AGGREGATION CRITERIA IN THE SERVICE OF RISKS' MODELING

Aurélien Couloumy – ADDACTIS Worldwide

Avenue Paul Henri Spaak 17
B-1060, Brussels, Belgium

aurelien.couloumy@addactis.com

Abstract

This paper presents a concrete solution to optimize input creation process for risk modeling while improving quality and thoroughness of model results by dealing with clustering on dataset liability. It proposes by the way an analysis of the influence of the aggregation criteria on these dataset liabilities.

The framework of the study will be the European prudential regulation Solvency 2 applied in a case of a unit-linked life insurance contract integrated in a "Standard Formula" model.

In the first part of the paper we will present a short reminder concerning classification and we will make some comparison of different clustering families in the case of an insurance application. Then, a presentation of the study setting up will be performed. This presentation will be accompanied by a process proposal whose objective will be to fix the main steps of an insurance liability clustering. Finally we will observe effects of aggregation criteria on capital requirements and model calculation time, and we will compare results between the different choices of clustering strategies and the per head case.

Key words

Aggregation criteria - Risk Modeling – Solvency 2 - Regulatory capital – Standard formula – Establishment of a model – Saving – Clustering standards – Model points – Classification – Hierarchical ascending classification –– K-means – Data analysis – Principal component analysis

1. Introduction

At the age of data digitization, the insurance market players see their volume of information multiplying massively. Data related to products, markets, companies' stocks, are industrialized in datawarehouses to guarantee at best the conservation of the information and its reliability. Paradoxically, when it is used, this level of exactitude of the information causes a problem during calculation of solvency capital requirement of insurance companies. Indeed, if the regulation is considered nowadays more relevant and closer to risks really incurred, it is not without confines for the insurers: the use of huge amount of information earlier in the process, the number and the frequency of publications intended to the legislator… Many technical and temporal issues that have to be compared to the heavy functioning of processes and models.

Facing this issue, technical teams had to propose simplified assumptions aimed at reducing the volume of initial data while preserving a right level of information. Most of the time, classical statistical methods like weighted average are used to deduce typical profiles, named model points (mp), among the liability table. They have the benefit to be easily understood as much mathematically as computationally. Nevertheless they are sometimes not satisfying in terms of results quality, quite binding, they can't propose an optimal solution and, above all, they are not under control because the information level which is lost during clustering is not known.

Thus, another solution will be proposed into this paper: data analysis and cluster analysis on dataset liability of models.

2. Choice of aggregation methods

2.1. Context

The classification concept, generally linked to data analysis is very useful in all scientific fields. The subject could be about animal species segmentation or individuals' behavior, but the goal remains often the same: to research segmentation for merging homogenous observations, called class, according to objective and selective criteria.

These issues frequently happen in the insurance environment and particularly in calculation connected to the new prudential European standard Solvency 2. However, few insurers use cluster classification, most of time because they think mathematically too complex to implement. In this paper, in addition of a classical method, 2 methods' families[1] have been retained: hierarchical ascending classification (HAC, with 4 variations: single link, average link, complete link, Ward

[1] Other methods could be considered like Gaussian Mixture Models (GMM). The goal of the GMM is to move from a geometric concept to a probabilistic concept by analyzing the density of each vector (or observation) of the dataset. But probabilities of class belonging in outputs of the algorithm could be hard to use. To have more elements about the GMM you can read BIERNACKI C. (2009), « Pourquoi les modèles de mélange pour la classification », CNRS & Université de Lille 1 or GOVAERT G. (2008), « Modèle de mélange et classification ».

criteria) and k-means aggregation. These 2 methods have been principally selected because they are set up in the most statistical software used by insurers and because literature on them is rich.

2.2. About classical method

The classical method which will be used consists by choosing discriminant variables, average variables and additive variables in order to reduce liability dimensions. More especially, discriminant variables are chosen according expert opinions, reserves, and premium. Other variables of the same type are considering additives and aren't taken into account in the classification choice, and the other variables, considering less important are simply averaged. So there are approximately as much model points as possibilities of existing values of discriminant variables.

2.3. Hierarchical Ascending Classification

The hierarchical ascending classification algorithm begins with a check of calculation distance between all observations by pair. Once each distance is determined, the partition of n singletons which correspond to n observations of dataset is used as starting point for successive aggregation of closer observations or classes. It's important to calculate again the distance between all objects and class newly clustered for each successive aggregation. The process is reapplied until obtaining one unique class. The main difficulty of this method is to refresh the distance matrix for each successive aggregation.

Several criteria (named links) exist to recalculate the distance matrix. Four usual criteria have been selected for the study: the single link, the average link, the complete link and the Ward criteria[2]. To be more specific according these criteria, let's take instances.

We consider Ω as a collection of individual, A and B are two classes of Ω, i and j are two ordinary elements of A and B. At the end of an iteration of the HAC, we can define the new distance D between A and B as:

- For the simple link:

$$D(A\,,B) = \min_{i\epsilon A\,,\ j\epsilon B} d(i,j)$$

- For the complete link:

$$D(A\,,B) = \max_{i\epsilon A\ ,j\epsilon B} d(i,j)$$

[2] More explanations and the corresponding algorithms are available in BOUBOU M. (2006), « Contribution aux méthodes de classification non supervisée via des approches pré topologiques et d'agrégations d'opinions » thesis, Université Claude Bernard Lyon 1.

- For the average link:

$$D(A\,,B) = \frac{1}{Card(A).Card(B)}.\sum_{i\epsilon A\ ,i\epsilon B} d(i,j)$$

- For the Ward criteria :

$$D(A\,,B) = \frac{Card(A).Card(B)}{Card(A)+Card(B)}.d(g_A\,,g_B)$$

Where g_A and g_B are the respective barycenter of A and B, and d the initial metric.

2.4. Concerning the k-means

For the k-means aggregation, a number of k classes is initially fixed and a partition is obtained from k centers. The choice of centers and the partition can be realized thanks to knowledge about class observations a priori, but this division could be done randomly too, by distributing at random observations in k classes. From that moment, the gravity center g_q is obtained for each class q which belongs to the group Q made of k classes. Each observation i is relocated to the class $C(i)$ which has the closest gravity center as presented below:

$$C(i) = q \text{ if and only if } d\left(i, g_q\right) = \min_{r\,\epsilon Q}(\,d\,(\,i, g_r))$$

The process is repeated until there is no more modification in the classes' composition and thus we can obtain the best final classification according to a fixed number of classes[3].

2.5. Benefits and disadvantages of methods

a. Basic comparisons

About the classical method, as said in the introduction, it's easy to implement. All the same, the information loss isn't controlled which means you can't measure your aggregation quality before modeling. In addition, according to the method explanation, it's all the more easy to realize the clustering when individuals have discriminate variables with few values. This means that the method won't be well fitted for inhomogeneous dataset which is very binding in the case of an insurance liability portfolio. Another issue is the obligation to define different types of variables (discriminant, additive, or average) and to make some choice in relation to which variable is "essential" (according expert opinions) for your modeling: that includes making sensitivity studies on your model which is not necessary for the other methods.

[3] For more information about this algorithm you can read ROUX M. (1985), *Algorithme de classification*, Edition Masson, (and more especially the demonstration of the Huygens theorem which explains the interest of the k-means algorithm), or the publication BENZECRI J.P (1964), « Analyse factorielle des proximités », Institut de Statistique de l'Université de Paris.

The most part of authors considers HAC as the classical cluster family: the algorithm is quite simple and it can be adapted to many distances and for different types of data. Its disadvantage is the necessity to recalculate the distance matrix at each iteration. This is restricting, because it's very expensive in term of calculation time and memory space. The first consequence of this phenomenon is the difficulty to use it for a voluminous dataset. Therefore, the method is not very suitable with our insurance issue and this is all the more important the dataset we will choose in the study has an important dimension. Otherwise the family method is well known to give good result quality.

On the contrary, the benefit of the k-means family consists exactly in the optimization of the calculation by avoiding the recalculation of the distance matrix. Nevertheless, two problems follow from this method. The first is the impossibility to obtain with certainty an absolute optimum, which is the most suitable solution. To overcome this lack, several solutions have been proposed. It is possible to realize the complete algorithm many times, with different initial partitions and to keep the most suitable. The second disadvantage is the initial partition. If information is available about the way to split dataset initially, it's essential to take it into account to begin the algorithm. But, as it is not always the case, the question of initial observation assignment has to be asked. These two disadvantages aren't very disturbing for achieving our goal.

b. Focus on class number selection

To remind, this statistical development is done in order to reduce time process and improve result quality by decreasing dimensions of liability data set. The final number of classes you would like to obtain is therefore an important element. Against this background, two options can be considered: you exactly know the number of class you want to achieve (because of technical requirements linked to the modeling process, or because of a preliminary study on information conservation maximization), or you don't know yet but you want to realize some tests by proposing several levels of clustering. So, in both case, the control of final class number is required.

Families' methods have different ways to treat this fact. First, in the classical method, you are not free to choose the exact number of final classes at the beginning (or at least, not without uncertain and time consuming manipulations on dataset) because of this number depends on the number of possibilities of your discriminant variables values. That is really disappointing because it means you can't pilot easily your degree of liability reduction. HAC family has an unsupervised approach which means you don't define a priori the number of clusters you want to obtain, but you can choose a posteriori the truncation level, among all truncation levels possibilities. And finally, the k-means family has a supervised approach, so you have to define at beginning the number of classes.

This fact shows one of the limitations of the classical methods compared to clustering algorithms.

For more information about distinguish between families' methods, a comparison sum up table, based on the study aggregation results (and a previous study[4]), is proposed in the appendix (appendix 1).

2.6. Use of cumulative strategies

Thus, each method has benefits without being optimal anymore. For instance, it's detrimental to not use HAC methods for important dataset while these methods present a high quality of aggregation. To compensate this fact and to combine effectively benefits of all classical methods' families, cumulative strategies are proposed in literature[5]. In the study, we propose to test, in addition of using each method separately, this cumulative strategy : use of k-means method then a HAC method which takes as initial points classes' barycenter previously founded in order to avoid dimension problem of the HAC method.

Of course, others cumulative strategies could be suggested like using a HAC method and then k-means (but this strategy suffers of the HAC dimensional limitation default) or using a k-means, then a HAC and then a k-means (but the operation would be a little chop logic for a very little increase of aggregation quality).

