AN EMPIRICAL COMPARISON BETWEEN GRADE OF MEMBERSHIP AND PRINCIPAL COMPONENT ANALYSIS

A. SULEMAN

Abstract. It is the purpose of this paper to contribute to the discussion initiated by Wachter about the parallelism between principal component (PC) and a typological grade of membership (GoM) analysis. The author tested empirically the close relationship between both analysis in a low dimensional framework comprising up to nine dichotomous variables and two typologies. Our contribution to the subject is also empirical. It relies on a dataset from a survey which was especially designed to study the reward of skills in the banking sector in Portugal. The statistical data comprise thirty polythomous variables and were decomposed in four typologies using an optimality criterion. The empirical evidence shows a high correlation between the first PC scores and individual GoM scores. No correlation with the remaining PCs was found, however. In addtion to that, the first PC also proved effective to rank individuals by skill following the particularity of data distribution meanwhile unveiled in GoM analysis.

1. **Introduction**

The aim of this paper is to contribute to the discussion on the relationship between principal component analysis (PCA) and the typological GoM analysis introduced by Woodbury and Clive in [21]. This issue was first addressed by Wachter in [19]. In his empirical work, the author used sets of three to nine dichotomous variables drawn from U.S. National Survey of Families and Households (NSFH). He found a strong correlation between the first PC scores and the individual GoM scores, and both scores proved effective to recover an age gradient implicit in NSFH data. In sequel, Wachter attempted to recast the underlaying GoM model with a geometrical formulation, thereby suggesting that the GoM analysis is somehow a version of PCA under certain metrics. These findings hold at least for low dimensional settings which comprise two typologies for any number from three to nine dichotomous variables where the referred study was confined to.

The GoM model was presented in 1974 by Woodbury and Clive as a statistical tool to represent fuzzy partitions. This has been the most used formulation of the model both in theoretical and applied frameworks so far (e.g. [17], [2]). However, it has already been mentioned in literature as an alternative PCA method for discrete data ([5]), although sometimes still in fuzzy context, namely as a mean

Received: November 2011; Revised: March 2012; Accepted: July 2012

Key words and phrases: Grade of Membership, Principal Component Analysis, Fuzzy Partition.

for identifying fuzzy clusters or typologies ([11]). Nevertheless, we did not find any other attempt to connect GoM to PCA as did Wachter in his work cited above.

The work presented in this paper is an attempt to link empirically GoM to PCA in a higher dimensional space than in Wachter's study. In particular, we were interested in understanding to what extent the results achieved in that study can be generalized and whether they provide potential guidelines to explicitly establish a theoretical linkage of the two approaches to data analysis. We kept the original formulation of the GoM model, that is, under the framework of fuzzy clustering. So, we did not consider herein the geometrical formulation referred above. For the purpose of our study, we used a dataset from a survey which was purposefully designed to study the impact of skills on wages in the Portuguese banking sector ([14]). In the survey, supervisors were asked to assess individual skills from a list of 30 skill items in a 5-point Likert scale. Besides that, they also provided professional and socio-demographic information on each individual. However, this additional information was not of concern in the specific context of our research and was not used at all therein.

The use of a fuzzy clustering approach has already been proved suitable to these data ([15]). In particular, if we assume that the population under study can be modeled by a fuzzy K-partition and use the GoM model to identify the underlined fuzzy clusters. The difficulty lies, however, in how to set the value of K prior to applying the GoM model to data as it must be. Differently from the just cited paper, here we used a more objective criterion to fix the value of K, which is based on a measure of the relative goodness of fit (i.e. a likelihood ratio test) of competing GoM models. The empirical evidence shows that a model with $K = 4$ fuzzy clusters provides the best fit to data among all the models considered in this study.

After submitting the dataset to a PCA, we estimated a linear regression model for each of 30 PCs to check out the extent to which these quantities can be predicted by 3 out of 4 non-redunctant GoM scores under ordinary least squares (OLS) method. The results achieved show an almost perfect fit to the first PC. But, on the other side, no other PC was found to fit a linear combination of GoM scores. It turns out, then, the GoM analysis can not be replaced by PCA in our particular study as was the case in Wachter's study which comprised two typologies and consequently only one non-redundant GoM score for each individual.

Apart from the results referred to herein before, the output of GoM model unveiled a hierarchical fuzzy partition of skills. As a post-hoc analysis, we explored then a geometrical counterpart of a fuzzy 4-partition, that is, the unit simplex with 4 vertices. The distribution of individuals on this convex set lead them to be ordered by skill through an utility function which is a linear combination of GoM scores. The coefficients of such combination must, however, meet some specific conditions so as to perform properly and provide individual rank. It is the case that the estimated linear regression model for the first PC meets these conditions and consequently the first PC could additionally be used to rank individuals by skill.

This paper is organised as follows. In the next section we present the data used in our empirical work. The third section describes the GoM model in a fuzzy sets theory perspective. A brief account on the theoretical background underlying PCA is given in section four. The fifth section is devoted to empirical analysis. Finally, some concluding remarks are presented in the last section.

2. **The Data**

The data used in our empirical analysis were compiled from a survey conducted by supervisors of the banking sector in Portugal. The survey was intended to analyse skills rewards of retail bankers in this country ([14]). The supervisors were asked to assess each retail banker in 30 different skill items. They accomplished this task rating individuals in each item through a Likert scale from 1 to 5, with the following meaning: 1: Very Low; 2: Low; 3: Medium; 4: High; and 5: Very High. The 30 variables comprise four different skill dimensions namely of Knowledge, Behaviour and Attitude toward Others, Behaviour and Attitude toward the Organisation and Cognitive and Technical Skills. Table 1 displays the number of variables in each dimension. Detailed specification of skill items is provided in Appendix A. Observed frequencies in each (variable, category) pair are accounted for in Appendix C, under the column labelled Frequency. The final sample size is $N = 600$.

Skill Dimension	Number of Variables
Knowledge	3
Behaviour and Attitude toward Others	6
Behaviour and Attitude toward the Organisation	12
Cognitive and Technical Skills	9
Total	30

Table 1. Number of Variables in Each Skill Dimension

In addition to those variables, the survey provides detailed information on human capital, demographic characteristics, job position and earnings of banking employees. We did not use this additional information here as we were focusing on the way how skill measurement variables cluster together on the basis of GoM analysis as well as of PCA. That is, our approach is methodological in the sense that it aims to analyse whether and the way one data decomposition technique turns into another. However, in a different setting, e.g. in applied research in economics ([15]), we should not ignore that professional and socio-demographic information on individuals as it is commonly used in feature selection process of estimated fuzzy partition sets ([1]).

