CrossMark

Is it optimal to group policyholders by age, gender, and seniority for BEL computations based on model points?

Pierre-Olivier Goffard[1] · Xavier Guerrault[2]

Abstract   An aggregation method adapted to life insurance portfolios is presented. We aim at optimizing the computation time when using Monte Carlo simulations for best estimate liability calculation. The method is a two-step procedure. The first step consists in using statistical partitioning methods in order to gather insurance policies. The second step is the construction of a representative policy for each aforementioned groups. The efficiency of the aggregation method is illustrated on a real saving contracts portfolio within the frame of a cash flow projection model used for best estimate liabilities and solvency capital requirements computations. The procedure is already part of AXA France valuation process.

1 Introduction

One of the challenges of stochastic asset/liability modeling for large portfolios of participating contracts is the running time. The implementation of such models requires a cumbersome volume of computations within the framework of Monte Carlo simulations. Indeed, the model has to capture the strong interactions existing between asset and liability through the lapse behavior and the redistribution of the

✉ Pierre-Olivier Goffard
   pierre.olivier.goffard@gmail.com; pierreolivier.goffard@axa.fr

   Xavier Guerrault
   xavier.guerrault@axa.fr

[1] AXA France and Aix-Marseille University, Marseille, France

[2] AXA France, Marseille, France

⚠ Springer

financial and technical result. For an overview we refer to Planchet et al. [19]. Using a stochastic asset/liability model to analyze large blocks of business is often too time consuming to be practical. Practitionners make compromises to tackle this issue.

Grouping methods, which group policies into model cells and replace all policies in each groups with a representative policy, is the oldest form of modelling techniques. It has been advised in [7] and is commonly used in practice due to its availability in commercial actuarial softwares such as MG-ALFA, GGY-Axis and TRICAST suite. In a recent Ph.D. thesis [20], representative portfolios have been drawn from a stratified sampling procedure over the initial portfolio. Computations are achieved on this smaller portfolio and the procedure is repeated to yield a final approximation. The idea remains to reduce the computation time by reducing the number of model cells.

Another idea consists in reducing the computation time for one policy. For instance one may think about reducing the number of economic scenarios. In [6] an algorithm, inspired by importance sampling, is presented to select the most relevant financial scenarios in order to estimate quantiles of interest. Applications to the estimation of remote (but not extreme) quantiles can be found in [4], in addition to consistency results regarding the estimator. However, we are interested in the calculation of the value of Best Estimate Liabilities (BEL) for participating contracts, which are the probability weighted average of future cash flows taking into account the time value of money. Therefore, Planchet et al. [17] presented a way to aggregate all the scenarios in a set of characteristic trajectories associated to a probability of occurence, well adapted to BEL computations. Another well-known modeling technique when dealing with financial sensitivities is the replicating portfolio method. As life insurance contracts share many features with derivative contracts, the idea of representing liabilities by a portfolio of financial products comes naturally. For an overview of the replicating portfolio method, we refer to [5, 21]. We also want to mention the work of Planchet et al. [2], where an analytical approximation formula for best estimate valuation of saving contracts is derived. The approximation is based on the evaluation of a coefficient, which when applied to the mathematical reserve yields the BEL. We believe that the ways of dealing with computation time issues are not in competition, for instance a grouping method combined with an optimization of financial scenarios may yield great results.

In this paper, we present a new grouping method and we focus on the calculation of the value of BEL for participating contracts via an asset/liability model. This work has been motivated by the revision of the valuation process in AXA France to be compliant with Solvency II. The aggregation method must meet many practical requirements such as easy implementation and understanding, applicability to the whole life insurance portfolio and good empirical and theoretical justification. We believe that many insurance companies face computation time problems, and this paper may help practioners to solve it. In Sect. 2, we give a description of the cash flow projection model being used. In Sect. 3, we describe a statistical based way to group policies, followed by procedures to construct the representative policy.

Section 4 illustrates the efficiency of the method on a portfolio of saving contracts on the French market.

2 Cash flows projection models and best estimate liabilities assessment

We consider a classical asset-liability model to project the statutory balance sheet and compute mathematical reserves for saving contracts. The settings and notations are closed to those in [2]. Consider a saving contract with surrender value $SV(0)$ at time $t = 0$, the surrender value at time t is defined as

$$SV(t) = SV(0) \times exp\left(\int_0^t r_a(s)ds\right), \tag{2.1}$$

where r_a denotes the instantaneous accumulation rate (including any guaranteed rate) modeled by a stochastic process. We are interested in the present surrender value at time t, we need therefore to add a discount factor in order to define the Present Surrender Value

$$
\begin{aligned}
PSV(t) &= SV(t) \times exp\left(-\int_0^t r_\delta(s)ds\right) \\
&= SV(0) \times exp\left(\int_0^t (r_a(s) - r_\delta(s))ds\right),
\end{aligned}
\tag{2.2}
$$