2.7. Quality indicators

As we said previously the aim is to give advice on choices among aggregation criteria and techniques to define best aggregation level on liability portfolio too. In this way, a special attention must be given to the result quality (in addition of initial inputs dimension and time of calculation) and the means to measure it.

In this way it seems good to define at least, one quality indicator. A good aggregation means obtaining a partition where observations of the same class are rather similar and where observations of different classes are definitely not the same. To testify mathematically of the clustering quality, R^2 indicator is proposed[6]. It's the proportion of explained variance of classes:

$$R^2 = \frac{I_{inter}}{I_{tot}} \leq 1$$

Other indicators could be suggested to observe quality results such as the Cubic Clustering Criterion (CCC)[7], the pseudo-F statistic or the pseudo-F t2 statistic. For brevity reason, we will only use the R^2 indicator in the study.

[4] Another study dealing approximately with the same subject has been realized two years ago: to refer to COULOUMY A. (2013), « Critères d'agrégations de jeu de données passif pour les calculs sous la Formule Standard de Solvabilité 2», actuarial master's thesis, ISFA.
[5] Source : Publication GONZALEZ P.L (2008), « Méthodes de classification », Cnam.
[6] Source : CHEVALIER F. and LE BELLAC J. (2013), « La classification », Université de Rennes 1.
[7] More information about CCC are available in SARLE W.S. (1983), « Cubic Clustering Criterion ».

3. Setting up of the application

3.1. Modeling and assumptions

The tool which has been used to make calculation is ADDACTIS Modeling, software dedicated to risk management modeling and actuarial calculations. The choice of this tool is motivated by the willingness to obtain the largest number of information in output of our model just by one click (even if just a part of them will be presented here).

About model, it allows modeling a saving contract with multiple pools, and managing asset and liability. As part of Solvency 2, it calculates *Best Estimate*, Net Asset Value (NAV), Solvency Capital Requirement (SCR) for each concerned risk, Basic Solvency Capital Requirement (BSCR) and so on...

This model takes into account several lines of the model points. The calculations are done in run-off, in other words there are no new underwritings during the simulation. Otherwise, the portfolio is dependent on reinsurance agreement and deferred taxes are not taken into account. It should be noted that the flows are updated with the risk-free rate given by the EIOPA yield curve. Assets and liabilities are projected simultaneously on 40 years.

The product which has been modeled is a unit-Linked life insurance contract. Premiums, free or scheduled, paid by the policyholder, are used to purchase unit-linkeds of one or more of the seven funds chosen at the constitution of the contract. The contract provides:

- A benefit in case of life at the term of the contract. This benefit is simply equivalent to the amount of the mathematical reserve at that moment;
- A benefit in case of redemption before the term of the contract. This benefit is also equivalent to the amount of the mathematical reserve at that moment minus a possible penalty;
- A benefit in case of death before the term of the contract. This benefit corresponds to the mathematical reserve.

Different types of expenses are considered such as: commissions, lapse fees, entry fees on premiums, management fees treated by act, discount fees for the different brokers.

We assume finally that there is no change of investment policy for the policyholder. The asset portfolio is composed of cash, mortgage loan, equities, bonds (corporate and government) and properties. The initial market value of theses assets are defined in the assumptions. The repartition which is proposed is a classical repartition of the market.

3.2. Dataset

Dataset which has been studied is a portfolio of 35 000 individuals, described by 29 variables including fifteen potential discriminant variables, nine additive variables and five other non-discriminant variables. We can mention: gender, birth date, contract start date, premiums, reserves, lapse fees, discount fees, entry fees, contract term…

Variables are all supposed quantitative: it is generally the case for calculation linked to the solvency.

3.3. Process proposal

This part contains a suggestion of a process based on dataset liability reduction before calculation for the standard formula of Solvency 2. Obviously this process could be adapted to any type of process which needs large table in model assumptions. In addition, some other points must be specified:

- The study doesn't take into account dataset quality issues: we assume data quality is good (that means: every individuals in the data set are conserved);
- There aren't specifying policies about exclusion of little classes.

a. Preliminary study

First, a preliminary study is realized on the dataset to know elementary statistics and to understand observations.

b. Observation on pretreatment

It may be important to do a pretreatment on the dataset for instance by standardizing variables[8]. In our case, standardization has been done.

Another aspect of the pretreatment for the dataset is to use principal component analysis (PCA). This concept is used to choose "relevant" variables among description variables. The term "relevant" is quoted here to remember a specific goal: having the best description of individuals. Taking irrelevant variables, forgetting relevant variables or not taking account of correlations could obscure the dataset structure by introducing noise, and could increase time of the next step of calculation. In addition, PCA allows us to use the Euclidian distance.

Eigen values

A table which sums up eigenvalues obtained for each factors for the PCA is available in the appendix (appendix 2). Now, number of factors must be chosen according to dimensional issues (for aggregation algorithms), and according to the loss dispersion tolerance. A split is visible on this chart after the factor 4. It shows a dispersion break for observations between factor 4 and factor 5. All the same, only 57% of the dispersion is included in the first four factors. The question is to know if it's enough to not deteriorate information (and so, results of the solvency calculation).

To study this question, we propose 2 options (in a normal case, we would just select one):

- In O1, 4 factors are used after the PCA. They conserve 57% of initial information and they substantially reduce dataset dimension;

[8] Some variables can have bigger values than others. Given this fact, they will have a significant role in distance calculations. To reduce this scale effect, standardization can be necessary.

- In O2, 15 factors are used after the PCA. They conserve 100% of initial information and they don't reduce dimension of dataset (same number of variables as the initial per head table).

Contributions

Another pertinent point can be observed about these two options: contribution of variables for each factor. A table which sums up the contributions is able just below:

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
Gender	0,493	0,341	0,104	7,236	0,015	3,473	73,812	13,872	0,461	0,049	0,000	0,022	0,096	0,027	0,000
birth_year	0,993	1,471	0,572	17,173	11,671	10,291	0,000	27,039	25,105	4,551	0,038	0,457	0,511	0,124	0,005
Startdate_year	0,005	0,022	0,471	53,801	0,049	0,028	6,082	0,533	16,955	12,213	0,138	8,645	0,420	0,001	0,636
ContractTerm	13,376	20,235	0,067	1,270	0,137	0,576	0,016	0,807	0,119	0,271	0,691	20,105	3,387	0,167	38,775
premiumendate_year	13,168	15,102	0,511	3,899	0,173	0,008	0,298	0,000	1,253	1,203	1,048	52,641	7,475	0,557	2,663
contractenddate_year	15,283	19,922	0,043	1,715	0,082	0,251	0,066	0,371	0,000	0,149	0,179	3,691	0,390	0,026	57,831
Lapse_fees	1,465	0,681	3,012	2,885	36,146	17,251	0,000	1,199	30,981	5,729	0,002	0,259	0,366	0,024	0,000
Index_lapse_fees	0,260	0,453	17,813	0,288	30,675	20,850	3,127	2,314	0,002	4,632	0,621	0,789	2,639	15,536	0,002
pct_Commission_CashFund	9,518	8,804	0,280	3,197	1,076	11,505	0,578	0,044	0,734	28,525	33,710	0,028	1,620	0,379	0,004
pct_Commission	11,789	10,394	0,847	2,768	3,043	1,678	0,626	1,212	1,238	7,776	49,855	0,019	8,262	0,496	0,000
pct_Commission_CashFund_Broker	14,848	7,949	2,672	2,773	0,068	2,407	0,291	0,344	1,208	17,015	11,363	5,729	25,853	7,406	0,076
pct_Commission_Broker	15,249	10,349	0,683	0,715	1,179	1,552	1,558	1,484	0,223	15,559	0,777	3,535	41,400	5,737	0,000
pct_Discount_Fees	1,995	1,641	2,756	1,897	10,018	1,703	9,941	47,392	20,538	0,001	1,186	0,608	0,277	0,047	0,000
pct_Entry_Fees	0,563	1,285	45,339	0,110	0,603	0,357	0,079	0,054	0,033	0,000	0,388	1,546	3,870	45,769	0,005
pct_Entry_Fees_to_Broker	0,997	1,350	24,830	0,275	5,066	28,069	3,527	3,333	1,151	2,328	0,004	1,927	3,435	23,707	0,002

Figure 1: Contributions matrix of initial variables for each factor of the PCA

In O2 variable contributions don't matter because all factors are conserved with all the variables information. Nevertheless, in O1, only the 4 first factors are kept. It's essential to analyze contributive variables for each factor.

According to the contributions table, we can notice that variables which deal with start date of the contract, commissions, entry fees on premiums have an important cumulative contribution on the 4 first factors (especially the start date of the contract with a contribution of 54% approximately). On the other side, gender, lapse fees and discount fees do not contribute much (less than 18% in each case).

Concretely, it means in O1, among the remaining information, the discriminate nature of the start date of the contract, commissions and entry fees on premiums will be more dominant than the gender, the birth year, lapse fees and discount fees. Actually this fact allows us to encourage more one variable than another during the aggregation process, by privileging the use of a factor for which the variable contributes more during the PCA. For example, some variables can be considered by an expert opinion, as more or less pertinent. In our case, it means individuals in O1 will be less discriminated through the sex and the age for instance.

c. Aggregations

When the PCA has been done, you don't dispose of a representation of individual given by "normal" variables, but you have a representation of individuals given by the factors of the PCA which have been conserved. This new matrix is used to realize aggregations.

Once the aggregations method is selected, you have to parametrize it. In the case of the HAC it consists by choosing what type of link you want. For the k-means you have to choose the initial centers, eventually the number of repeats of the aggregation (if you have chosen random initial centers), the number of iterations to converge even further the optimal intra-class inertia, and the final degree of clustering. Because of the significance of the parameters of aggregation methods, a complementary sensitivity study could be realized.

d. Completion of liability tables

After the aggregation, you obtain your different classes. To use these classes in the saving model, it's essential to have the corresponding characteristic. However, after the PCA and aggregations, model points are described with barycentric values of classes, following the PCA factor. These values are very different to initial variables values and they are unexploitable by the model. An efficient hypothesis is proposed: to choose the barycentric object characteristics of each cluster as characteristics. This hypothesis permits to complete quickly and effectively liabilities tables. Another benefit is to insert existing values in the model: it could be detrimental to not use the exact value of some variables (for technical reasons like model implementation reason, or regulation reason like fixed guaranteed rate).

