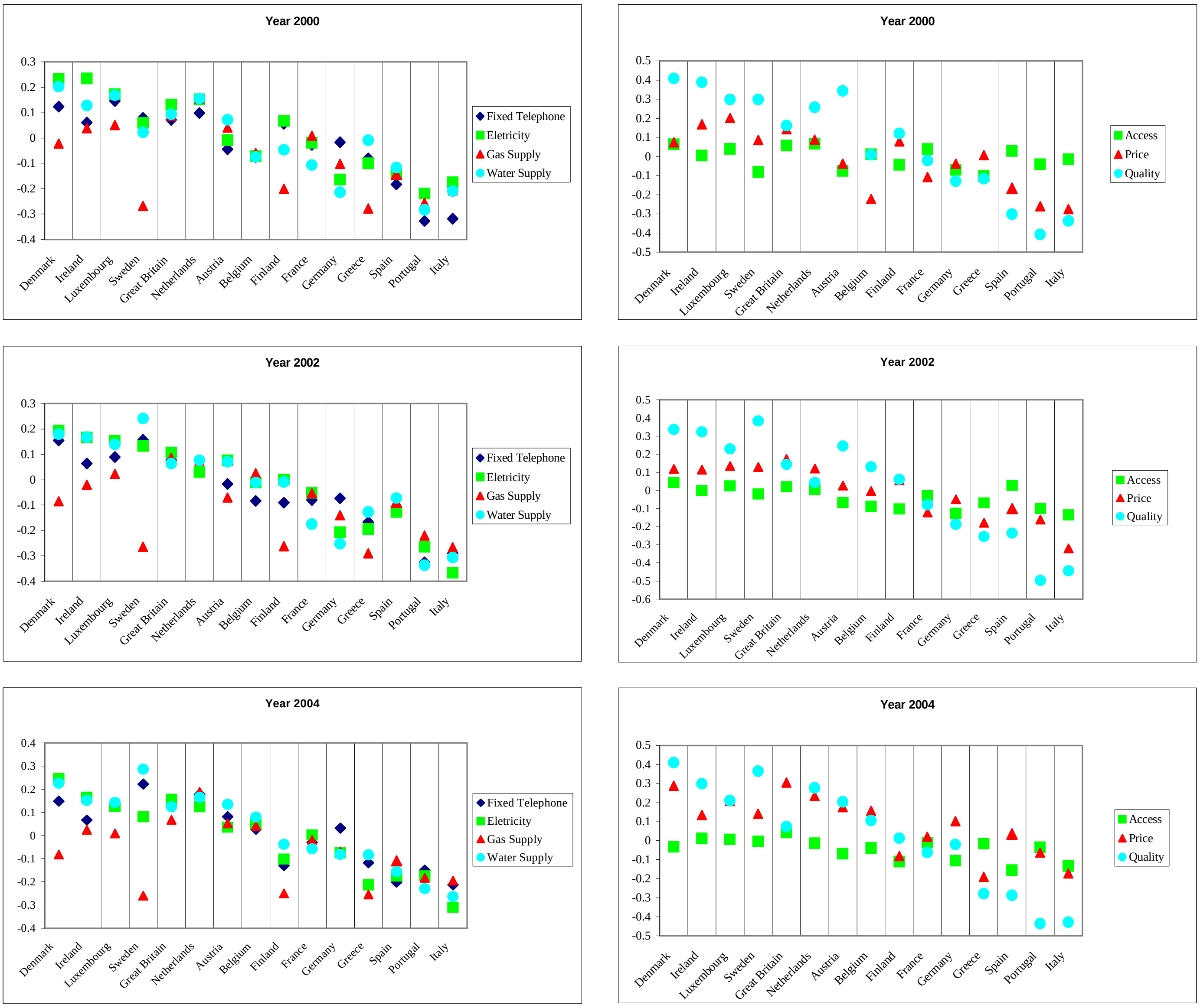


Figure 15. Averages satisfaction for each year, each service, each aspect of service

4. Conclusions

Our starting point is the Consumers' perception of service quality. First we consider the customer satisfaction survey as a measuring instrument of service quality and examine the