where r_δ denotes the instantaneous discount rate, also modeled by a stochastic process. The spread between the accumulation and the discount rates in (2.2) is then a stochastic process. We assume that these stochastic processes are governed by a probability measure denoted by Q^f. We denote by $\mathbf{F}$ a financial scenario, drawn from Q^f, which corresponds to a trajectory of r_δ and r_a. Let $\tau|\mathbf{F}$ be the early surrender time, modeled by a random variable having probability density function $f_{\tau|\mathbf{F}}$ on the positive half line, absolutely continuous with respect to the Lebesgue measure. The probability density function can be interpreted as an instantaneous surrender rate between times t and $t + dt$, which depends on the characteristics of the policyholder like for instance his age, his gender, the seniority of his contract and the financial environment that may influence policyholders behavior. More specifically, in the case of a saving contract, the payment of the surrender value occurs in case of early withdrawal due to lapse, death, or expiration of the contract. In the cash flow projection model definition, an horizon of projection is usually specified. There are also life insurance contracts with a fixed term. Both of these times are deterministic. We denote by T the minimum of the expiration date of the contract and the horizon of projection. The real surrender time $\tau \wedge T|\mathbf{F} = min(\tau|\mathbf{F}, T)$ is a random variable associated with a probability measure divided into the sum of a singular part and a continuous part

$$dP_{\tau \wedge T|\mathbf{F}}(t) = f_{\tau|\mathbf{F}}(t)d\lambda(t) + \overline{F_{\tau|\mathbf{F}}}(T)\delta_T(t), \tag{2.3}$$

where λ is the Lebesgue measure on $[0, T]$, δ_T is the Dirac measure at T and $\overline{F_\tau | \mathbf{F}}(t)$ denotes the survival function associated to the random time $\tau | \mathbf{F}$. The BEL at time $t = 0$ associated to the financial scenario $\mathbf{F}$ is defined as

$$BEL^\mathbf{F}(0, T) = E^{P_{\tau \wedge T | \mathbf{F}}}(PSV(\tau \wedge T | \mathbf{F})). \tag{2.4}$$

We refer to [10, 19] for this definition of BEL. BEL, given a financial scenario, are written as

$$
\begin{aligned}
BEL^\mathbf{F}(0, T) &= E^{P_{\tau \wedge T | \mathbf{F}}}(PSV(\tau \wedge T | \mathbf{F})) \\
&= \int_0^{+\infty} SV(0) \times exp\left(\int_0^t (r_a(s) - r_\delta(s))ds\right) dP_{\tau \wedge T | \mathbf{F}}(t) \\
&= \int_0^T SV(0) \times exp\left(\int_0^t (r_a(s) - r_\delta(s))ds\right) f_{\tau | \mathbf{F}}(t)dt \\
&\quad + \overline{F_{\tau | \mathbf{F}}}(T) \times SV(0) \times exp\left(\int_0^T (r_a(s) - r_\delta(s))ds\right).
\end{aligned} \tag{2.5}
$$

In order to avoid tedious calculations of the integral in (2.5), time is often discretized. BEL are therefore written as

$$
\begin{aligned}
BEL^\mathbf{F}(0, T) \approx &\left[\sum_{t=0}^{T-1} p^\mathbf{F}(t, t+1) \prod_{k=0}^t \frac{1 + r_a(k, k+1)}{1 + r_\delta(k, k+1)}\right. \\
&\left. + p^\mathbf{F}(T) \prod_{k=0}^{T-1} \frac{1 + r_a(k, k+1)}{1 + r_\delta(k, k+1)}\right] SV(0),
\end{aligned} \tag{2.6}
$$

where $p^\mathbf{F}(t, t+1)$ is the probability that surrender occurs between time t and $t + 1$, and $r_a(t, t+1)$ and $r_\delta(t, t+1)$ are the accumulation and discount rates between time t and $t + 1$. Monte Carlo method for BEL evaluation consists in generating a set of financial scenarios under Q^f and compute the BEL for each one of them. The final estimation is the mean over the set of all scenarios. This procedure is fast enough for one policy, it becomes time consuming for a large portfolio. The formula given in (2.6) is commonly used by practionners and compliant with Solvency II. However, we know that there is room for discussion and improvement regarding the "economic" scenario generation, the assumptions on the liability side or the link between assets and liabilities. Nevertheless, the efficiency of our grouping methodology is not affected by these aspects hence we choose to stick to AXA France assumptions without discussing it further.