We can imagine that this hypothesis will create a difference between aggregated table results and per head table results. Indeed, barycentric objects and barycentric points are not necessarily mixed up. A good measure in a first approach could be to define a cumulative distance between barycentric objects and barycentrics point to not exceed. Otherwise the best idea would remain to "convert" the final matrix in "true" values by using the initial Eigen values of the PCA, but it requires another step of calculations, that's why we won't' proceed like that.

e. Automation of calculations

This part of the process is developed especially because we make a study, otherwise it wouldn't be essential. At the end, we dispose of a large number of model points table to realize the study. The work would be too important if we change by hand each liability assumption to calculate capital requirement and other results for each aggregation case. Thus, we use ADDACTIS Workflow which allows us to automate the parametrization of cases. Finally, we will obtain results very easily, and they will be compared to each other.

4. Application and studies of aggregation criteria

4.1. Aggregations strategies panel

In this part we will describe the panel of possibilities that we have selected to highlight influence of aggregation criteria on capital charges. We will sum up just below the choice of aggregation strategies. Obviously, in a non-study case, only one strategy would have to be chosen.

The use of HAC at the beginning is outlawed because the dimension and more exactly the number of individuals exceeds the acceptation limit. Thus this method won't be used neither alone nor in a first part of a cumulative strategy. Nevertheless, as we said previously, we will use it after a k-means aggregation. Thus we propose the following list:

- Classical method only, by defining that discriminant variables are the age, the sex and the fiscal maturity;
- K-means only, by choosing random initial centers, 10 repeats of the global aggregation and 500 iteration to achieve the optimal result;
- K-means which creates 3000 model points, then a HAC with simple link;
- K-means which creates 3000 model points, then a HAC with complete link;
- K-means which creates 3000 model points, then a HAC with average link;
- K-means which creates 3000 model points, then a HAC with Ward criteria.

These different aggregation specific cases (with the exception of the classical method) are multiplying by the number of options we have retained during the PCA study, which means multiplying by two.

Concerning clustering levels, several levels are arbitrary considered to observe continuously evolution of indicators: 2, 10, 100, 500, 1000, 1500 and 2000 model points.

At the end, 77 possibilities, and therefore 77 different datasets will be created. These datasets will be inserted in turn in the saving model, to notice sensitivities of solvency indicators.

4.2. Remarks

For the cumulative strategies, the number of 3000 classes has been chosen to be closest to the dimensional acceptation limit noticed for a HAC as early as possible, and by reducing the loss of information. Besides, 10 repetitions of the aggregation have been configured to obtain several final partitions and to retain the most relevant, which is the partition where the internal inertia is the most minimal.

Finally, for the classical method, we have chosen to keep the birth year and the contract start date as discriminant variables. It gives us different numbers of model points in order to compare to classification methods.

4.3. Aggregation results

a. Aggregation time

Calculation time has been observed to know if there is a difference between methods' types, and if the choice of factors number for the PCA could have an impact on calculation time. Regarding the comparison of calculation time, the method which uses Ward criteria is slightly faster among HAC methods : this follows from the use of the reciprocal neighbor algorithm. The k-means method remains faster by avoiding the recalculation of the distance matrix.

Moreover, concerning the different clustering levels for the HAC in cumulative strategies, a decreasing effect of calculation time is visible when the class number is growing. It can be explained by the fact that the more class there is, the more the upper part of the hierarchical tree is underdeveloped, and so the more the number of iteration is reduced.

Regarding the difference between O1 and O2, calculation times are clearly similar. Thus, the diminution of the number of factors for the PCA hasn't an influence on calculation time.

Compared to the classical method which is not expensive in terms of computing time but expensive in manipulation time, aggregation times for classification methods which are at least three minutes seem to be fast enough.

Finally, given the fact that the order of magnitude of aggregation times is the minute, we can neglect it compared to consistent calculation time of a classical per-head model, and so aggregation time doesn't seem to be a relevant indicator for choosing methods.

b. Macro study of result quality

About the aggregation quality, the calculation of the R^2 gives these results (results table are available in the appendix, appendix 3):

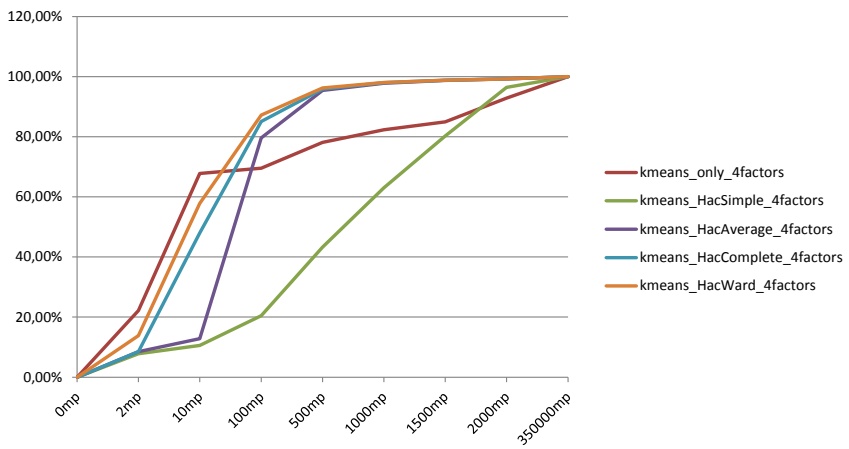

Figure 2: R^2 depending on the aggregation level for each strategy used in the case O1

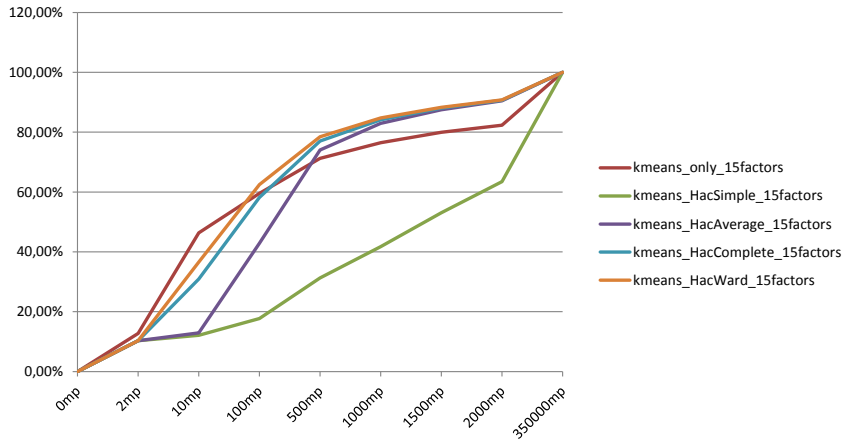

Figure 3: R^2 depending on the aggregation level for each strategy used in the case O2

In the case O1 and O2, this hierarchy is visible: strategies using the Ward criterion are better fitted than those which use the complete link which are better fitted than those which use the average link which are better fitted than those which use the simple link (as matter of fact, strategies using simple link are clearly distinguished by a poor quality). About the use of k-means' method only, quality according to R^2 is evolving, better than other families for a little number of clusters and worst for an important number of classes.

Concerning the segmentation level, it is reached when the growth of the R^2 is no longer noteworthy. It corresponds approximately to the aggregation of 1000 model points in O1 and 2000 model points for O2, for the strategies using Ward criterion, the complete link and the average link. For the other methods, optimal number of model points isn't easy to define. For the strategy using the simple link it could be in O1 2000 model points. For the strategies using only k-means, the notion "R^2 is no longer noteworthy" is hard to materialize, so we advise to not retain optimal results in this case too.

About the comparison between O1 and O2, the representative curves of R^2 for O1 seems to be more concave than the representative curves of R^2 for O2. It means for O1 the R^2 converges faster around 100% than for O2. In other words, when the truncation level grows, the intra-class inertia grows faster in O1 than in O2. So, the aggregation is "better" in the case O1. This conclusion, once more, is to compare with the information quality contained in O1 (case with 57% of the observation cloud dispersion) compared to the case O2, and to compare to the fact that, in O1, there are fewer ways to discriminate individuals than in O2 (because of the use of 4 and not 15 factors), so the discrimination is lower. Thus, optimal aggregation results are different according to the methods and PCA choices too, and sometimes these optimal solutions can't be easily reached by using R^2.

c. Micro study of result quality

Indicators like R^2 are a first approach to note the quality of aggregations, but the best study remains to observe evolution of individual characteristics. So, in order to better understand the potential impact of aggregations on solvency indicators, representations of the weighted average characteristics convergence of aggregated portfolio for different clustering level have been made. These convergences are

realized in percentage of variable value for the per head table, in this way, the study of the per-head dataset, realized in the beginning, is used as a reference. Some figures are explained just below. About the contract start date:

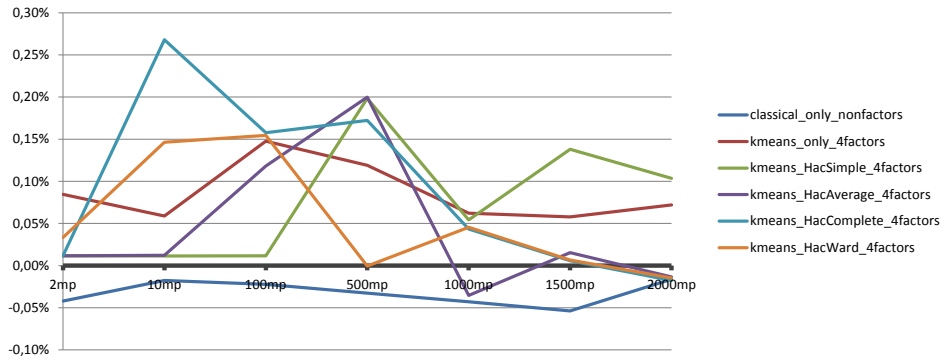

Figure 4: Difference of the contract start date for each aggregative strategy, depending on the aggregation level, in the case O1