3. **Grade of Membership Analysis**

We assume that the universe U of banking employees is decomposed into $K \geq 2$ fuzzy sets or fuzzy clusters of skills, $A_1,\ A_2, ...,\ A_K$, that form a fuzzy $K-$partition of U, although we do not know in advance the value of K. Each individual of U, say individual i, is represented in this structure by his / her vector of GoM scores

$$\mathbf{g}_i = (g_{i1}, g_{i2}, ..., g_{iK}) \tag{1}$$

This coordinate vector belongs to the unit simplex

$$S_K = \left\{ \mathbf{a} = (a_1, a_2, ..., a_K) : a_k \geq 0 \wedge \sum_{k=1}^{K} a_k = 1 \right\} \tag{2}$$

The generic GoM score g_{ik} of the vector g_i, in (1), stands for the grade of membership of individual i in fuzzy set A_k. We further assume that each partition set is normal, i.e.,

$$\forall_{1\leq k\leq K}\ \exists_i : g_{ik} = 1$$

In other words, every fuzzy set $A_k,\ 1 \leq k \leq K$, has at least one full member.

In order to estimate the fuzzy $K-$partition that presumably underlies the data, we used the statistical model introduced by Woodbury and Clive ([21]), and known by the acronym GoM (Grade of Membership). It assumes that the number of fuzzy partition sets, K, is fixed a priori. The model can be formulated as follows. Let

$$\mathbf{X}_i = \left(X_{i1}, X_{i2}, ..., X_{iJ}\right), 1 \leq i \leq N \tag{3}$$

be the vector of outcomes or attributes of individual i in J measurement variables, where $X_{ij} \in \{1, 2, .., L_j\}, 1 \leq j \leq J$, is a categorical variable with $L_j \geq 2$ number of categories, and N is the sample size. In our case, $N = 600$, $J = 30$, X_{ij} is the individual i assessment in Likert scale on j^{th} skill item, and $L_j = 5$ for all skill items.

In GoM model the coordinates of X_i (3), that is the outcomes X_{ij}, are, by assumption, ruled in latent form by the vector of GoM scores g_i, as in (1). Given g_i, they are considered independent from each other. This means, $X_{ij}|g_i$ and $X_{ij'}|g_i$ $(j \neq j')$ are independent random variables. Denote by λ_{kjl} the probability of a full member of fuzzy set A_k has the outcome l in j^{th} variable, i.e.,

$$\lambda_{kjl} = \Pr\left[X_{ij} = l \mid g_{ik} = 1\right]$$

where $1 \leq i \leq N; 1 \leq j \leq J; 1 \leq k \leq K; 1 \leq l \leq L_j$. Being probabilities, the λ_{kjl} verify the two conditions

$$\lambda_{kjl} \geq 0 \text{ and } \sum_{l=1}^{L_j} \lambda_{kjl} = 1, \text{ for each } k \text{ and each } j \tag{4}$$

The quantities λ_{kjl} account in probabilistic terms for the importance of different attributes in fuzzy clusters. Their estimates are therefore used in practice to identify typologies associated with those clusters ([12]). In this perspective, fuzzy clusters may be referred to as typologies. A few other very common names for those clusters in the context of GoM model are pure types ([21]) or extreme profiles ([10]).

The basic assumption of GoM is that, given g_i, the probability p_{ijl} of individual i having the outcome l in j^{th} variable is

$$p_{ijl} = \Pr\left[X_{ij} = l \mid \mathbf{g}_i\right] = \sum_{k=1}^{K} g_{ik}\lambda_{kjl} \tag{5}$$

Given a fixed value for the number of fuzzy partition sets, that is K, the model parameters, namely g_{ik} and λ_{kjl}, are estimated by maximization of the logarithm of the multinomial likelihood function,

$$\mathbf{L}_K = \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{l=1}^{L_j}\left(\sum_{k=1}^{K} g_{ik}\lambda_{kjl}\right)^{\eta_{ijl}} \tag{6}$$

subject to the above referred constraints, (2) and (4). In (6), η_{ijl} is an indicator variable which is equals to 1 if $X_{ij} = l$ and 0 otherwise. The optimal solution is achieved iteratively: in each step, the likelihood function is maximized with one set of parameters fixed, say λ_{kjl}, to obtain estimates for g_{ik}, and then the obtained g_{ik} are used in the next step to update λ_{kjl}. This process is repeated until one criterion from a set of predefined criteria for convergence is met (e.g. difference between consecutive estimates, maximum number of iteration).

The model fitness is assessed by a likelihood ratio test

$$T = -2\ln\left(\frac{\mathbf{L}_1}{\mathbf{L}_K}\right) \tag{7}$$

which measures the deviance of the solution of $K \geq 2$ fuzzy clusters L_K (6) from the so-called independence model with a single cluster solution L_1. Under the null hypothesis, i.e. when L_1 is the true model, the statistic T in (7) is approximated by a chi-square distribution with the degrees of freedom equal to number of estimated parameters (see [10], for details). If we denote by υ the number of degrees of freedom then, for a given K,

$$\upsilon = N \times (K-1) + K \times \sum_{j=1}^{J} (L_j - 1) = 600 \times (K-1) + K \times 30 \times 4 \tag{8}$$

Due to high number of degrees of freedom commonly associated with GoM model application, in practice we use a transformation which converts a statistic with a chi-square distribution to a statistic with a distribution closely approximating a standard normal distribution. In Section 5, we will report some results on the referred approximation procedures as well as our option for a particular solution.

4. **Principal Component Analysis**

Principal Component Analysis (PCA) is a statistical method of data reduction. The basic idea of PCA is to reduce the dimensionality of a dataset, which comprises a large number of interrelated variables, so that the variation present in the dataset can be expressed through a smaller number of latent variables called principal components (PCs). The PCs are uncorrelated, and are constructed as linear combinations of the input variables. The number of PCs is equal to the number of original variables. However, PCs are ordered variables in the sense that the first component has as high variance as possible. All succeeding components have the highest possible variance but are subject to the constraint of being uncorrelated with the preceding components. As such, the first few PCs may account for the most of the variability present in all of the original variables. If it is the case, these few PCs can alternatively be used to examine the universe under study. This can be stated formally as follows.