3 Presentation of the aggregation procedure

The goal of aggregation procedures is to reduce the size of the input portfolio of the cash flow projection model. The first step consists in creating groups of policies that share similar features. The second one is the definition of an "average" policy, called Model Point (MP), that represents each group and forms the aggregated

portfolio. The initial surrender value of the MP is the sum of the initial surrender values over the represented group. The method must be flexible so as to generate the best aggregated portfolio under the constraint of a given number of MPs. Let us consider two contracts having identical guarantees and characteristics (same age, same seniority, . . .). Therefore they have identical surrender probabilities. We build a contract having these exact characteristics and whose initial surrender value is the sum of the initial surrender values of the two aforementionned contracts. The BEL of this contract, given a financial scenario, is

$$
\begin{aligned}
BEL_{MP}^{\mathbf{F}}(0,T) &= \left[\sum_{t=0}^{T-1} p^{\mathbf{F}}(t,t+1) \prod_{k=0}^{t} \frac{1+r_a(k,k+1)}{1+r_\delta(k,k+1)} + p^{\mathbf{F}}(T) \prod_{k=0}^{T-1} \frac{1+r_a(k,k+1)}{1+r_\delta(k,k+1)}\right] \\
&\quad \times SV_{MP}(0) \\
&= \left[\sum_{t=0}^{T-1} p^{\mathbf{F}}(t,t+1) \prod_{k=0}^{t} \frac{1+r_a(k,k+1)}{1+r_\delta(k,k+1)} + p^{\mathbf{F}}(T) \prod_{k=0}^{T-1} \frac{1+r_a(k,k+1)}{1+r_\delta(k,k+1)}\right] \\
&\quad \times \sum_{i=1}^{2} SV_{C_i}(0) \\
&= \left[\sum_{t=0}^{T-1} p^{\mathbf{F}}(t,t+1) \prod_{k=0}^{t} \frac{1+r_a(k,k+1)}{1+r_\delta(k,k+1)}\right] SV_{C_i}(0) \qquad (3.1) \\
&\quad + \sum_{i=1}^{2} \left[p^{\mathbf{F}}(T) \prod_{k=0}^{T-1} \frac{1+r_a(k,k+1)}{1+r_\delta(k,k+1)}\right] SV_{C_i}(0) \\
&= \sum_{i=1}^{2} BEL_{C_i}^{\mathbf{F}}(0,T), \qquad (3.2)
\end{aligned}
$$

where $BEL_{C_i}^{\mathbf{F}}(0,T)$ and $SV_{C_i}(0)$ are the best estimate liability and initial surrender value of the contract $i \in \{1,2\}$. The idea behind the grouping strategy lies in this linearity property of the BEL. The aggregation of contracts having the same surrender probabilities leads to an exact evaluation of the BEL of the portfolio. The creation of an aggregated portfolio by grouping the policies having identical characteristics leads to a portfolio that is usually still too big to perform multiple valuations (with different sets of hypothesis). However one particular valuation might be doable using this aggregated portfolio in order to get a benchmark value for the BEL and assess the accuracy of the aggregation procedure in the validation phase. One may also note that the use of partitioning algorithms on the aggregated portfolio is faster because it is much smaller than the initial one. The AXA France portfolio of saving contracts contained different products associated to different guarantees. Obviously we cannot group together two contracts that belong to different lines of product. The variable PRODUCT divides the portfolio into sub-portfolios on which the aggregation procedure is used separately. We get aggregated sub-portfolios that are concatened to yield the final aggregated portfolio. In the first subsection, we explain how to gather contracts having similar surrender

probabilities and in the second subsection how to build a representative contract for each group.

3.1 The partitioning step

A portfolio is a set of contracts $\mathcal{P} = \{\mathbf{x}_i\}_{i \in 1,\ldots,n}$, where n is the size of the portfolio. We aim at partitioning n contracts into k sets $\mathcal{C} = \{C_1, \ldots, C_k\}$. We use clustering algorithms to identify sub-portfolios. A choice has to be made concerning the variables that characterize each observation and the metric to measure the dissimilarity between two observations. In order to get closer to the additivity of the BEL, every individual in the portfolio is represented by its sequence of surrender probabilities

$$\mathbf{x}_i = \begin{pmatrix} p_i^{\mathbf{F_1}}(0,1) & \cdots & p_i^{\mathbf{F_1}}(T) \\ \vdots & \ddots & \vdots \\ p_i^{\mathbf{F_N}}(0,1) & \cdots & p_i^{\mathbf{F_N}}(T) \end{pmatrix}, \tag{3.3}$$

where $\{\mathbf{F_1}, \ldots, \mathbf{F_N}\}$ denotes the set of N financial scenarios. Policies are therefore represented by matrices of size $N \times (T+1)$. The determination of $p^{\mathbf{F}}(t, t+1)$ necessitates the evaluation of a mortality and a lapse rate from one period to another using classical actuarial tools, see [18]. For saving contracts, lapse behavior is usually split between two components: a structural part linked to endogeneous factors and a dynamic part driven by the financial environment. the structural part is derived using historical data at product level. Figure 1 displays the lapse rate according to the seniorities of contracts in the portfolio we will consider later. Figure 1 clearly shows a step when reaching the 4th anniversary year and a peak at the 8th anniversary of the contract. The explanation lies in the french market specific tax rules that depend on the seniority.