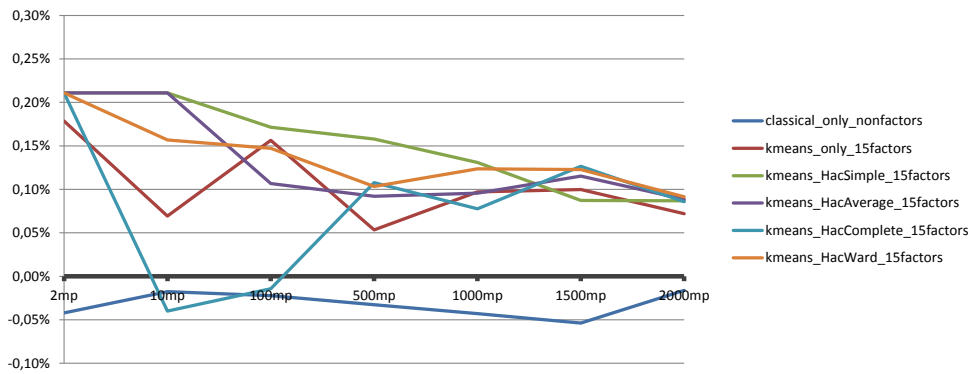

Figure 5: Difference of the contract start date for each aggregative strategy, depending on the aggregation level, in the case O2

About birth year:

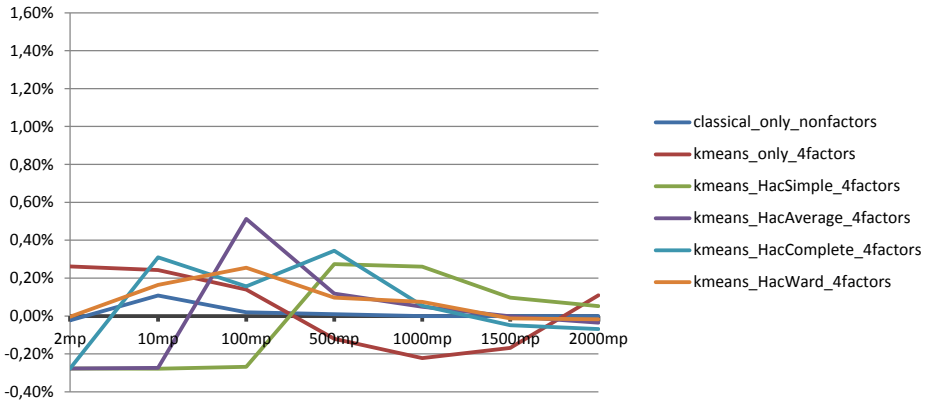

Figure 6: Difference of birth date for each aggregative strategy, depending on the aggregation level, in the case O1

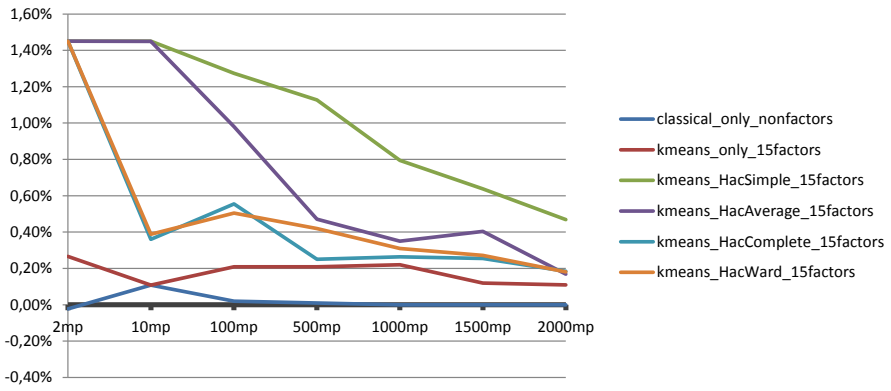

Figure 7: Difference of birth date for each aggregative strategy, depending on the aggregation level, in the case O2

The first remark is that deltas between characteristics obtained with aggregation and those of the per-head table are very low which is a good sign of aggregation quality. Then, we notice instability when you have few model points. This point can be explained by the blatant inequality in the number of observations by class for these levels. The study of the standard deviation of number of individual by class for each aggregation strategy shows big standard deviations for these low number of model points. This phenomenon is amplified when the variable has few possible values and an appreciable proportion of standard deviation because of stronger choices.

In addition, it's possible to see a hierarchy in the speed of characteristic convergence compared to the per head table characteristic (even if all is relative given the fact deltas are very low): strategies which use the complete link, the average link and the Ward criteria converge more quickly than those which use the simple link. Classical method sounds to have great performance concerning birth year and to a lesser extent contract start date, but it seems quite obvious because they were the exact discriminant variables of the method.

The percentage of dispersion conserved or not for the PCA has an impact on characteristics of clusters. It seems that when dispersion isn't taken into account completely (O1), characteristics are more unstable on average, for the different aggregation levels. To the opposite, when the dispersion is taken in its entirety (O2), the convergence is clear without value jumping. Despite the volatility for O1, the fact that the first four factors of the PCA have a strong contribution of the contract start date, generates aggregations principally based on this discriminant variable and the consequence is that O1 converges faster than O2 (here it's a good example of how, in case of expert opinions, giving priority to a variable).

Finally, the exclusive use of a k-means method gives results as good as cumulative strategies. However, cumulative strategies have been made to obtain better results than simple use of classification method: the transition between the first method and the second one for the cumulative strategies has probably affected quality results. So this transition must be treated very carefully. This phenomenon could be explained by the use of barycentric object instead of using barycentric point of aggregated classes too.

4.4. Analysis of the model results

a. Calculation time

The evolution of calculation time (values are available in appendix, appendix 4) depends on the number of model points which exists in the liability table. Initially, the per-head model runs in 54 minutes and 16 seconds. Only one table line is represented in the appendix. Indeed, the difference between O1 and O2, and for each strategy, has no effect on calculation time: only the number of lines in the model points table prevails. It's important to not forget that aggregation must be used to make the model more reactive. As expected, the calculation time increases when the number of model points grows. We notice a plateau between 2 and 100 model points (probably causes by a calculation cost not inherent to the model point's number). Then, relative gain of time decreases in a quasi-linear way.

b. BE

Results of differences between central Best Estimate[9] of aggregated tables and central Best Estimate of the per-head table are presented just below:

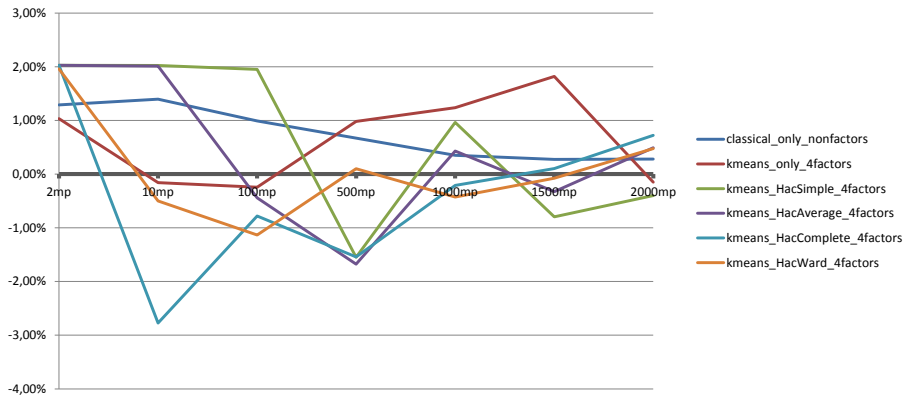

Figure 8: Evolution of the central Best Estimate depending on each aggregation strategy – Case O1

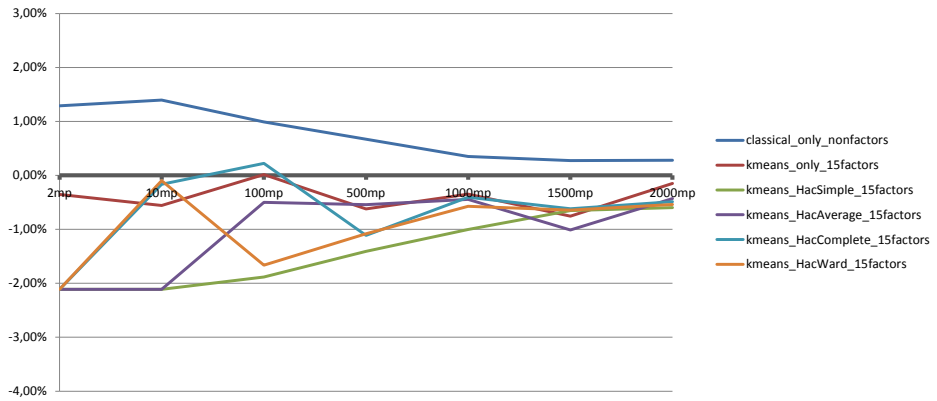

Figure 9: Evolution of the central Best Estimate depending on each aggregation strategy – Case O2

[9] Results of BE and other solvency indicators are available in appendix, appendix 5 to 10.

The first observation is that the Best Estimates of aggregated tables seem very close to Best Estimates of per-head tables. Indeed, the difference in terms of absolute value doesn't exceed 3% of the central Best Estimate value of the per-head table, after 1000 model points the delta is close to 0.5%.

Some information must be highlighted in the distinction of strategies:

- O2 is globally more close to the per-heard value of Best Estimate than O1;
- The volatility of O1 results is still visible;
- The same hierarchy between result quality of different cumulative strategies can be noticed : strategies which use the complete link, the average link and the Ward criteria converge more quickly than those which use the simple link;

Moreover the sign of the Best Estimate difference can be observed too. In O1, the different is positive then it becomes negative or close to zero. In O2, the different remains negative. That means in O1, future liabilities are overestimated in a first time, then underestimate whereas in O2, future liabilities are always underestimated. These facts are not easy to well understand. It comes from the fact, in the model after 8 years of contract maturity, no more fees are charged on redemption, whereas expenses remain. That means old contracts aren't benefit for the insurer. All the same, as presented previously, in our aggregated cases (except the classical method) the contract start date is increasing and so contracts are younger that per-head case. As the global aggregated portfolio is more benefit for the insurer the Best Estimate decreases.

c. Δ NAV

In this part we will present some Δ NAV results to study aggregation effects in a more global comprehensive. We will focus on life Δ NAV given the fact aggregations have been made on liabilities and so potential difference of capital charge could be observed for life risks.