Eurobarometer survey data. With this goal in mind, two different techniques are proposed, both able to take into account the order of categories without establishing an *a priori* difference between them and both able to pick up on nonlinear relationships as well. They are: the Rasch Model (RM) and the Nonlinear Principal Components Analysis (NLPCA). The first assumes a model entirely known except for the values of parameters which have to be estimated; the second is instead connected to an algorithmic procedure, no data generating process is assumed but the best representation of data is searched. The two methods allow us a ranking of items to be established, on one hand based on the perceived quality and on the other based on the importance. Moreover they allow for synthetic indicators of the level of satisfaction to be provided for subsequent further analysis. The NLPCA analysis shows that the service aspects play different roles in the European countries evaluation, the analysis of preferences using for example Conjoint Analysis technique (Green and Srinisavan, 1978) could be a helpful approach in measuring and improving the quality of a Service of General Interest as perceived by its users. In our feature researches we intend to realize a simulation study that explains the potential of conjoint analysis in the context of service of general interests. The most difficult task in this case is the definition of service options. Usually a careful review of existing levels of service and alternatives can identify and define the appropriate commodity for the study. In general, preparatory activities and focus groups can uncover important features of the selected service options.

The RM model results suggest that the Eurobarometer questionnaire presents some measuring problems, and therefore a recalibration could be in order or else a different way to measure the Service's quality. In particular for a better measure of accessibility aspect, the question in the survey could be formulated differently (see Special Eurobarometer 226 / Wave 63.1 – TNS Opinion & Social realized for the 10 New European Member State). The fact that consumers that have no the *access* to the service do not give an answer on *quality* and *price* may generate a sample selection bias, statistical methods have to consider this problem.

As a future prospective useful for economist the quantitative variables obtained (θ_i of RM and objective scores for NLPCA) can be used as response variables in interpretative model instead of 12 qualitative variables or a simple aggregation of these (linearity, equidistance between categories and equal weight for the different aspects are implicitly assumed in this case). Covariates in the model could be individual variables (income, sex, etc.) and macro economical variables (GDP, privatization, etc.) as in Fiorio et. al (2007).

5. References

Andrich D. (1978a), *A rating formulation for ordered response categories*, Psychometrika, XLIII, n. 4, 561-573.

Andrich D. (1988a), *Rasch models for measurement*, Sage, Beverly Hills.

Andrich D. (1988b), "A general form of Rasch's extended logistic model for partial credit scoring", in *Applied Measurement in Education* I, 363-378.

Andrich D., Sheridan B., Lyne A. and Luo G. (2000), *RUMM: A windows-based item analysis program employing Rasch unidimensional measurement models*. Perth, Australia: Murdoch University.

Barone S., Lombardo A. (2004). "Service quality design through a smart use of conjoint analysis". *Asian Journal on Quality*, 5, 1, pp. 34-42

Bertrand M. and Mullainathan S. (2001). *Do People Mean What They Say? Implications for Subjective Survey Data.* Papers and Proceedings of the Hundred Thirteenth Annual Meeting of the American Economic Association, Economics and Social Behavior, *The American Economic Review*, Vol. 91, No. 2, 67-72.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests, *Psychometrika*, 16, 297-334.

De Battisti F., Nicolini G., Salini S., (2005). *The Rasch Model to Measure the Service Quality*, The Journal of Services Marketing, vol. III, n° 3, pp. 58-80 ISSN: 0887-6045

De Battisti F., Nicolini G., Salini S., (2007), *Rasch Model In Customer Satisfaction Survey* (2007), submitted.

De Battisti F., Salini S., Crescentini A. (2006), *Statistical Calibration of Psychometric Test*, Statistica & Applicazioni, Vol. IV, n. 2.

Efron, B. (1979). *Bootstrap method: another look at the jackknife*, Annals of Statistics, 7, 1-26.

Eurobarometer (2002) Report No. 58 "Consumers' Opinions about Services of General Interest", December 2002.

Eurobarometer (2004) Report No. 53, "The people of Europe and Services of general interest", October 2000.