Remark 1 The shape of the lapse rate curve in Fig. 1 may entail questions regarding the use of a continuous probability density function to model surrendering time in (2.3). This is another practical advantage of using a discretized expression to

Fig. 1 Lapse rate curve depending on the seniority of contracts in a saving contract portfolio
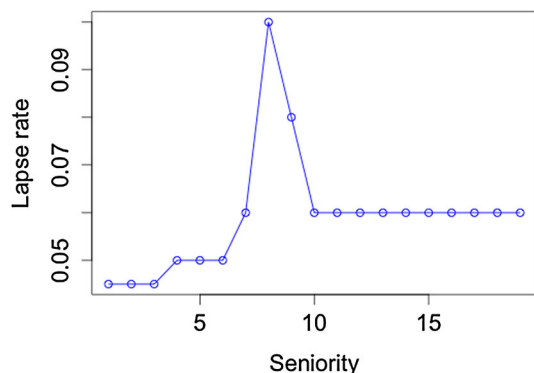
compute best estimates liabilities. However, there are in the applied probability literature interesting probability distributions on $\mathbb{R}^+$ that allow great flexibility. they are called phase-type distributions and might be appropriate to model surrender time, an overview is given in Chapter 8 of [1].

The dynamic part introduces more lapses when the client is not satisfied with the annual performance of his contract either because it is less than expected given past performances or because competitors offer more. Both reasons are governed by the financial situation and modeled accordingly. We notice an interesting fact that allows to reduce the operationnal burden that occurs when working with matrices as in (3.3). The financial scenario has a very limited impact on the proximity of contracts. Namely, two contracts close to each other given a financial scenario are also close with respect to another financial scenario. If there exist $k \in \{1, \ldots, N\}$ such that $\mathbf{x}_i^{\mathbf{F}_k} \approx \mathbf{x}_j^{\mathbf{F}_k}$ then we have approximately $\mathbf{x}_i^{\mathbf{F}_l} \approx \mathbf{x}_j^{\mathbf{F}_l}$ for all $l \in \{1, \ldots, N\}$. This is due to the fact that we consider saving contracts that belong to the same line of product. Note also that it consumes less ressources to store a vector for each policy than a matrix which is very important from an operational point of view. Thus we consider for each contract a vector of surrender probablities given one of the financial scenarios. Every policy in the portfolio is now represented by a vector of surrender probabilities

$$\mathbf{x}_i = (p_i(0, 1), p_i(1, 2), \ldots, p_i(T - 1, T), p_i(T)). \tag{3.4}$$

The variables are quantitative and fall between 0 and 1. The natural dissimilarity measure between two observations is the euclidean distance defined as

$$||\mathbf{x}_i - \mathbf{x}_j||_2 = \sqrt{\left(p_i(0, 1) - p_j(0, 1)\right)^2 + \cdots + \left(p_i(T) - p_j(T)\right)^2}. \tag{3.5}$$

Remark 2 It is worth noting that dealing with matrices instead of vectors is not a problem. Indeed, the techniques presented below still work by considering a dissimilarity measure between matrices.

Partitioning algorithms are designed to find the k-sets partition that minimizes the Within Cluster Sum of Square (WCSS), which characterises the homogeneity in a group,

$$\tilde{\mathcal{C}} = \arg\min_{\mathcal{C}} \sum_{j=1}^{k} \sum_{\mathbf{x} \in C_j} ||\mathbf{x} - \mu_j||_2^2, \tag{3.6}$$

where μ_j is the mean of the observations that belong to the set C_j. The two foreseen methods are the so called KMEANS procedure and the agglomerative hierarchical clustering procedure with Ward criterion. These two standard methods are described in [8, 11], and Ward criterion has been introduced in [22].

The KMEANS algorithm starts with a random selection of k initial means or centers, and proceeds by alternating between two steps. The assignment step permits to assign each observation to the nearest mean and the update step consists

in calculating the new means resulting from the previous step. The algorithm stops when the assignements no longer change. The algorithm that performs an agglomerative hierarchical clustering uses a bottom up strategy in the sense that each observation forms its own cluster and pairs of cluster are merged sequentially according to Ward criterion. At each step, the number of clusters decreases of one unit and the WCSS increases. This is due to Huygens theorem that divides the Total Sum of Square (TSS), into Between Cluster Sum of Square (BCSS) and WCSS,

$$\sum_{\mathbf{x}\in\mathcal{P}}||\mathbf{x}-\mu||_2^2 = \sum_{j=1}^{k}\sum_{\mathbf{x}\in C_j}||\mathbf{x}-\mu_j||_2^2 + \sum_{j=1}^{k}||\mu_j-\mu||_2^2$$

$$TSS = WCSS(k) + BCSS(k)$$

Note that TSS is constant and that BCSS and WCSS evolve in opposite directions. Applying Ward's criterion yields the aggregation of the two individuals that goes along with the smallest increase of WCSS at each step of the algorithm. The best way to visualize the data is to plot "surrender trajectories" associated with each policy as in Fig. 5. Our problem is analogous to the problem of clustering longitudinal data that arises in biostatistics and social sciences. Longitudinal data are obtained by doing repeated measurements on the same individual over time. We chose a nonparametric approach, also chosen in [9, 12]. A parametric approach is also possible by assuming that the dataset comes from a mixture distribution with a finite number of components, see [16] for instance. We believe that the nonparametric approach is easier to implement and clearer from a practionner point of view.