Mortality shock

Charts just below represent the evolution of Δ NAV difference of mortality shock for level aggregation and for each strategy used in O1 and O2 (in percentage of the Δ NAV for the per-head table):

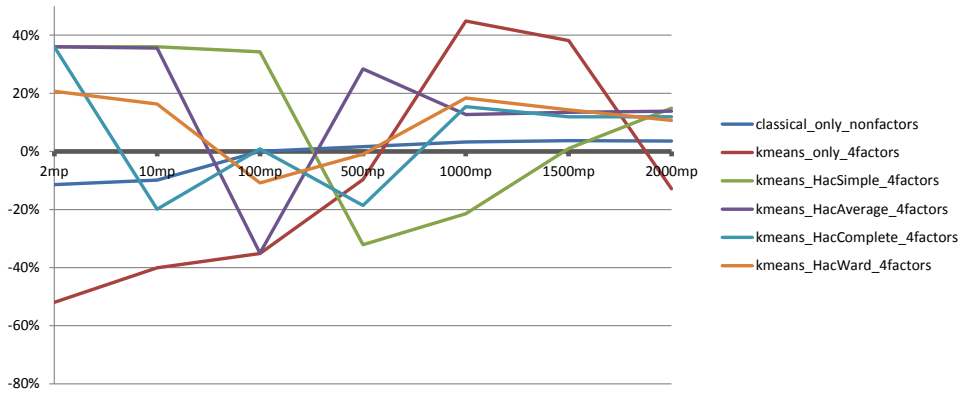

Figure 10: Delta NAV difference for the mortality shock – Case O1

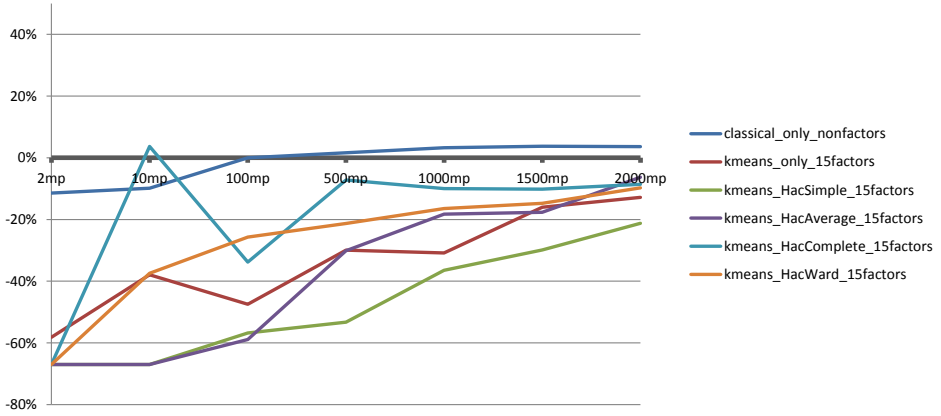

Figure 11: Delta NAV difference for the mortality shock – Case O2

The convergence is visible but not perfect for this shock: at the end we obtain for classification strategies a difference around 10%. Classical method is better because of exact value of age due to the choice as discriminant value. We note hierarchy of methods quality is conserved too. Finally, volatility in O1 is more visible than in O2 (as remind: birth date doesn't contribute much in the first four factors of the PCA).

To give more details about curves' monotonies, let reminds that mortality risk corresponds to the uncertainty linked to the insured death. The insurer has to do a shock to measure the loss risk or the deterioration risk of liabilities value of the insurer caused by an increase of mortality rate. This mortality rate comes from generational mortality table and depends on two factors: the insured age and his sex. To understand delta NAV variations, we must observe differences on the average sex and the average age previously calculated for each strategy. Let focuses on the birth date curve presented at the beginning of this chapter. In O1, strategies using HAC with simple link and with average link have average birth date older than the per-head case. As people older it sounds natural the mortality risk grows too, that's why in these scenarios BE is bigger than per-head BE. For the other methods, insured are younger so mortality risk is less important. In O2, birth year is always bigger than per-head case. Thus, as insured are younger, mortality risk in aggregated scenarios is reduced and so delta NAV difference is negative for little number of model point, before converging.

To conclude, delta NAV is negatively correlated to the value of average birth year, and suffers through this link of deterioration when model points are too few. All the same, classification methods show interested capacity of convergence (from 1000 model points).

Lapse shocks

The capital charge for the lapse risk is calculated by observing the maximal variation of NAV after the up and down shocks application on the lapse rate. In the study, the lapse up shock strongly impacts the BSCR calculation.

Charts just below represent the difference between delta NAV of the lapse up shock for aggregated tables (for different aggregation levels and for each strategy) and

delta NAV of the lapse up shock for the per-head table (in percentage of the delta NAV for the per-head table):

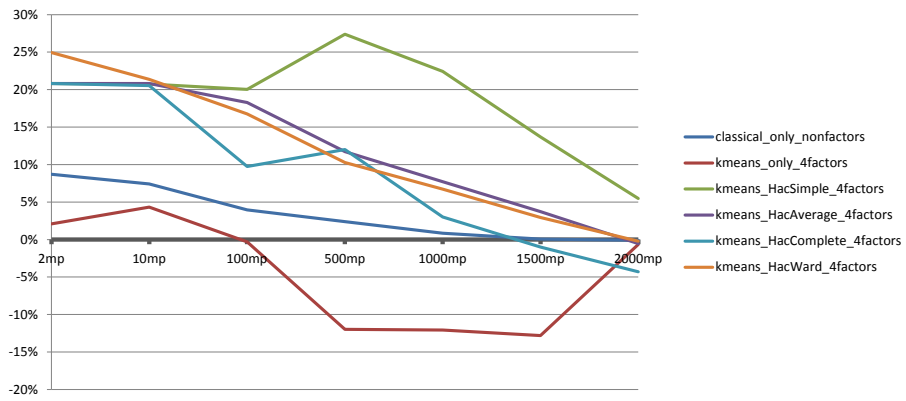

Figure 12: Delta NAV difference for the lapse up shock – Case O1

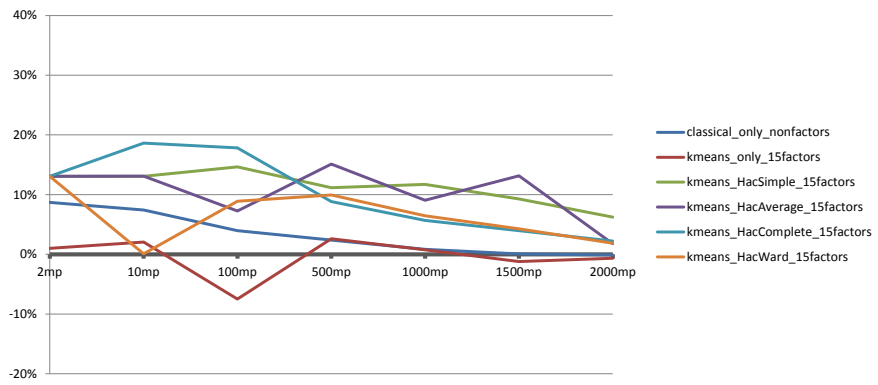

Figure 13: Delta NAV difference for the lapse up shock – Case O2

Delta NAV of the lapse up shock is one of the main posts for the life undertaking risk. We observe small deviations of delta NAV in both PCA options and a convergence is well visible in 2000 model points at least for cumulative strategies. Again, strategies using Ward criteria, average link and complete link have good results compared to strategies using simple link which don't converge at 2000 model points. Classical method has interested results with a fast convergence between 1000 and 1500 model points, but it remains logical because the second discriminant variable of the method is the contract start date, variable linked to lapse rate calculations.

Otherwise, to explain variations and speed of aggregation convergences, it is essential to look at the origin of the risk. The lapse risk is caused by uncertainty on policy lapse rate, abrogation, or cessation of premium payments. The lapse up shock corresponds to an increase of 50% of the lapse rate. In the aggregated table scenarios, we have seen contracts were younger, and so insurer gain was higher than the per-head example because lapse fees activation allowed him to maintain his benefits. That's why an important redemption is more impacting in term of delta NAV in the aggregated case: the insurer loses higher margins contracts than in the per head scenario. A last remark: contrary to the birth date case, O1 is quite fast to converge compared to O2 too because contract start date contribute a lot in the first four factors of the PCA.

Expense shock

Charts below represent the evolution of the delta NAV difference for the expense risk for different truncation levels and for each strategy employed in O1 and O2 (in percentage of the delta NAV for the per-head table):