Suppose $\mathbf{X}$ is a vector of J random variables,

$$\mathbf{X} = \left(X^{(1)}, X^{(2)}, ..., X^{(J)}\right) \tag{9}$$

The first PC $Y^{(1)}$ is obtained as a linear function $\mathbf{a}_1^{'}\mathbf{X}$ of components of $\mathbf{X}$, where $\mathbf{a}_1^{'} = \left[\begin{array}{cccc} a_{11} & a_{12} & ... & a_{1J} \end{array}\right]$, that is,

$$Y^{(1)} = \mathbf{a}_1^{'}\mathbf{X} = a_{11}X^{(1)} + a_{12}X^{(2)} + ... + a_{1J}X^{(J)}$$

and has maximum variance subject to the condition $\mathbf{a}_1^{'}\mathbf{a}_1 = 1$. The coordinates of the vector $\mathbf{a}_1$ are referred to as first principal component loadings. The second PC $Y^{(2)}$ is given similarly as a linear combination of components of $\mathbf{X}$, that is,

$$Y^{(2)} = \mathbf{a}_2^{'}\mathbf{X} = a_{21}X^{(1)} + a_{22}X^{(2)} + ... + a_{2J}X^{(J)}$$

which has the highest possible variance subject to the two constraints $\mathbf{a}_1^{'}\mathbf{a}_2 = 0$ and, in the same way as before, $\mathbf{a}_2^{'}\mathbf{a}_2 = 1$. All the remaining components are obtained following the same procedure. If $Y^{(j)}$ and $Y^{(j')}$, $1 \leq j < j' \leq J$, are any two PCs with the vector of loadings $\mathbf{a}_j$ and $\mathbf{a}_{j'}$, respectively, then

$$\text{Var}\left[Y^{(j)}\right] \geq \text{Var}\left[Y^{(j')}\right] \tag{10}$$

and $\mathbf{a}_j^{'}\mathbf{a}_j = 1$ and $\mathbf{a}_{j'}^{'}\mathbf{a}_j = 0$. This later condition means that $Y^{(j)}$ and $Y^{(j')}$ are uncorrelated random variables. In (10), Var $\left[Y^{(j)}\right]$ means variance of $Y^{(j)}$.

Having defined the PCs, the concern is now on the way these new variables are found. Suppose that the vector of original variables $\mathbf{X}$ has a known covariance matrix $\mathbf{\Sigma}$. It can be shown that if $Y^{(j)} = \mathbf{a}_j^{'}\mathbf{X}$, $j = 1, 2, ..., J$, is the j^{th} PC, $\mathbf{a}_j$ is an eigenvector of $\mathbf{\Sigma}$ corresponding to its largest j^{th} eigenvalue α_j. Furthermore, if the loadings vector $\mathbf{a}_j$ is chosen to have unit length, i.e. $\mathbf{a}_j\mathbf{a}_j^{'} = 1$, then Var $\left[Y^{(j)}\right] = \alpha_j$. So, $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_J$.

In most practical cases the covariance matrix $\mathbf{\Sigma}$ is unknown and it is replaced by the sample covariance matrix $\mathbf{S}$. It is also common in practice to derive the PCs from the correlation matrix ρ using the standardised version of observed variables instead. Readers interested in more details about PCA may consult [8] where this section was mostly derived from.

Even not explicitly mentioned, PCA is designed for real value data, that is, $\mathbf{X}$ in (9) is a vector of $\mathbb{R}^J$. It turns out that our data are categorical though ordinal. So, prior to apply PCA to our dataset we shall assume that the skill items are vectors of $\mathbb{R}^J$ ($J = 30$), where the Euclidean distance makes sense as does, consequently, the calculation of empirical covariance or correlation matrices. Although this is convenient for practical purposes, we have no formal statistical justification for such assumption.

The question now being asked is how to link GoM to PCA. Extending the work developed in [19], we aim to find a linear regression model that expresses a relation between the two data analysis techniques. For j^{th} PC, we calculate individual PC scores beforehand as follows,

$$Y_i^{(j)} = \mathbf{a}_j^{'}\mathbf{X}_i = a_{j1}X_{ij} + a_{j2}X_{ij} + ... + a_{jJ}X_{ij}, \quad 1 \leq i \leq N; 1 \leq j \leq J$$

where $\mathbf{X}_i$ is individual i observed skill items vector as in (3). Then we estimate a regression model between each PC and $(K - 1)$ non-reductant GoM scores,

$$Y_i^{(j)} = \beta_0^{(j)} + \beta_1^{(j)}g_{i1} + \beta_2^{(j)}g_{i2} + ... + \beta_{(K-1)}^{(j)}g_{i(K-1)} + \varepsilon_i^{(j)} \tag{11}$$

where $\varepsilon_i^{(j)}$ are the error terms and, by assumption, are normally and independently distributed with zero mean and σ^2 variance. The intercept term $\beta_0^{(j)}$ and the regression coefficients $\beta_1^{(j)}$, ..., $\beta_{(K-1)}^{(j)}$ are estimated by ordinary least squares

(OLS) method and the model (11) goodness of fit is assessed through the adjusted coefficient of determination, R^2_{adj}.

5. **Empirical Evidence**

5.1. **GoM - PCA Relation.** The estimation process of GoM model parameters, g_{ik} and λ_{kjl}, was conducted through DsiGoM Software ([6]). In order to define the optimal number of fuzzy clusters K, we first ran this computer program on a trial basis, ranging the value of K from 2 to 5, and realised that the best fit to the data was achieved for $K = 4$. With the exception of the model based on $K = 3$ fuzzy clusters, the optimal solution was attained in less than 100 iterations, for a maximum fixed error of 10^{-6}. In case of $K = 3$, it took $1,285$ iterations. It may be the case that the estimation algorithm is hindered to estimate all independent parameters involved in GoM model. Whenever this happens some degrees of freedom are lost. The DsiGoM Software reports the effective number of degrees of freedom involved in each estimation process. Details about practical estimation aspects of GoM model are found in [10].

The goodness of fit was assessed through the test statistic T as in (7) using an approximation of chi-square distribution to normal distribution due to excessive number of degrees of freedom (df) or, equivalently, the number of parameters involved. In this regard we considered three different approximations to standard normal distribution ([4]) as follows:

$$\begin{aligned}
Z_1 &= \sqrt{2T} - \sqrt{2\upsilon - 1} \\
Z_2 &= \frac{(T/\upsilon)^{1/3} - (1 - 2/9\upsilon)}{\sqrt{2/9\upsilon}} \\
Z_3 &= Z_2 + 8Z_2^5 \ / \ 100\upsilon^3
\end{aligned}$$

where T is the value of test statistic obtained through the relation (7) and υ is the number of degrees of freedom. The approximation Z_1 is due to Fisher ([7], quoted in [4]), the approximation Z_2 is due to Wilson and Hilferty ([20]) and the approximation Z_3 is due to Kelley ([9], quoted in [4]). In any case, the bigger the value of the statistic Z_m $(m = 1, 2, 3)$ the higher the deviance from the independence model. Table 2 displays the empirical values found for each statistic for different values of K. The GoM model with $K = 4$ fuzzy clusters seems to fit the data consistently better than any other competing model.