Eurobarometer (2006) Report No. 62.1: The Constitutional Treaty, Economic Challenges, Vocational Training, Information Technology at Work, Environmental Issues, and Services of General Interest, October-November 2004.

Ferrari P, Annoni P., Salini S. (2005). *A comparison between alternative models for environmental ordinal data: Nonlinear PCA vs Rasch Analysis*, IWSM 2005, Sydney.

Ferrari P. A., Manzi G. (2007), N*onlinear Principal Component Analysis as a tool for the evaluation of customer satisfaction*, (2007), submitted.

Ferrari, P. A., Annoni, P. and Manzi, G. (2007). *European opinion on services: evaluation and comparison, submitted*, submitted.

Fiorio C.V., Florio M. (2007), *Opening Public Utilities to the market makes people happier? Don't give it for sure,* submitted.

Fiorio C.V., Florio M., Salini S., Ferrari P.A. (2007), *Consumers' Attitudes on Services of General Interest in the EU:Accessibility, Price and Quality 2000-2004* - In: Servizi Pubblici: Nuove Tendenze nella Regolamentazione, nella Produzione e nel Finanziamento ; [a cura di] M. Marrelli, F. Padovano, I. Rizzo. - [s.l.] : Franco Angeli, 2007.

Gifi, A. (1990). *Nonlinear Multivariate Analysis*. Wiley, New York.

Green, P.E. and Srinivasan, V. (1978) 'Conjoint analysis in consumer research: issues and outlook', *Journal of Consumer Research,* Vol. 5, No. 2, pp.103–123.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Press.

Heiser, W. J. and Meulman, J.J. (1994). *Homogeneity analysis: Exploring the distribution of variables and their nonlinear relationship*, in Greenacre M., J. and Blasius, J. (eds.): Correspondence analysis in the social sciences: recent developments and applications, Academic Press, Harcourt Brace & Company New York, 179-209.

Likert, R. (1932). *A Technique for the Measurement of Attitudes*, Archives of Psychology, 140, 595-639.

Marcucci E. And Gatta V. (2007), *Quality and Public Transport Service Contracts*, Società Italiana di Economia Pubblica, Working Paper, No 586.

Masters G.N., (1982), A Rasch model for partial credit scoring, *Psychometrika*, XLVII, n.2, pp. 149-174.

Masters G.N., (1988a), The analysis of partial credit scoring, *Applied Measurement in Education*, I, pp. 279-297.

Masters G.N., (1988b), Measurement models for ordered response categories, in Langeheine R., Rost J. (a cura di) 1988, *Latent trait and latent class models*, Plenum Press, New York.

Michailidis, G. and De Leeuw, J. (1998). The Gifi system of descriptivemultivariate analysis. *Statistical Science*, 13, 307-336.

Nicolini G., Salini S. (2006), *Customer Satisfaction in the Airline Industry: The case of British Airways*, Qual. Reliab. Engng. Int., 22, 1-9.

Pattanayak, S.K., E. Sills, L. Carrasco, J.-C. Yang, C. van den Berge, C. Agarwal, H. Gunatilake, H. Bandara, and T. Ranasinghe. 2005b. The Economic Geography of Water and Poverty: Evidence from Southwest Sri Lanka. RTI Working Paper.

Rasch G., (1960/1980), *Probabilistic models for some intelligence and attainment tests*, (Copenhagen, Danish Institute for Educational Research),expanded edition (1980) with foreward and afterword by B.D. Wright. The University of Chicago Press, Chicago.
Research, Vol. 40, No. 4, pp.406–420.

Special Eurobarometer 219 Wave 62.1 - TNS Opinion & Social, http://europa.eu.int/comm/public_opinion/

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y. M. and Lauro, C. (2005). *PLS Path Modeling*, Computational Statistics and Data Analysis, 48, 159-205.

Wright B.D., Masters G.N. (1982), *Rating scale analysis*, MESA Press, Chicago.

Zanella A. (2001), Valutazione e modelli interpretativi di customer satisfaction: una presentazione d'insieme, *Istituto di statistica, Università Cattolica del Sacro Cuore*, Milano, Serie E.P. n. 105.