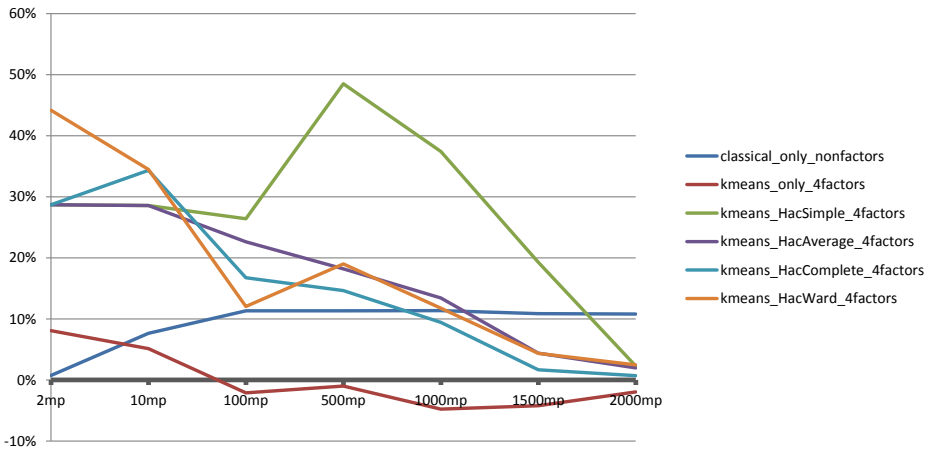

Figure 14: Delta NAV difference for the expense shock – Case O1

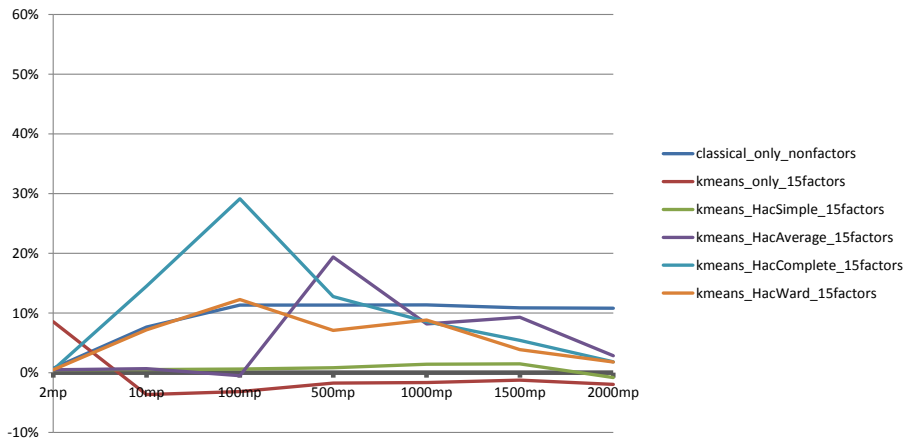

Figure 15: Delta NAV difference for the expense shock – Case O2

The delta NAV for the expense shock is the second risk for the life undertaking risk in our study. In O1, differences are quite important (up to 50% for the delta NAV of the per-head table). Despite this fact, a convergence is visible. In O2, the trend is more precise again. In addition, there isn't a significant evolution for low model points number. The usual hierarchy between methods is visible. Moreover, we observe this time, that classical method is clearly worse than classification methods: it comes from the lack of taking into account of fees in discriminant variables: here is the limit of the classical method.

Otherwise, the expense risk is due to an increase of the management fees of insurance policies. Unlike the previous risk, criteria which have to be taken into account for the analysis of the expense shock are very numerous: management fees,

switches fees…It appears that postulates are difficult to deduct concerning the influence of characteristics. Despite this, we can say that the decrease of contract fiscal maturity and the increase of birth date in aggregated cases makes that contracts stay longer in the portfolio than those of per-head case. However, in the shocked scenario each contract cost much more because of the increase of expenses, thus insurer will have to support this loss during longer period which makes grown the delta NAV.

Other shocks

Delta NAV difference of others shocks are not showed in this study for several reasons. First, some shocks have negative delta NAV and so they aren't involved in the calculation of the capital requirement (longevity shock for instance). Then, some shocks have a positive delta NAV but they aren't significant, and they aren't interesting for the study of aggregation quality: shocks as the equity shock has an unevolved delta NAV between calculation with per-head table and calculation with aggregated tables. We see the same type of result for other sub-posts of the market risk like property shock or spread shock.

d. SCR Market, SCR Life, SCR Default and BSCR

SCR Market

Given the fact, few remarks would have been made because most of shocks aren't impacted by liability aggregations, we have chosen to not study it the SCR Market (in addition it's the second post of the BSCR).

SCR Life

By definition, the study of the SCR Life is a simple remind of the previous results: delta NAV of the lapse shock represents more than 85% of the cumulated delta NAV for this risk, so monotony of SCR life and delta NAV of lapse shock is strictly the same. We note a positive accentuation effect of deviation dues to delta NAV of expense shock.

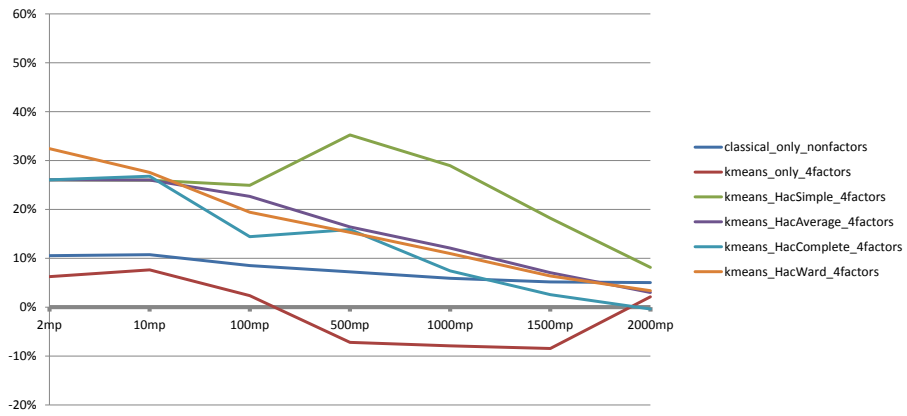

Figure 16: SCR Life differences – Case O1

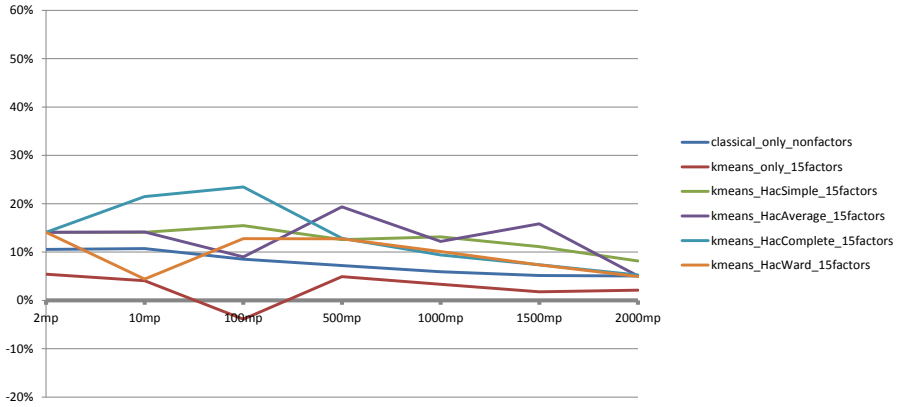

Figure 17: SCR Life differences – Case O2

BSCR

BSCR includes, in the case of the study, the life undertaking risk and the market risk. The difference between this indicator for aggregated tables and the per-head table represents a good way to measure the global influence (and possible compensation) of aggregations on regulatory requirements. Charts bellows represent the BSCR difference for each number of model points:

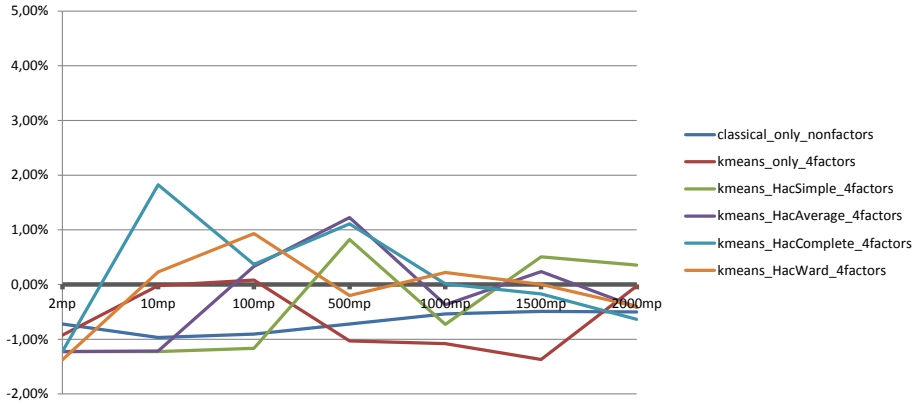

Figure 18: Study of the BSCR difference between per-head scenario and the multiple strategies depending on the clustering level – Case O1

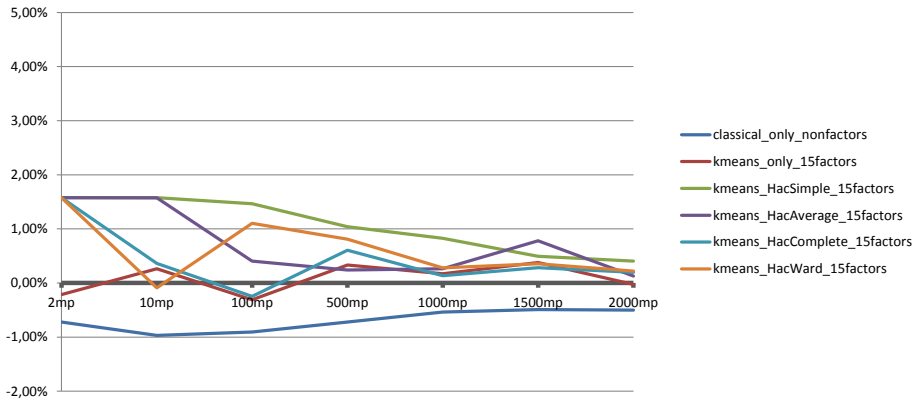

Figure 19: Study of the BSCR difference between per-head scenario and the multiple strategies depending on the clustering level – Case O2

Convergence is quite good given the fact we stay between -1% and 2% of difference compared to the per-head scenario. Regarding O1 and O2 comparison, we note again the volatility of O1 compared to O2 which is closer to the real per-head portfolio.

Otherwise, classical method appears less robust than it would seem at the beginning of the result study: actually results were good uniquely for mortality risk (which is not the aim post). And cumulative strategies converge quite well with always the same order of convergence quality.

In addition, monotonies don't correspond exactly to SCR life monotonies as we could expect. Actually the peaks on the O1 curves and in a more mitigated way for O2 are due to SCR Market evolution. At the end, we conclude that BSCR for aggregated tables is very close to BSCR of the per-head scenario.

Lastly we remark this little offset between -1% and 2% which exist for classification method convergences can be a consequence of approximated transition between the first method and the second method for the cumulative; it could be explained by the use of barycentric object instead of using barycentric point of aggregated classes too.

5. Limits

It's important to insist on the fact that some information and conclusion on this study must be put into perspective, especially concerning the choice of the dataset, modeling and methodology.

5.1. About data

The choice of dataset has a strong impact on aggregation results. In this way, the hierarchy realized on aggregative strategy, calculation time and differences are verified in the case of the study and some nuances can be said otherwise.

Regarding the line number, which is the number of people, the study deals with a dataset of 35 000 individuals. This number depends on the treated portfolio. It has been shown that this number was essential for the logical of the study. Indeed, methods and strategies aren't the same and depend on this number. For instance, in the case of a small portfolio, the study can be done with classical strategies only and thus it can be very brief. To the opposite if the portfolio is bigger, complex aggregative strategies have to be used, and the proposal process will be well fitted.