K	T	υ (df)	Z_1	Z_2	Z_3
2	$10,007$	750 (840)	102.75	79.71	79.79
3	$13,806$	$1,500$ $(1,560)$	111.41	90.03	90.05
4	$16,837$	$2,250$ $(2,280)$	116.43	96.20	96.21
5	$17,662$	$3,000$ $(3,000)$	110.49	93.62	93.62

Table 2. GoM Model Goodness of Fit Assessment for Different Values of K. Column 3 Displays Both the Effective Number of Degrees of Freedom and, in Parenthesis, the Number of Freedom Implied by Formula (8)

Needless to say, all four tests lead to the rejection of the null hypothesis, that is the true model is the one with a single cluster, for the level of significance of 5% ($Z_m > 1.645,\ m = 1, 2, 3$).

Having obtained the estimates of GoM model based on a solution of $K = 4$ fuzzy clusters, we submitted our dataset to a PCA with the aim of assessing how the two data analysis techniques are related to each other. The PCs were derived from both the sample covariance matrix and correlation matrix (see results in Appendix B). In sequel, we calculated individual PC scores based on the two forms of PCs. The results achieved so far show a closer relationship to PCs derived from the correlation matrix, in particular to the first PC. So, this form of PCs are considered in more detail here.

We estimated a regression model for the first PC scores to understand how this quantity can be predicted by GoM scores. The model is represented by the equation

$$Y_i^{(1)} = \beta_0^{(1)} + \beta_1^{(1)} g_{i1} + \beta_2^{(1)} g_{i2} + \beta_3^{(1)} g_{i3} + \varepsilon_i^{(1)},\ i = 1, 2, ...N = 600 \tag{12}$$

which is adapted from (11) to the particular case of $K = 4$. The OLS estimates of model parameters as well as relevant associated statistics, related to the first PC computed from the correlation matrix, are displayed in Table 3. All parameters were found to be significant at a 5% significance level. The goodness of fit of the regression model (12) was assessed through the adjusted coefficient of determination, R^2_{adj}. The empirical value found for R^2_{adj} is 0.985, which shows almost a perfect fit of the first PC scores to a linear combination of GoM scores. When we used the first PC derived from the sample covariance matrix we found a slightly lower value for that statistic, namely 0.970. Recall that PCA is not scale invariant. The inspection of the distribution of residuals, i.e. predicted minus observed values, through a normal probability plot (Figure 1), gives some room for questioning the normality assumption of error terms in (12). The fact that GoM scores are constrained to lie within unit simplex S_k (2) is probably one of the factors contributing to curve residuals away from the normal line in places. The S shape curve as seen on Figure 1 signals heavier tails than a normal distribution.

			95% Confident Interval	
Parameter	Estimate	Std. Error	Lower Bound	Upper Bound
β_0	1.995	0.021	1.953	2.037
β_1	−4.446	0.031	−4.506	−4.386
β_2	−2.661	0.025	−2.710	−2.612
β_3	−1.341	0.028	−1.396	−1.287

TABLE 3. First PC Regression Model Estimates

In the continued work on the relation between PCA and GoM, we estimated a regression model similar to (12) for the remaining 29 PCs. Curiously, the values of the adjusted coefficient of determination were very low and some negative values were found. The highest value was 0.030. Before our research, we thought $(K - 1)$ out of J PCs would be linearly related to $(K - 1)$ non-reductant GoM parameters somehow in the line with Wachter's work ([19]). This would potentially lead to the estimation of GoM parameters through PC scores. However, the empirical evidence

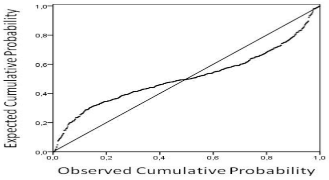

Figure 1. Probability Plot of Standardised Residuals

shows that this is not the case. Thus, PCA is not an alternative to GoM at least for our particular study.

5.2. **Pos-Hoc Analysis.** In the second stage of our empirical analysis, we explored the estimated fuzzy 4−partition from a distributional point of view. For this purpose we looked at the estimates of individual vector of GoM scores $\mathbf{g}_i = (g_{i1}, ..., g_{i4})$ and made use of the unit simplex S_4 (2) which can be represented geometrically by a tetrahedron. Full fuzzy set members position on the vertices of this convex set; partial members of two fuzzy clusters position on the edges of S_4 while individuals who share three fuzzy clusters are placed on the faces. All remaining individuals position in the interior of S_4. We can number the vertices of S_4 so that the vertex k corresponds to fuzzy cluster A_k, $k = 1, 2, 3, 4$. As such, the edge $1-2$ corresponds to individuals who are partial members of fuzzy clusters A_1 and A_2; the edge $2-3$ to individuals who share fuzzy clusters A_2 and A_3, and so forth. The same reasoning applies to faces and to the interior points of S_4. Table 4 gives account of the results obtained from the application of GoM model to our dataset. The same information is provided graphically in Figure 2.

Part of S_4						Total (600)
Vertex						44 (7.3%)
	1	2	3	4		
	12	13	7	12		
Edge						300 (50.0%)
$1-2$	$2-3$	$3-4$	$1-3$	$1-4$	$2-4$	
60	157	81	2	0	0	
Face						211 (35.2%)
	$1-2-3$	$2-3-4$	$1-3-4$	$1-2-4$		
	100	104	5	2		
Interior						45 (7.5%)
45						

Table 4. Decomposition of the Unit Simplex S_4

After analysing the distribution of retail bankers in the unit simplex S_4, we may conclude that 57% lie on the path of edges $1-2-3-4$, including the vertices. Combined with this, a further 31% lie in a 0.85 vicinity of this path, that is their estimated GoM scores obey the relation

$$g_{ik} \geq 0.85, \quad k = 1, 2, 3, 4 \text{ or } g_{ik} + g_{i(k+1)} \geq 0.85, \quad k = 1, 2, 3 \tag{13}$$

The value 0.85 should be understood as a threshold for "strong agreement" ([3]) with either one or two consecutive fuzzy clusters and it is set subjectively. Explicitly, we first consider all members whose grade of membership in any fuzzy set is equal or greater than 0.85. Next, we account for the members for whom the sum of two consecutive GoM scores obey the same condition, i.e. ≥ 0.85. In each step, we exclude all members who have already been accounted for on earlier steps to avoid double counting. It can therefore be concluded that about 90% of individuals position close or on the path of edges $1-2-3-4$. This is an important finding as we shall see further ahead.