The KMEANS method, that takes the number of clusters as a parameter, seems to be more suited to the problem than the agglomerative hierarchical clustering method. In the Agglomerative Hierarchical Clustering (AHC) algorithm the partition in k sets depends on the previous partition. Furthermore, the KMEANS algorithm is less greedy in the sense that fewer distances need to be computed. However the KMEANS algorithm is an optimization algorithm and the common problem is the convergence to a local optimum due to bad initialization. To cope with this problem, we initialize the centers as the means of clusters builded by the AHC, thus we ensure that the centers are geometrically far from each other. The value of the BEL is highly correlated to the initial surrender value, thus a bad representation of policies having a significant initial surrender value gives rise to a significant negative impact on the estimation error after aggregation. We decide to define a weight according to the initial surrender value as

$$w_{\mathbf{x}} = \frac{SV_{\mathbf{x}}(0)}{\sum_{\mathbf{x}\in\mathcal{P}}^{n} SV_{\mathbf{x}}(0)}, \tag{3.7}$$

and we define the Weighted Within Cluster Inertia (WWCI) as

$$WWCI(k) = \sum_{j=1}^{k}\sum_{\mathbf{x}\in C_j} w_{\mathbf{x}}||\mathbf{x}-\mu_j||_2^2, \tag{3.8}$$

where μ_j becomes the weighted mean over the set C_j. Each surrender path is then stored in a group.

Remark 3 A time continuous approach within the cash flow projection model would leave us with probability density function to group. The problem would be analogous to the clustering of functional data that have been widely studied in the literature. A recent review of the different techniques has been done in [13].

3.2 The aggregation step

The aggregation step leads to the definition of a representative policy for each group resulting from the partitioning step. Probabilities of surrender depend on characteristics of the contracts and of the policyholders. The best choice as a representative under the least square criterion is the barycenter. Its surrender probabilities are defined through a mixture model

$$f_{\tau_C}(t) = \sum_{\mathbf{x} \in C} w_i f_{\tau_{\mathbf{x}}}(t), \tag{3.9}$$

where C is a group of policies and τ_C is the random early surrender time for every member of C. The equivalent within a discrete vision of time is a weighted average of the surrender probabililities with respect to each projection year. The probability density function of surrender of a given contract is associated to its age and seniority. The PDF defined in (3.9) is not associated to an age and a seniority. This fact might give rise to an operational problem if every MP in the aggregated portfolio must have an age and a seniority. The number of suitable combinations of age and seniority fall into a finite set given the possible features of policies. It is then possible to generate every "possible" surrender probability density functions in order to choose the closest to the barycenter. This optimal density function might be associated with a policy (or equivalently a combination of one age and one seniority) that does not exist in the initial portfolio.

4 Illustration on a real life insurance portfolio

The procedure is illustrated within the frame of a saving contracts portfolio extracted from AXA France portfolio. The mechanism of the product is quite simple. The policyholder makes a single payment when subscribing the contract. This initial capital is the initial surrender value that will evolve during the projection, depending on the investment strategy and the financial scenario. The policyholder is free to withdraw at any time. In case of death, the surrender value is paid to the designated beneficiaries. The contractual agreement does not specify a fixed term. The projection horizon is equal to 30 years. BEL are obtained under a discrete vision of time as in (2.6). Mortality rates are computed using an unisex historical life table. Mortality depends therefore only on the age of the insured. Lapse probabilities are computed with respect to the observed withdrawal depending on the seniority and given a financial scenario that do not trigger any dynamic lapse. The main driver that explains the withdrawal behavior is the specific tax rules applied on French life insurance contracts, see Fig. 1. Financial scenarios

are generated through stochastic modeling. The instantaneous interest rates are stochastic processes and simulations are completed under a risk neutral probability. Table 1 gives the number of policies and the amount of the initial reserves in the portfolio. Surrender probabilities depend only on the age and the seniority and thus the heterogeneity of the trajectories depends on the distribution of ages and seniorities in the portfolio. Table 2 gives a statistical description of the variable AGE. The minimum is 1 because parents can subscribe life insurance contracts for their children. Table 3 gives a statistical description of the variable SENIORITY. The range of the variable AGE and SENIORITY are completely linked to the number of trajectories in the portfolio as there is a one to one correspondence between a trajectory and a combination of age and seniority. Figure 5 displays the surrender trajectories in the portfolio. An increase of the maximum of the variable SENIORITY would increase the number of trajectories in the portfolio. Partitioning algorithms are designed to cope as far as possible with an increase of the heterogeneity of the data. This makes our aggregation procedure quite robust from one portfolio to another that might have different AGE and SENIORITY distributions.

In Sect. 3, it has been pointed out that an exact evaluation of the BEL is obtained with a portfolio that aggregates policies having the same key characteristics. In our modeling, policies that have identical age and seniority are grouped together in order to get a first aggregation of the portfolio that provides an exact value of BEL. Table 4 gives the number of policies and the BEL resulting from the valuation of this first aggregation of the portfolio.