About the number of columns, 29 variables are available in the liability dataset which has been used. To model, insurer can use more or less variables. In the scenario he uses less variables, classical method remains interesting because of the limiting number of discriminant variables to choose. But in a scenario he uses more, classification can be very helpful but has some limits too. Indeed, the more variables are numerous the more it was complicated to link modeling effects with aggregation effect

Only variables which have been used to do aggregation have been studied to explain the evolution of solvency indicators. But, the study of delta NAV has shown that it

was complicated to refer just to liability discriminatory variables. In addition it could be pertinent to study others information available in the model like shocked flows of Best Estimates, or the present value of the future profit (PVFP) for instance.

In the study, quantitative variables have been used. Nevertheless, it's feasible to have qualitative variables in the liability dataset. If this type of variables was used, the way to proceed would have been very different. Indeed, contingency tables would have been necessary and preliminary study could be more fastidious.

5.2. About the model

The choice of the model has a strong influence on aggregation results too. To be more specific, the perfection of the model has impacts on our choices for the study. First, some insurers work only with per-head data because its models are very successful and thus they don't have trouble concerning optimization (like time saving as it exposed in the introduction issues of the paper). To the contrary, others insurers have simplified models and they have to do aggregation to respect both quality and time issues.

The second point deals with the fact that generally complex models need bigger inputs. But bigger dataset increases the complexity of the process aggregation and its setting up as we saw previously. In this way the more the model is sophisticated the more it will be difficult to do aggregations. To make one last point about models: to explain as better as possible convergence levels, model needs to produce as far outputs as possible.

5.3. About methodologies

It is important to give some additional information about methods used, classification strategies, options for the PCA, assumptions… Indeed, alternatives way could be considered. The use of some methods is the result of dataset technical necessities (use of k-mean aggregation for example) or model technical necessities. Other methods are arbitrary choices: it is the case for aggregations families used. In the study, only 2 families have been used. This choice has limited possibilities of strategies combinations, and maybe, constraints about aggregations strategies could have been fewer.

About PCA, factors' choices have shown that it was possible to "encourage" one variable rather another one by using a factor for which the variable contributes as a deciding way. Thus, actuarial hypothesis could be used (for instance some variables are considered less important than others) to guide the study. These hypotheses must be used carefully because we have seen that results volatility for important aggregation was increasing by making choices of this type.

Regarding R^2, we have seen this indicator was not always adapted. In this context it should be clarified quality aggregation must be measured, either by another indicator like CCC or by a strong complementary characteristics study as it has been done in the paper.

It could be pertinent to measure sensitivities of solvency capital requirements facing with hypothesis which have been done: the choice of barycentric objects as reference points for aggregated classes, the first loss of information linked to k-means aggregation… Indeed, influence of these hypotheses on Solvency 2 indicators is visible. As part of a further study, it could be interesting to observe the consequence the loss of one basic point of the inter-class inertia during the first aggregation, on the final difference between indicators of the per-head table and indicators of the aggregated tables.

Moreover, the choice concerning the use of barycentric objects as reference point of aggregated classes must be quantified. It could be interesting to study the distance between barycentric objects of aggregated classes and barycentrics points of these classes. We could have a great quantification of the approximation degree.

Finally, truncation levels aren't numerous: 2, 10, 100, 500, 1000, 1500 and 2000. To describe results in the best conditions, it could be useful to work with more truncation levels. This proposition has to be studied with attention because, increasing truncation level means increasing number of study, number of cases and so, process and calculation time.

6. Conclusion

To conclude we have seen classification was a thorough solution to reduce calculation time while improving model results. First, the theoretical remind concerning classification allowed us to make comparisons on different clustering methods and classical method and to conclude on the interest of using cumulative strategies to combine benefits of classification families. In addition, to study aggregations results, it sounded essential to define qualities indicators like R^2, but also to observe dataset characteristics' evolution.

Then, in order to setting up the study, we have presented a saving model which simulated risk with an annual frequency and a projection on 40 years. The liability portfolio, reference point of the study has been presented too: a table of 35 000 individuals, described though 29 variables. A proposal process to create aggregated table has been suggested too. It had consisted to standardize the initial table and to use PCA before realizing aggregation, to avoid correlation, scale effect, eventually to reduce dimensions: two options have been retained to observe influence of dispersion conservation. Finally, we have obtained 77 different aggregated tables to insert in the saving model.

Before observing modeling results, we have taken a look on aggregations results. First we have noted that aggregation time could be neglect because of its low importance (about the minute). Concerning quality results an optimal aggregation has been remarked for 1000 model points in O1 and 2000 model points for O2 using R^2, but we couldn't get this information for each classification methods. Thus, the other solution was to observe directly a part of portfolio average characteristics:

- A hierarchy in term of quality result is visible among cumulative strategies : strategies which use the complete link, the average link and the Ward criteria converge more quickly than those which use the simple link;

- Instability for few number of model points can be explained by standard deviation of number of individual very important at these aggregation levels;
- The percentage of dispersion conserved or not for the PCA has an impact on characteristics of clusters: convergence in O2 is less precise but more stable whereas in O1 there is a good convergence of studied characteristics too but values are volatile. In addition the use of reduce number of factors allows us to encourage more a variable than another during the aggregation process.

These conclusions have been confirmed in the study of modeling results. We have observed (apart from the fact there is an average gain of 94% on calculation time) Best Estimates were directly underestimated because of characteristics deformation. Despite this, results were good given the fact we obtained as expected after the aggregative study, a convergence close to 0.5% of the per-head case Best Estimate. About classical method results, as supposed, they are often worse than classifications methods. To go ahead we have studied delta NAV shocks for life risk. They show more important deviation than the previous indicators. In some case, correlations with characteristic alteration are obvious (mortality risk with birth date), other time these correlations are more subtle (influence of fiscal maturity on the expense which impacts capital charge of lapse risk), and sometimes effects are very complicated to understand. Finally the analysis of SCR Life and BSCR demonstrate global solvency indicators for aggregated table are closed to the indicator value for the per-head table. It means that aggregations are very satisfying in a global point of view, thus it remains important to not begin by these markers, in order to not be duped by quality aggregation.

7. Bibliography

BACCINI A. (2010), « Statistique Descriptive Multidimensionnelles ».

BENZECRI J.P (1964), « Analyse factorielle des proximités », Institut de Statistique de l'Université de Paris.

BIERNACKI C. (2009), « Pourquoi les modèles de mélange pour la classification », CNRS & Université de Lille 1.

BOUBOU M. (2006), « Contribution aux méthodes de classification non supervisée via des approches pré topologiques et d'agrégations d'opinions » thesis, Université Claude Bernard Lyon 1.

CHEVALIER F. and LE BELLAC J. (2013), « La classification », Université de Rennes 1.

COULOUMY A. (2013), « Critères d'agrégations de jeu de données passif pour les calculs sous la Formule Standard de Solvabilité 2», actuarial master's thesis, ISFA.

GONZALEZ P.L (2008), « Méthodes de classification », Cnam.

GOVAERT G. (2008), « Modèle de mélange et classification ».

Mc LACHLAN G. J. and KRISHMAN K. (1997), « The EM Algorithm Wiley ».

ROUX M. (1985), A*lgorithme de classification*, Edition Masson.

SARLE W. S. (1983), « Cubic Clustering Criterion ».

8. Appendix

	Time calculation	Quality	Dimensional acceptance	Definition of the targeted class number
Classical method	+	+ +	+ + +	Impossible
HAC with simple link	+	+	+	Non supervised
HAC with complete link	+	+ + +	+	Non supervised
HAC with average link	+	+ + +	+	Non supervised
HAC with Ward criteria	+	+ + +	+	Non supervised
K-means	+ + +	+ +	+ + +	Supervised
Cumulative strategy : k-means then HAC	+ + +	+ + +	+ + +	Mix

Appendix 1: Sum up table of classification methods comparison

	F1	F2	F3	F4	F5	F6	F7
Eigen values	3,002	2,449	1,886	1,283	1,052	0,979	0,971
Variability (%)	20,012	16,328	12,576	8,553	7,015	6,526	6,473
% cumulated	20,012	36,339	48,916	57,468	64,483	71,009	77,482

	F8	F9	F10	F11	F12	F13	F14	F15
Eigen values	0,915	0,800	0,628	0,395	0,249	0,218	0,152	0,020
Variability (%)	6,103	5,333	4,190	2,633	1,659	1,453	1,016	0,132
% cumulated	83,584	88,917	93,107	95,740	97,399	98,852	99,868	100,000

Appendix 2: Eigen values obtained after the PCA

	0mp	2mp	10mp	100mp	500mp	1000mp	1500mp	2000mp	350000mp
kmeans_only_4factors	0	22,1%	67,7%	69,5%	78,1%	82,3%	85,0%	92,9%	100,0%
kmeans_HacSimple_4factors	0	7,8%	10,5%	20,4%	43,3%	63,0%	80,2%	96,4%	100,0%
kmeans_HacAverage_4factors	0	8,5%	12,9%	79,6%	95,4%	97,8%	98,7%	99,2%	100,0%
kmeans_HacComplete_4factors	0	8,5%	48,0%	85,1%	95,8%	98,0%	98,8%	99,2%	100,0%
kmeans_HacWard_4factors	0	13,8%	57,8%	87,2%	96,2%	98,1%	98,8%	99,2%	100,0%
kmeans_only_15factors	0	12,7%	46,4%	59,6%	71,3%	76,5%	79,9%	82,3%	100,0%
kmeans_HacSimple_15factors	0	10,4%	12,1%	17,7%	31,3%	41,8%	53,1%	63,5%	100,0%
kmeans_HacAverage_15factors	0	10,4%	12,9%	42,9%	74,1%	82,9%	87,5%	90,5%	100,0%
kmeans_HacComplete_15factors	0	10,4%	30,9%	58,1%	77,0%	84,2%	88,0%	90,7%	100,0%
kmeans_HacWard_15factors	0	10,4%	36,6%	62,5%	78,5%	84,8%	88,3%	90,8%	100,0%

Appendix 3: Value of the R^2 of each clustering strategy and each aggregation levels

	2mp	10mp	100mp	500mp	1000mp	1500mp	2000mp	350000mp
Calculation time in seconds	112	122	125	152	183	212	244	3256
Relative gain of calculation time	96,6%	96,3%	96,2%	95,3%	94,4%	93,5%	92,5%	0,0%