At the other extreme of GoM representation of our dataset, we looked at the λ_{kjl} parameter estimates so as to identify skill typologies latent in the Portuguese banking sector. Empirical evidence shows that the four typologies are somehow ordered by skill from lowest (A_1) to highest (A_4). Indeed, we found out that full fuzzy sets members are likely to have increasing skill rates as we move from A_1 to A_4 in all but one observed variable, namely variable number 21, and in particular for fuzzy set A_3 (see Appendix C for the estimates of λ_{kjl} parameters). This finding therefore leads us to label fuzzy sets A_1, A_2, A_3 and A_4 as Low (skill), Medium, High and Very High, respectively. As a consequence, individuals who lie on the oriented path of edges going from the vertex 1 to the vertex 4, that is the path $1 \to 4$ of unit simplex S_4, are increasingly more skilled (Figure 2). This fact allows, in particular, to rank these individuals by skill as we shall subsequently see.

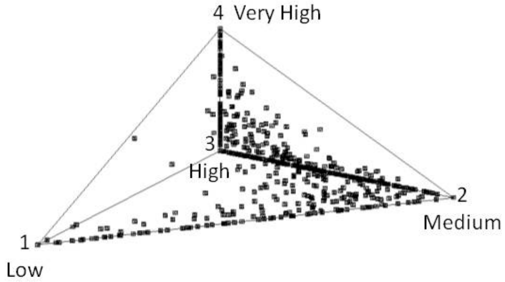

Figure 2. Estimated Distribution of Retail Bankers on the Unit Simplex S_4

Consider the following function of GoM scores

$$\rho\left(i\right)=b_0+\sum_{k=1}^{4}b_k g_{ik} \tag{14}$$

It can be shown that ρ strictly increases in the oriented path $1 \to 4$ as long as $b_k < b_{k'}$, where $1 \leq k < k' \leq 4$ (see [15], for a demonstration of an equivalent result). That is, ρ is an (ordinal) utility function for the oriented path $1 \to 4$. Curiously, the first PC turns out to meet the conditions that have been just mentioned, that is $b_k < b_{k'}$. Explicitly, looking at β estimates in Table 3, we realise that not only they are statistically significant but also verify the inequalities $\beta_1 < \beta_2 < \beta_3$. Furthermore, all of them are negative. Setting $\beta_4 = 0$ and replacing each b coefficient in (14) by corresponding β estimates leads to the first PC. Thus, the first PC can be used as an analytical tool to provide a point estimate of individual skills rank for, at least, the 57% of individuals estimated to lie on the oriented path $1 \to 4$ ([15]). If we extend this possibility to individuals with strong agreement with either one or two consecutive fuzzy sets in the same sense as in (13), the first PC would fairly behave as an ranking device for almost all sampled individuals (Fig. 2). Of course, this PC could be used for ranking purposes without resorting the fuzzy approach. However, the individual PC score would probably not sound the same in that case. Furthermore, the first PC gives by itself no indication about the underneath hierarchical skill structure as does the GoM analysis. In summing up, combining GoM with PCA led to deeper insight into skill data structure.

6. Concluding Remarks

In this study, we have tried to contribute to the discussion about the relationship between PCA and GoM following the pioneer work by Wachter. Differently from this author, we used a rather more complex dataset and have found that GoM and PCA are linearly related to each other only through the first PC and not otherwise. The same relation found by the referred author in two-dimensional space appears now to be unrelated to the space dimension. The exclusive relation to the first PC could be explained by the fact that PCs are uncorrelated by nature. However, there is no any apparent theoretical impediment to, for example, the second PC be linearly related to GoM as does the first PC, being their respective coefficients orthogonal in the similar way as are their loadings. Quite apart from this, the results achieved so far encourage practitioners to use both techniques in a complementary way as well as give pointers for a theoretical research on the subject concerned.

In our on going research, we are trying to ”invert” the GoM analysis in a way to construct new variables that directly relate the latent parameters g_{ik} and λ_{kjl} and the observed variables $\mathbf{X}_i$. Our purpose is to find out any particular relation that leads such variables to PCs. The results we achieved so far are not conclusive and need further development.

Meanwhile, we believe that using PCA and GoM techniques in a complementary basis potentially enhance the quality and the granularity of data analysis. Clearly, this is the main message of our study.

Acknowledgements. The author would like to thank Fátima Suleman for providing the data on skill assessment in banking sector in Portugal, and Marina Coutinho and Elizabeth Reis for their many suggestions and comments. An earlier version of this paper was presented in The 58th World Statistics Congress of ISI - 2011 ([16]) and the author is indebt to peers for their queries. He is also grateful to FCT - Foundation of Science and Technology of Ministry of Science, Technology and Higher Education for the financial support under FLEX FCT grant PTDC/EGE-ECO/108547/2008.

Appendix A. **List of Skill Items**

The 30 skill items used to access retail bankers belong to four different dimensions, namely Knowledge; Behaviour and Attitude toward Others; Behaviour and Attitude toward the Organisation; Cognitive and Technical Skills, as follows.

Skill Dimension	Items Specification
Knowledge	1. General technical knowledge 2. Specific technical knowledge 3. Foreign languages
Behaviours and Attitudes toward Others	4. Relationship with colleagues 5. Team working 6. Communication skills 7. Willingness to help others 8. Negotiation skills 9. Persuasion skills
Behaviours and Attitudes toward Organisation	10. Perseverance and goal-oriented attitudes 11. Client-oriented attitudes 12. Autonomy 13.Responsibility 14. Adaptability 15. Innovative attitudes 16. Favourable learning attitudes 17. Proactive attitudes toward learning 18. Following rules and procedures 19. Cooperation (with organisational goals) 20. Working time flexibility 21. Punctuality
Cognitive and Technical Skills	22. Work planning 23. Computer skills 24. Analytical skills 25. Ability to select and process information 26. Problem solving 27. Learning ability 28. Ability to transfer knowledge and experience 29. Ability to understand the banking specificities 30. Ability to understand corporation strategy

Appendix B. **Principal Component Analysis Output**

We present the PCs computed from the sample correlation matrix and omit the components derived from covariance matrix to avoid information overload. In the table below, the column 3 labelled "% of Variance" contains the percent of variance accounted for by each PC, and column 4 labelled "Cumulative %" contains the cumulative percentage of variance accounted for by the current and all preceding PCs. Only four PCs have eigenvalues equal to 1 or greater. These four components together account for 69.16% of the total variance.