The error is defined as the relative difference between the exact BEL and the BEL obtained with an aggregated portfolio. We also compare the two aggregation ways discussed in Sect. 3.2. One corresponds to the exact barycenter of each group (METHOD = BARY), the other being the closest-to-barycenter policy associated with an age and a seniority (METHOD = PROXYBARY).

Table 1 Number of policies and amount of the initial surrender value of the portfolio

Number of policies	Mathematical provision (euros)
140,790	2,632,880,918

Table 2 Statistical description of the variable AGE in the portfolio

Variable: AGE			
Mean	Standard deviation	Minimum	Maximum
49.09	18.57	1	102

Table 3 Statistical description of the variable SENIORITY in the portfolio

Variable: SENIORITY			
Mean	Standard deviation	Minimum	Maximum
4.10	1.63	1	7

Table 4 Number of MP and best estimate liability of the aggregated portfolio

Number of policies	BEL (euros)
664	2,608,515,602

Table 5 Best estimate liabilities error with 28 model points depending on the aggregation method

BEL error (euros) METHOD = BARY	BEL error (euros) METHOD = PROXYBARY	BEL error (euros) METHOD = Naive
−10,880	−199,734	1,074,983

These procedures are compared to a more "naive" grouping method (METHOD = NAIVE) that consists in grouping policies having the same seniority and belonging to a given class of age. The classes are defined by the quartiles of the distribution of ages in the portfolio. The ages and seniorities of the MP associated with each group are obtained through a weighted mean. The weights are defined as in (3.7). The "naive" method leads to an aggregated portfolio with 28 MPs. Table 5 reports the errors for the three methods with 28 MP. The two proposed methods outperform greatly the naive one. Figure 2 shows the error on BEL computation according to the number MPs created whitin the frame of BARY and PROXYBARY. The proposed methods permit a better accuracy even for a smaller number of MP. The use of BARY is the best choice. The question of the optimal choice of the number of clusters arises naturally. The optimal number of clusters has been widely discussed in the literature, and there exists many indicators. The main idea is to spot the number of clusters for which the WWCI reaches a sort of plateau when the number of clusters is increasing. Figure 3 displays the WWCI depending on the number of clusters. The optimal number of clusters seems to be 3 or 6. In order to automatize the choice, we can use indicators. The relevance of such indicators often depends on the partitionning problem. We need to choose the best suited to our problem. Among the indicators recommended in the literature, we find the index due to Calinsky and Harabasz [3] defined as

$$CH(k) = \frac{WBCI(k)/(k-1)}{WWCI(k)/(n-k)}, \qquad (4.1)$$

where n is the number of observations, WBCI and WWCI denote the Weighted Between and Weighted Within Cluster Inertia. The idea is to find the number of clusters k that maximises CH. Note that $CH(1)$ is not defined. This indicator is quite simple to understand, as a good partition is characterized by a large WBCI and a small WWCI. Another indicator has been proposed in [15] as follows: first define the quantity

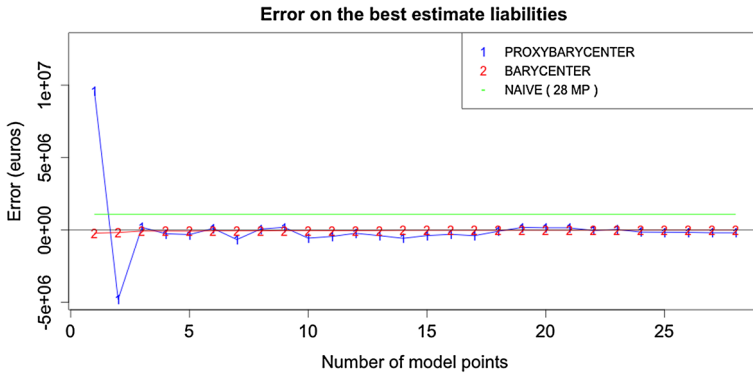$$DIFF(k) = (k-1)^{2/p} \times WWCI(k-1) - k^{2/p} \times WWCI(k), \qquad (4.2)$$

Fig. 2 Error on the BEL evaluation depending on the number of model points and the aggregation method
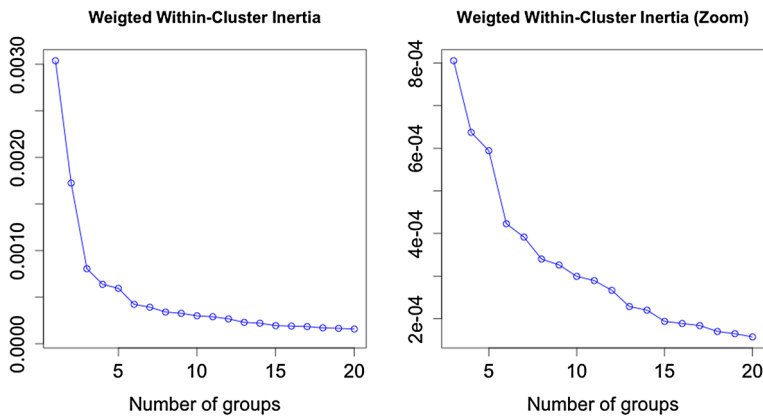