Appendix 4: Calculation time of the model for each model points number

BE central Relative value	2mp	10mp	100mp	500mp	1000mp	1500mp	2000mp
classical_only_nonfactors	1,29%	1,40%	0,99%	0,67%	0,35%	0,27%	0,28%
kmeans_only_4factors	1,03%	-0,16%	-0,24%	0,98%	1,24%	1,82%	-0,15%
kmeans_HacSimple_4factors	2,03%	2,02%	1,95%	-1,55%	0,96%	-0,80%	-0,40%
kmeans_HacAverage_4factors	2,03%	2,01%	-0,44%	-1,67%	0,43%	-0,33%	0,49%
kmeans_HacComplete_4factors	2,03%	-2,77%	-0,78%	-1,54%	-0,21%	0,10%	0,72%
kmeans_HacWard_4factors	1,96%	-0,50%	-1,13%	0,10%	-0,43%	-0,08%	0,48%
classical_only_nonfactors	1,29%	1,40%	0,99%	0,67%	0,35%	0,27%	0,28%
kmeans_only_15factors	-0,35%	-0,56%	0,02%	-0,62%	-0,35%	-0,76%	-0,15%
kmeans_HacSimple_15factors	-2,11%	-2,11%	-1,88%	-1,41%	-1,00%	-0,65%	-0,60%
kmeans_HacAverage_15factors	-2,11%	-2,11%	-0,50%	-0,54%	-0,45%	-1,01%	-0,43%
kmeans_HacComplete_15factors	-2,11%	-0,16%	0,22%	-1,11%	-0,41%	-0,62%	-0,49%
kmeans_HacWard_15factors	-2,11%	-0,10%	-1,66%	-1,08%	-0,57%	-0,64%	-0,54%

Appendix 5: Best estimate relative values for each classification methods and aggregation levels

SCR Mort relative value	2mp	10mp	100mp	500mp	1000mp	1500mp	2000mp
classical_only_nonfactors	-11,44%	-9,86%	-0,02%	1,62%	3,25%	3,72%	3,59%
kmeans_only_4factors	-51,98%	-40,04%	-35,15%	-9,71%	44,86%	38,11%	-12,84%
kmeans_HacSimple_4factors	36,04%	35,99%	34,29%	-32,11%	-21,48%	0,96%	14,91%
kmeans_HacAverage_4factors	36,03%	35,54%	-35,04%	28,39%	12,67%	13,47%	13,81%
kmeans_HacComplete_4factors	36,03%	-19,91%	0,85%	-18,63%	15,38%	11,87%	11,91%
kmeans_HacWard_4factors	20,66%	16,31%	-10,87%	-0,95%	18,34%	14,30%	10,70%
classical_only_nonfactors	-11,44%	-9,86%	-0,02%	1,62%	3,25%	3,72%	3,59%
kmeans_only_15factors	-58,17%	-37,93%	-47,46%	-29,92%	-30,82%	-16,06%	-12,84%
kmeans_HacSimple_15factors	-66,99%	-66,98%	-56,76%	-53,27%	-36,46%	-29,89%	-21,26%
kmeans_HacAverage_15factors	-66,99%	-66,97%	-58,87%	-30,12%	-18,27%	-17,65%	-6,30%
kmeans_HacComplete_15factors	-66,99%	3,67%	-33,79%	-7,22%	-10,00%	-10,15%	-8,61%
kmeans_HacWard_15factors	-66,99%	-37,45%	-25,70%	-21,29%	-16,46%	-14,77%	-9,76%

Appendix 6: SCR Mortality relative values for each classification methods and aggregation levels

SCR Exp relative value	2mp	10mp	100mp	500mp	1000mp	1500mp	2000mp
classical_only_nonfactors	0,75%	7,65%	11,33%	11,34%	11,35%	10,85%	10,81%
kmeans_only_4factors	8,07%	5,14%	-2,10%	-0,98%	-4,78%	-4,22%	-1,97%
kmeans_HacSimple_4factors	28,68%	28,58%	26,40%	48,50%	37,41%	19,28%	2,28%
kmeans_HacAverage_4factors	28,69%	28,58%	22,63%	18,20%	13,43%	4,38%	1,99%
kmeans_HacComplete_4factors	28,69%	34,32%	16,75%	14,63%	9,45%	1,67%	0,71%
kmeans_HacWard_4factors	44,15%	34,49%	12,05%	18,99%	11,75%	4,35%	2,46%
classical_only_nonfactors	0,75%	7,65%	11,33%	11,34%	11,35%	10,85%	10,81%
kmeans_only_15factors	8,54%	-3,67%	-3,17%	-1,73%	-1,66%	-1,24%	-1,97%
kmeans_HacSimple_15factors	0,51%	0,53%	0,60%	0,82%	1,43%	1,48%	-0,81%
kmeans_HacAverage_15factors	0,51%	0,67%	-0,49%	19,38%	8,16%	9,28%	2,83%
kmeans_HacComplete_15factors	0,51%	14,50%	29,11%	12,74%	8,55%	5,43%	1,83%
kmeans_HacWard_15factors	0,51%	7,16%	12,26%	7,08%	8,83%	3,87%	1,80%

Appendix 7: SCR Expense relative values for each classification methods and aggregation levels

SCR lapse relative value	2mp	10mp	100mp	500mp	1000mp	1500mp	2000mp
classical_only_nonfactors	8,70%	7,42%	3,95%	2,40%	0,84%	0,05%	-0,09%
kmeans_only_4factors	2,08%	4,34%	-0,30%	-11,98%	-12,07%	-12,81%	-0,65%
kmeans_HacSimple_4factors	20,82%	20,75%	20,03%	27,38%	22,42%	13,69%	5,47%
kmeans_HacAverage_4factors	20,82%	20,81%	18,28%	11,73%	7,71%	3,73%	-0,55%
kmeans_HacComplete_4factors	20,82%	20,53%	9,73%	12,00%	3,02%	-1,01%	-4,29%
kmeans_HacWard_4factors	24,93%	21,36%	16,75%	10,27%	6,72%	2,94%	-0,22%
classical_only_nonfactors	8,70%	7,42%	3,95%	2,40%	0,84%	0,05%	-0,09%
kmeans_only_15factors	1,01%	2,06%	-7,49%	2,61%	0,74%	-1,21%	-0,65%
kmeans_HacSimple_15factors	13,07%	13,07%	14,66%	11,17%	11,71%	9,28%	6,24%
kmeans_HacAverage_15factors	13,07%	13,10%	7,26%	15,11%	9,06%	13,16%	1,76%
kmeans_HacComplete_15factors	13,07%	18,64%	17,82%	8,82%	5,66%	3,96%	2,21%
kmeans_HacWard_15factors	13,07%	0,09%	8,88%	9,95%	6,46%	4,27%	1,87%

Appendix 8: SCR Lapse relative values for each classification methods and aggregation levels

SCR Life Relative value	2mp	10mp	100mp	500mp	1000mp	1500mp	2000mp
classical_only_nonfactors	10,53%	10,72%	8,52%	7,22%	5,92%	5,16%	5,03%
kmeans_only_4factors	6,25%	7,62%	2,36%	-7,20%	-7,91%	-8,44%	2,12%
kmeans_HacSimple_4factors	26,03%	25,96%	24,93%	35,22%	28,95%	18,22%	8,13%
kmeans_HacAverage_4factors	26,03%	26,01%	22,65%	16,40%	12,10%	7,06%	3,01%
kmeans_HacComplete_4factors	26,03%	26,78%	14,41%	15,89%	7,41%	2,56%	-0,39%
kmeans_HacWard_4factors	32,41%	27,56%	19,42%	15,28%	10,97%	6,38%	3,36%
classical_only_nonfactors	10,53%	10,72%	8,52%	7,22%	5,92%	5,16%	5,03%
kmeans_only_15factors	5,43%	4,06%	-3,90%	4,90%	3,33%	1,78%	2,12%
kmeans_HacSimple_15factors	14,09%	14,10%	15,47%	12,56%	13,15%	11,11%	8,13%
kmeans_HacAverage_15factors	14,09%	14,15%	9,00%	19,35%	12,18%	15,84%	5,05%
kmeans_HacComplete_15factors	14,09%	21,46%	23,47%	12,85%	9,40%	7,38%	5,23%
kmeans_HacWard_15factors	14,09%	4,41%	12,78%	12,72%	10,12%	7,35%	4,94%

Appendix 9: SCR Life relative values for each classification methods and each aggregation levels

BSCR Relative value	2mp	10mp	100mp	500mp	1000mp	1500mp	2000mp
classical_only_nonfactors	-0,72%	-0,97%	-0,90%	-0,72%	-0,54%	-0,49%	-0,50%
kmeans_only_4factors	-0,93%	-0,02%	0,08%	-1,03%	-1,08%	-1,37%	-0,03%
kmeans_HacSimple_4factors	-1,23%	-1,23%	-1,17%	0,82%	-0,73%	0,51%	0,36%
kmeans_HacAverage_4factors	-1,23%	-1,22%	0,33%	1,22%	-0,37%	0,24%	-0,42%
kmeans_HacComplete_4factors	-1,23%	1,82%	0,37%	1,11%	0,01%	-0,17%	-0,64%
kmeans_HacWard_4factors	-1,38%	0,23%	0,93%	-0,20%	0,22%	0,00%	-0,40%
classical_only_nonfactors	-0,72%	-0,97%	-0,90%	-0,72%	-0,54%	-0,49%	-0,50%
kmeans_only_15factors	-0,22%	0,26%	-0,31%	0,33%	0,17%	0,38%	-0,03%
kmeans_HacSimple_15factors	1,58%	1,58%	1,47%	1,04%	0,83%	0,49%	0,40%
kmeans_HacAverage_15factors	1,58%	1,57%	0,40%	0,24%	0,26%	0,78%	0,13%
kmeans_HacComplete_15factors	1,58%	0,36%	-0,25%	0,61%	0,13%	0,28%	0,20%
kmeans_HacWard_15factors	1,58%	-0,09%	1,10%	0,81%	0,28%	0,36%	0,22%

Appendix 10: BSCR relative values for each classification methods and each aggregation levels