Component	Eigenvalue	% Variance	Cumulative %
1	**16.88**	56.25	**56.25**
2	**1.61**	5.35	**61.61**
3	**1.20**	4.01	**65.62**
4	**1.06**	3.54	**69.16**
5	0.77	2.57	71.73
6	0.70	2.33	74.06
7	0.64	2.12	76.17
8	0.62	2.06	78.23
9	0.52	1.74	79.97
10	0.50	1.65	81.62
11	0.46	1.53	83.15
12	0.43	1.43	84.59
13	0.40	1.32	85.91
14	0.36	1.21	87.12
15	0.35	1.17	88.29
16	0.34	1.14	89.44
17	0.33	1.11	90.54
18	0.32	1.06	91.60
19	0.29	0.96	92.56
20	0.27	0.90	93.47
21	0.26	0.85	94.32
22	0.24	0.79	95.11
23	0.22	0.74	95.85
24	0.21	0.71	96.56
25	0.21	0.70	97.27
26	0.19	0.63	97.89
27	0.18	0.61	98.51
28	0.16	0.55	99.05
29	0.15	0.51	99.56
30	0.13	0.44	100.00

Appendix C. **Partial GoM Model Output**

We present here the estimates of λ_{kjl} parameters denoted by $\hat{\lambda}_{kjl}$, for $k = 1, 2, 3, 4$. For the purpose of fuzzy clusters characterization we follow closely the criteria established in [2]. Thus, excepting for high frequencies, the variable-category (j, l) pair is considered to substantively contribute to discriminate the fuzzy cluster A_k if $\hat{\lambda}_{kjl} \geq (1 + \delta) \times f_{jl}$, where f_{jl} is the observed relative frequency of that pair. We fixed $\delta = 0.2$ in this empirical study. For high frequencies, the (j, l) pair is tagged as a discriminant condition whenever the conjunction $\left(\hat{\lambda}_{kjl} \geq f_{jl}\right) \wedge \left(\hat{\lambda}_{kjl} \geq 0.9\right)$ holds. The observed frequencies are listed under the column labelled ”Frequency”. The discriminant conditions are displayed in boldface figures. Where applicable, a code 0 is added for missing data. Non-observed categories are omitted. All figures are represented in ×100% format.

Variable (j)	Category (l)	Frequency	$\hat{\lambda}_{1jl}$	$\hat{\lambda}_{2jl}$	$\hat{\lambda}_{3jl}$	$\hat{\lambda}_{4jl}$
1. General technical knowledge	0. missing	1.33	0.00	3.93	0.00	0.00
	1. Very Low	2.87	**17.41**	1.35	0.00	0.00
	2. Low	17.23	**74.69**	19.70	0.00	0.00
	3. Medium	39.36	0.00	**75.70**	27.07	17.90
	4. High	34.97	7.90	3.26	**72.93**	36.43
	5. Very High	5.57	0.00	0.00	0.00	**45.67**
2. Specific technical knowledge	0. missing	0.67	0.00	1.96	0.00	0.00
	1. Very Low	1.68	**10.46**	0.88	0.00	0.00
	2. Low	17.79	**80.05**	**21.62**	0.00	0.00
	3. Medium	36.58	9.49	**77.50**	20.92	9.45
	4. High	38.09	0.00	0.00	**79.08**	**48.18**
	5. Very High	5.87	0.00	0.00	0.00	**42.38**
3. Foreign languages	0. missing	10.00	**15.74**	11.43	7.54	8.34
	1. Very Low	4.81	**37.57**	0.00	0.00	0.00
	2. Low	19.44	**51.79**	**24.53**	10.28	0.93
	3. Medium	44.63	7.53	**58.20**	46.81	37.47
	4. High	27.22	3.10	17.27	**40.60**	**36.57**
	5. Very High	3.89	0.00	0.00	2.30	**25.04**
4. Relationship with colleagues	2. Low	5.17	**48.33**	0.00	0.00	0.00
	3. Medium	21.17	**39.86**	**53.10**	0.00	0.00
	4. High	55.83	11.81	46.90	**100.00**	3.19
	5. Very High	17.83	0.00	0.00	0.00	**96.81**
5. Team working	1. Very Low	0.67	**5.84**	0.64	0.00	0.00
	2. Low	7.83	**68.57**	0.00	0.00	0.00
	3. Medium	23.67	25.60	**70.42**	0.00	0.00
	4. High	53.33	0.00	28.93	**100.00**	0.00
	5. Very High	14.50	0.00	0.00	0.00	**100.00**
6. Communication skills	1. Very Low	0.50	**4.52**	0.00	0.00	0.00
	2. Low	6.17	**55.81**	0.00	0.00	0.00
	3. Medium	34.17	34.36	**82.34**	0.00	0.00
	4. High	50.83	5.31	17.66	**100.00**	25.91
	5. Very High	8.33	0.00	0.00	0.00	**74.09**
7. Willingness to help others	0. missing	0.17	**0.75**	0.54	0.00	0.00
	1. Very Low	0.83	**6.83**	0.00	0.00	0.00
	2. Low	8.68	**71.06**	0.00	0.00	0.00
	3. Medium	23.04	22.11	**65.82**	0.00	0.00
	4. High	53.26	0.00	34.18	**100.00**	5.56
	5. Very High	14.19	0.00	0.00	0.00	**94.44**
8. Negotiation skills	0. missing	1.00	0.00	1.98	0.00	2.71
	1. Very Low	2.36	**15.22**	0.00	0.00	0.00
	2. Low	13.13	**84.78**	0.00	0.00	0.00
	3. Medium	40.07	0.00	**100.00**	8.66	0.00
	4. High	36.70	0.00	0.00	**91.34**	35.78
	5. Very High	7.74	0.00	0.00	0.00	**64.22**
9. Persuasion skills	0. missing	1.00	0.00	2.43	0.00	1.46
	1. Very Low	2.53	**14.42**	0.00	0.00	0.00
	2. Low	14.98	**85.58**	0.00	0.00	0.00
	3. Medium	44.78	0.00	**100.00**	31.69	0.00
	4. High	31.99	0.00	0.00	**68.31**	**51.43**
	5. Very High	5.72	0.00	0.00	0.00	**48.57**
10. Perseverance and goal-oriented attitudes	0. missing	0.17	0.00	0.49	0.00	0.00
	1. Very Low	2.67	**20.51**	0.00	0.00	0.00
	2. Low	10.35	**79.49**	0.00	0.00	0.00
	3. Medium	34.56	0.00	**100.00**	3.53	0.00
	4. High	41.74	0.00	0.00	**96.47**	0.00
	5. Very High	10.68	0.00	0.00	0.00	**100.00**
11. Client-oriented attitudes	0. missing	0.33	0.00	0.00	0.00	2.76
	1. Very Low	0.84	**6.58**	0.56	0.00	0.00
	2. Low	9.87	**77.69**	0.00	0.00	0.00
	3. Medium	30.60	15.72	**84.10**	0.00	0.00
	4. High	45.65	0.00	15.34	**100.00**	7.28
	5. Very High	13.04	0.00	0.00	0.00	**92.72**
12. Autonomy	0. missing	0.17	0.00	0.49	0.00	0.00
	1. Very Low	1.84	**13.21**	0.00	0.00	0.00
	2. Low	11.69	**84.05**	0.00	0.00	0.00
	3. Medium	35.73	2.74	**89.52**	13.64	0.00
	4. High	42.90	0.00	10.48	**86.36**	27.68
	5. Very High	7.85	0.00	0.00	0.00	**72.32**
13. Responsibility	0. missing	0.17	0.40	0.59	0.00	0.00
	1. Very Low	1.34	**11.82**	0.00	0.00	0.00
	2. Low	7.85	**69.47**	0.00	0.00	0.00
	3. Medium	30.72	18.71	**83.37**	0.00	0.00
	4. High	48.58	0.00	16.63	**100.00**	6.59
	5. Very High	11.52	0.00	0.00	0.00	**93.41**
14. Adaptability	0. missing	0.17	0.40	0.59	0.00	0.00
	1. Very Low	1.34	**12.32**	0.00	0.00	0.00
	2. Low	8.01	**73.94**	0.00	0.00	0.00
	3. Medium	28.38	13.74	**100.0**0	0.00	0.00
	4. High	53.26	0.00	0.00	**100.00**	0.00
	5. Very High	9.02	0.00	0.00	0.00	**100.00**
15. Innovative attitudes	0. missing	0.17	0.00	0.49	0.00	0.00
	1. Very Low	2.17	**10.66**	0.00	0.00	0.00
	2. Low	18.20	**89.34**	0.00	0.00	0.00
	3. Medium	43.91	0.00	**100.00**	30.13	0.00
	4. High	32.55	0.00	0.00	**69.87**	**72.63**
	5. Very High	3.17	0.00	0.00	0.00	**27.37**