Fig. 3 WWCI evolution depending on the number of clusters

and choose k to maximize

$$KL(k) = \left| \frac{DIFF(k)}{DIFF(k+1)} \right|. \qquad (4.3)$$

This permits to compare the decreasing of WWCI in the data with the decreasing of WWCI within data uniformly distributed through space. The silhouette statistic has been introduced in [14] and is defined, for each observation i, as

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}, \qquad (4.4)$$

where $a(i)$ is the average distance between i and the others points in its cluster, and $b(i)$ is the average distance from i to the data points in the nearest cluster besides its own. A point is well clustered when $s(i)$ is large. The optimal number of clusters

maximises the average of $s(i)$ over the data set. Figure 4 displays the different indicators computed for every partition ranging from one to twenty groups. The different indicators seems to retain 3 clusters (except for KL, that is maximized for 6 clusters but still have a large value for 3 clusters). Figure 5 offers a visualization of the 3-groups partition of the portfolio. Table 6 report the errors on the BEL, which have been normalized by its exact value and expressed in percentage. The presented methods outperform greatly the naive one with less model points. Our aggregation procedure does not only performed an accurate BEL evaluation but manages also to replicate the cash flows dynamics throughout the entire projection. Figure 6 shows the accuracy on the expected present value of the exiting cash flow associated with each projection year for the 3-MP aggregated portfolio.

From a practical point of view, the optimal number of clusters should be associated with a level of error chosen by the user. We did not manage to establish a clear link between WWCI and the error on the BEL evaluations. Maybe it does not exist and we need to define another indicator, instead of WWCI, that we can
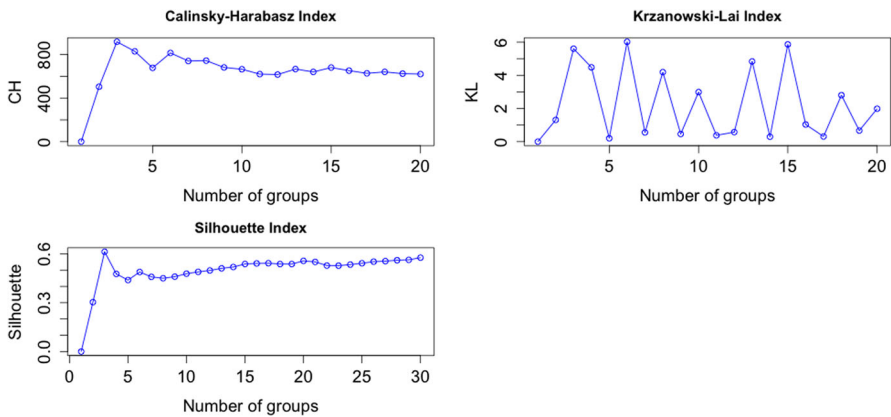

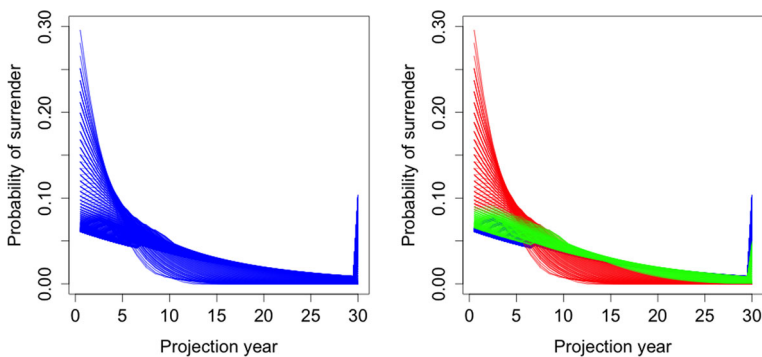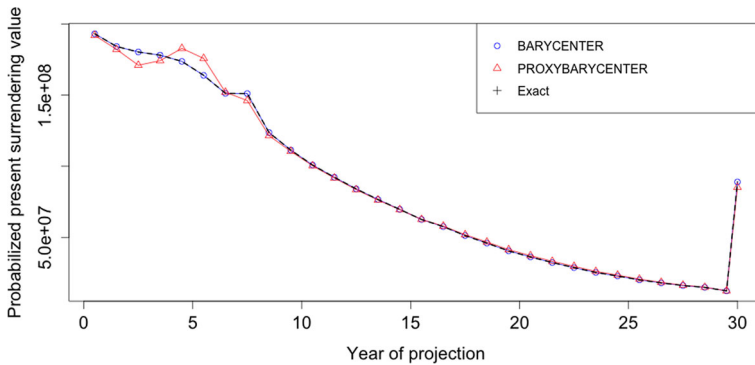**Fig. 4** Partitionning quality indicators variations depending on the number of clusters