Variable (j)	Category (l)	Frequency	$\hat{\lambda}_{1jl}$	$\hat{\lambda}_{2jl}$	$\hat{\lambda}_{3jl}$	$\hat{\lambda}_{4jl}$
16. Favourable learning attitudes	1. Very Low	1.67	**14.97**	0.00	0.00	0.00
	2. Low	7.17	**64.38**	0.00	0.00	0.00
	3. Medium	24.33	20.64	**82.26**	0.00	0.00
	4. High	54.33	0.00	17.74	**100.00**	0.00
	5. Very High	12.50	0.00	0.00	0.00	**100.00**
17. Proactive attitudes toward learning	0. missing	0.17	1.31	0.00	0.00	0.00
	1. Very Low	1.84	**14.10**	0.00	0.00	0.00
	2. Low	11.19	**85.90**	0.00	0.00	0.00
	3. Medium	27.05	0.00	**100.00**	0.00	0.00
	4. High	49.75	0.00	0.00	**100.00**	0.00
	5. Very High	10.18	0.00	0.00	0.00	**100.00**
18. Following rules and procedures	1. Very Low	0.50	**4.53**	0.60	0.00	0.00
	2. Low	3.83	**34.72**	0.00	0.00	0.00
	3. Medium	26.00	**51.03**	**56.84**	4.82	0.00
	4. High	59.33	9.72	42.57	**95.18**	0.00
	5. Very High	10.33	0.00	0.00	0.00	**100.00**
19. Cooperation (with organisational goals)	0. missing	0.17	0.00	0.49	0.00	0.00
	1. Very Low	1.17	**10.91**	0.00	0.00	0.00
	2. Low	9.35	**87.27**	0.00	0.00	0.00
	3. Medium	29.55	1.82	**100.00**	0.00	0.00
	4. High	47.91	0.00	0.00	**100.00**	0.00
	5. Very High	12.02	0.00	0.00	0.00	**100.00**
20. Working time flexibility	0. missing	0.67	3.06	0.82	0.00	0.00
	1. Very Low	2.01	**15.36**	0.00	0.00	0.00
	2. Low	6.21	**47.37**	0.00	0.00	0.00
	3. Medium	18.46	10.46	**52.39**	0.00	0.00
	4. High	51.85	22.00	47.61	**80.54**	4.64
	5. Very High	21.48	5.23	0.00	19.46	**95.36**
21. Punctuality	0. missing	0.33	0.00	0.00	0.81	0.00
	1. Very Low	1.00	**8.00**	0.00	0.00	0.00
	2. Low	3.01	**9.48**	5.49	0.00	0.00
	3. Medium	11.20	**14.49**	**25.19**	2.68	0.00
	4. High	38.96	39.18	**51.50**	43.88	0.00
	5. Very High	45.82	28.86	17.82	53.44	**100.00**
22. Work planning	0. missing	0.83	0.00	0.82	1.34	0.00
	1. Very Low	1.51	**9.78**	0.00	0.00	0.00
	2. Low	13.95	**90.22**	0.00	0.00	0.00
	3. Medium	41.18	0.00	**100.00**	0.00	0.00
	4. High	37.14	0.00	0.00	**100.00**	43.44
	5. Very High	6.22	0.00	0.00	0.00	**56.56**
23. Computer skills	1. Very Low	1.33	**12.15**	0.00	0.00	0.00
	2. Low	6.83	**62.29**	0.00	0.00	0.00
	3. Medium	32.17	25.38	**71.77**	9.29	0.00
	4. High	45.83	0.17	28.23	**82.17**	18.70
	5. Very High	13.83	0.00	0.00	8.54	**81.30**
24. Analytical skills	0. missing	0.50	0.00	0.81	0.55	0.00
	1. Very Low	1.01	**7.06**	0.00	0.00	0.00
	2. Low	13.23	**92.94**	0.00	0.00	0.00
	3. Medium	43.05	0.00	**100.00**	0.00	0.00
	4. High	38.02	0.00	0.00	**100.00**	**57.33**
	5. Very High	4.69	0.00	0.00	0.00	**42.67**
25. Ability to select and process information	0. missing	0.33	0.00	0.98	0.00	0.00
	1. Very Low	0.67	**5.26**	0.00	0.00	0.00
	2. Low	12.04	**94.74**	0.00	0.00	0.00
	3. Medium	42.14	0.00	**100.00**	0.00	0.00
	4. High	39.97	0.00	0.00	**100.00**	**53.16**
	5. Very High	5.18	0.00	0.00	0.00	**46.84**
26. Problem solving	0. missing	0.17	0.00	0.49	0.00	0.00
	1. Very Low	1.00	**7.06**	0.00	0.00	0.00
	2. Low	13.19	**92.94**	0.00	0.00	0.00
	3. Medium	39.40	0.00	**100.00**	0.00	0.00
	4. High	39.57	0.00	0.00	**100.00**	36.25
	5. Very High	6.84	0.00	0.00	0.00	**63.75**
27. Learning ability	1. Very Low	1.00	**9.11**	0.00	0.00	0.00
	2. Low	5.33	**48.58**	0.00	0.00	0.00
	3. Medium	27.00	**42.31**	**82.26**	0.00	0.00
	4. High	54.83	0.00	17.74	**100.00**	0.00
	5. Very High	11.83	0.00	0.00	0.00	**100.00**
28. Ability to transfer knowledge and experience	0. missing	0.33	2.28	0.00	0.00	0.00
	1. Very Low	1.17	**12.73**	0.00	0.00	0.00
	2. Low	8.03	**82.27**	0.00	0.00	0.00
	3. Medium	39.97	0.00	**100.00**	0.00	0.00
	4. High	46.15	0.00	0.00	**100.00**	48.66
	5. Very High	4.68	0.00	0.00	0.00	**51.34**
29. Ability to understand the banking specificities	1. Very Low	0.67	**6.72**	0.53	0.00	0.00
	2. Low	6.67	**67.18**	0.00	0.00	0.00
	3. Medium	36.50	26.10	**95.37**	0.00	0.00
	4. High	50.00	0.00	4.10	**100.00**	19.02
	5. Very High	6.17	0.00	0.00	0.00	**80.98**
30. Ability to understand corporation strategy	1. Very Low	1.67	**17.24**	0.00	0.00	0.00
	2. Low	8.00	**82.76**	0.00	0.00	0.00
	3. Medium	37.17	0.00	**100.00**	0.00	0.00
	4. High	46.17	0.00	0.00	**100.00**	29.45
	5. Very High	7.00	0.00	0.00	0.00	**70.55**

References

[1] A. Andreotti, N. Minicuci, P. Kowal and S. Chatterji, *Multidimensional profiles of health status: an application of the grade of membership model to the world health survey,* PLoS ONE 4(2): e4426 (2009) (DOI:10.1371/journal.pone.0004426).

[2] L. Berkman, B. Singer and K. Manton, *Black / White differences in health status and mortality among eldery*, Demography, **26** (1989), 661–678.

[3] C. J. Bezdek, *Cluster validity with fuzzy sets*, Journal of Cybernetics, **3(3)** (1974), 58–73.

[4] J. R. Brown, *Error analysis of some normal approximations to the chi-square distribution,* Journal of the Academy of Marketing Science, (DOI: 10.1007/BF02729388), **2(3)** (1974), 447–454.

[5] W. Buntine and A. Jakulin, *Applying discrete PCA in data analysis*, Proceeding UAI '04 Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, (available on-line http://portal.acm.org/citation.cfm?id=1036851), (2004), 59–66.

[6] Decision System, Inc., *User documentation for DSIGoM*, Version 1.0, 1999.

[7] R. A. Fisher, *Statistical methods for research workers*, Oliver and Boyd, Edinburgh, 1925.

[8] I. T. Jolliffe, *Principal component analysis*, Springer-Verlag New York Inc., 2nd Edition, 2002.

[9] L. T. Kelley, *Fundamentals of statistics*, Harvard University Press, Cambridge, 1947.

[10] K. G. Manton, M. A. Woodbury and D. Tolley, *Statistical applications using fuzzy sets*, John Wiley & Sons, Inc, 1994.

[11] K. G. Manton and X. Gu, *Disability declines and trends in medicare expendidure*, Ageing Horizons, **2** (2005), 25–34.

[12] A. L. McCutcheon, *Latent class analysis*, Sage Publications, 1987.

[13] E. Mehdizadeh, S. Sadi-Nezhad and R. Tavakkoli-Moghaddam, *Optimization of fuzzy clustering criteria by a hybrid pso and fuzzy c-means clustering algorithm,* Iranian Journal of Fuzzy Systems, **5(3)** (2008), 1–14.

[14] F. Suleman, *O valor das competências: um estudo aplicado ao sector bancário*, Livros Horizonte, Lisboa, 2007.

[15] A. Suleman and F. Suleman, *Ranking by competence using a fuzzy approach*, Quality and Quantity, **46(1)** (2012), 323-339.

[16] A. Suleman, *Grade of membership and principal components analysis: a comparative empirical study*, (available on-line http://isi2011.congressplanner.eu/pdfs/950735.pdf).

[17] D. Tolley and K. G. Manton, *Large sample properties of estimates of a discrete grade of membership model*, Annals of Institute of Statistical Mathematics, **44** (1992), 85–95.

[18] R. Viertl and D. Hareter, *Fuzzy information and stochastics,* Iranian Journal of Fuzzy Systems, **1(1)** (2004), 43–56.

[19] K. W. Wachter, *Grade of membership models in low dimensions*, Statistical Papers, **40** (1999), 439–457.

[20] E. B. Wilson and M. M. Hilferty, *The distribution of chi-square*, Proceeding of the National Academy of Sciences, **17** (1931), 684–688.

[21] M. A. Woodbury and J. Clive, *Clinical pure types as a fuzzy partition*, Journal of Cybernetics, **4** (1974), 111–121.

Abdul Suleman, Department of Quantitative Methods, Instituto Universitário de Lisboa (ISCTE - IUL), BRU-UNIDE, Av. Forcas Armadas, Lisbon, Portugal

E-mail address: abdul.suleman@iscte.pt