Fig. 5 Portfolio visualization through its trajectories of surrender

Table 6 Best estimate liabilities error with 3 model points depending on the aggregation method

BEL error % BARYCENTER	BEL error % PROXYBARYCENTER	BEL error % NAIVE (28 MP)
−0.003 %	0.007 %	0.0412 %

Fig. 6 Expected present value of surrender during the projection

compute from the probabilities of surrender and that is more linked to the evaluation error. This a topic of current research. We may also add that the optimal number of clusters is not necessarily a problem that needs much thoughts from a practitionner point of view. The number of model points in the aggregated portfolio is often constrained by the capability of computers to deal with the valuation of large portfolios. The number of model points is then a parameter of the aggregation process. Another technical requirement is to take into account the different lines of business. Two life insurance products cannot be grouped together because they may have characteristics, besides their probabilities of surrender, that impact BEL computations. A solution is to allocate a number of model points for each line of business in proportion to their mathematical provision.

5 Conclusion

The aggregation procedure permits a great computation time reduction that goes along with a limited loss of accuracy. The method is easy to understand and to implement as the statistical tools are available in most datamining softwares. The application may extend to the entire scope of life insurance business, and can be tailored to many cash flow projection model based on the general definition of BEL given in the introduction. The production of aggregated portfolios can also be combined to optimization methods concerning economic scenarios. A reduction of the number of financial scenarios would permit to increase the number of model

points. The optimal gain of accuracy would result from a trade-off between the number of model points and the number of financial scenarios.

This work represents the successful outcome of a Research and Development project in the industry. It is already implemented in AXA France, from a portfolio that contains millions of policies, the output is a portfolio of only a few thousands model points for a 0.005% error on the BEL. It remains many rooms for improvement, especially at the partitioning level where distance other than euclidean might be better suited to quantify the distance between trajectories. The definition of an indicator that gives insight on the error resulting from the aggregation would be as well a great improvement. For instance, the computation of this indicator may provide an optimal number of model points allocated to each line of business and therefore the whole portfolio.

References

1. Asmussen S, Albrecher H (2010) Ruin probabilities, vol 14. World scientific
2. Bonnin F, Planchet F, Julliard M (2014) Best estimate calculations of savings contracts by closed formulas: application to the ORSA. Eur Actuar J 4(1):181–196
3. Calinski T, Harabasz J (1974) A dendrite method for cluster analysis. Commun Stat 3:1–27
4. Chauvigny M, Devineau L, Loisel S, Maume-Deschamps V (2011) Fast remote but not extreme quantiles with multiple factors: application to Solvency II and Enterprise Risk Management. Eur Actuar J 1(1):131–157
5. Devineau L, Chauvigny M (2011) Replicating portfolios: calibration techniques for the calculation of the Solvency II economic capital. Bull Français d'Actuariat 21:59–97
6. Devineau L, Loisel S (2009) Construction d'un algorithme d'accélération de la méthode des "simulations dans les simulations" pour le calcul du capital économique Solvabilité II. Bull Français d'Actuariat 10(17):188–221
7. EIOPA (2010) Quantitative Impact Studies V: technical specifications. Technical report, European comission, Brussels
8. Everitt BS, Landau S, Leese M (1995) Cluster analysis, 4th edn. A Hodder Arnold Publication
9. Genolini C, Falissard B (2010) KML: K-means for longitudinal data. Comput Stat 25(2):317–328
10. Gerber HU (1990) Life insurance mathematics. Springer, Berlin
11. Hartigan J (1975) Clustering algorithms. Wiley, New York
12. Hejblum B, Skinner J, Thiebaut R (2012) Application of gene set analysis of time-course gene expression in a HIV vaccine trial. In: 33rd Annual conference of International Society for Clinical Biostatistics, Bergen
13. Jacques J, Preda C (2013) Functional data clustering: a survey. ADAC 8(3):231–355
14. Kaufman L, Rousseeuw PJ (1990) Finding groups in Data: an introduction to cluster analysis. Wiley, New York
15. Krzanowski WJ, Lai YT (1988) A criterion for determining the number of groups in a data set using Sum-of-Squares clustering. Biometrics 44:23–34

16. Nagin DS (1999) Analysing developmental trajectories: a semiparametric group based approach. Psychol Methods 4:139–157
17. Nteukam TO, Planchet F (2012) Stochastic evaluation of life insurance contracts: model point on asset trajectories and measurements of the error related to aggregation. Insur Math Econ 51(3):624–631
18. Petauton P, Kessler D, Bellando JL (2002) Théorie et pratique de l'assurance vie: manuel et exercices corrigés. Dunod
19. Planchet F, Therond P, Juillard M (2011) Modèles financiers en assurance. Analyses de risque dynamique, 2nd edn. Economica, Paris
20. Sarukkali M (2013) Replicated stratified sampling for sensitivity analysis. Ph.D. thesis, University of Connecticut
21. Schrager D (2008) Replicating portfolios for insurance liabilities. AENORM 59:57–61
22. Ward JH (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58:236–544