METHOD FOR FAST ESTIMATION OF LIFE INSURANCE LIABILITIES WITH RESPECT TO DIFFERENT INVESTMENT STRATEGIES

Jan Fojtík – Jiří Procházka – Pavel Zimmerman – Markéta Švehláková – Simona Macková

Abstract

The goal of this paper is to find approximate methods for estimating liabilities for life insurance contracts. Traditionally used methods in this field are usually very time-consuming and it can take days or even weeks to obtain estimates of liability for an average-sized portfolio. In general insurance companies test the sensitivity of liability estimation with respect to the change of many assumptions. Reduction of computational time could improve the efficiency of the task of testing different assumptions because the estimations need to be fast and should reflect current market development. Furthermore, a fast calculation is a crucial condition for successful optimization of investment strategy and deeper analyses of liability sensitivity. A new approximating approach which reduces computational time and provides results with a tolerable difference from the results of the traditional methods is more than welcome in an actuary practice. This contribution focuses on the time aspect of the modeling liability estimation based on a cluster analysis and the quality of the estimation.

Key words: life insurance, estimation of BEL, cluster analyses, scenario analysis

JEL Code: C38, C63, G22

Introduction

The value of liability is an essential metric used for determining economic profit and solvency position of an insurance company. The main goal of actuary work is to calculate the best estimate of liability (BEL). Correct estimation of BEL follows certain assumptions which should be based on realistic, current and credible information (e.g. expected development of interest, lapse or mortality rates) see (Selimovic, 2010). Important task is to capture changes of liability estimate on different initial assumptions or on changes in investment strategies (Giamouridis, 2016; Kaucic, Daris, 2015). Typically, the more scenarios are tested the better information about liability distribution is obtained. Hundreds or thousands of different scenarios can usually be created to study the sensitivity of liability. When calculating high number of scenarios, the main issue is calculation time. Supposing that liability estimation of medium sized portfolio (300 000 model-point) lasts about 10 minutes,[1] simulation of thousands of scenarios would take 10 000 minutes which is almost a week. Results derived with such a delay may be outdated or not satisfying actual assumptions. Therefore, time aspect is an important factor for liability calculation. An actuary has usually only a few possibilities to reduce computational time, such as purchasing faster hardware or software, buying more machines or test fewer scenarios. These options may bring extra expenses or a loss of important information. Moreover, these solutions are only temporary and no single one solves the entire problem. Several statistical or data mining researchers have already been studying this topic see (Mohammed, et al. 2016). Cluster analysis seems to be a useful alternative for valuating life insurance portfolio (Freedman, 2008). Cluster analysis is only an approximate method to cash flow approach and therefore the results may be inaccurate but derived in a shorter time.

The main idea of cluster analyses is to create a few weighted reference model points representing original insured portfolio which is faster than using all the model points in the portfolio. The primary focus of this paper is to study the sensitivity of computational time with respect to the number of clusters.

1 Input model points

For the purpose of this paper real market data are used. The data include model points representing a real portfolio of life insurance products. Insurance companies usually have hundreds of thousands of contracts. Therefore, the size of a portfolio of 106 524 contracts could be described as a medium-sized portfolio for a smaller insurance company. Each contract represents one model point with personal information about a client. Due to the fact, that the used portfolio is a real market portfolio and includes confidential information about clients, the adjustments have been made to secure personal information. A sample of the portfolio is presented in Table 1.

The presented portfolio consists of four different product types, A, B, C and D. Each product differs in settings of charges, surrender fees or surrender period. *Entry age* is the age of the insured person at the beginning of the contract. The *Annual premium* is the amount of money being paid by the policyholder to the insurance company for coverage of risk. Gender

[1] Assumption for such fast calculation is using specialized actuarial software and hardware with high performance

is given by the variable *sex*; 0 stands for male and 1 for female. The length of policy period is presented in months in the column *Policy term in months*. *Duration in force* represents how long the contract has been in force. *Sum assured* is the guaranteed amount of money the policyholder receives in case of death of the insured person. *Fund value* represents client's account of savings, usually reinvested.

Tab. 1: Sample of insured portfolio

Product type	Entry age	Annual premium	Sex	Policy term in months	Duration in force	Sum assured	Fund value
C	22	2 321	1	588	252	93 745	53 927
A	30	2 629	1	444	288	117 024	61 930
B	24	2 364	1	480	180	93 267	43 902
A	41	7 955	1	384	204	233 706	128 731
C	39	4 122	1	264	204	101 248	40 148

Source: the author's

All input data described above include information only about basic characteristics of each client (e.g. age, premium, sum assured) but no information about its economic profit. In the next section, methods to calculate metrics determining economic profit of each contract are introduced.

2 Methodology

Several methods can be used for determining values of economic profit of life insurance product many of which are based on analyzing cash flow (CF) projections (e.g. present value of future cash flow or present value of future profit). To calculate values of economic profit, the cash flow model needs to be built first. Basic introduction to cash flow model construction is described in (Cipra, 2014). This section introduces a brief description of commonly used metrics of economic profit and presents a basic cash flow model.

The cash flow model, as well as the whole study, is created in R software. Cluster analysis is performed using package *Cluster* see (Maechler, et al. 2017). The authors are aware of advantages and disadvantages of R programming. Therefore, some steps to minimize calculation time are considered (e.g. avoiding unnecessary loops and vectorization).

2.1. Cash flow projection

We present several variables usually used in the actuar practice:

- present value of future cash flows (PVFC);
- present value of profit and loss (PVPL);
- present value of premium (PVPrem);
- present value of distributive earnings (PVDE).

Clustering algorithm uses these metrics to group individual model points into clusters. Economic metrics are based on the cash flow projection of insured portfolio. In the first step, the cash flow model is built. To calculate the economic profit of the whole portfolio, the cash flow model is applied on each model point separately and the results are cumulated. In case of many model points, this procedure may be time-consuming. As presented in the results section, calculation of economic profits of presented portfolio by cash flow methods lasts more than 2 hours[2]. The basic idea of cash flow model is as follows:

$$\mathrm{CF}_t = E\mathrm{Prem}_{t-1} - E\mathrm{Surrender}_t - E\mathrm{Death}_t - E\mathrm{Maturity}_t - E\mathrm{Commission}_t - E\mathrm{Expenses}_t \quad (1)$$

Each component of the cash flow model is the expected value of particular random variables. For example, expected death benefit ($E\mathrm{Death}_t$) is the expected value of benefit paid in case of death at time t adjusted by the probability of dying at time t. $E\mathrm{Prem}_{t-1}$ stands for expected premium at the beginning of the period t, $E\mathrm{Surrender}_t$ stands for expected surrender value at the end of period t, $E\mathrm{Maturity}_t$ stands for expected value at maturity, $E\mathrm{Commission}_t$ represents expected commission and $E\mathrm{Expenses}_t$ stands for expected expenses. For simplicity, the model is built on an annual basis and index t stands for the year of the projection. Values with index t are at the end of projected year.

2.2. Scenario testing

Testing the sensitivity of liabilities or the economic profit is an important aspect of the actuarial work (Jensen, 2016). To provide a quality test of clustering model, results of several scenarios need to be compared, where one scenario representing changes of various assumptions such as lapse rate, mortality rate, expenses or interest rate.

The list of scenarios is presented in Table 2. Later, the results of all scenarios for the approximate model and the cash flow model are compared to demonstrate the quality of the estimation. The portfolio of shocks was inspired by Solvency II directive (EIOPA, 2009), as a mandatory guide of risk analyses for insurance companies.

[2] Calculation performed in R 3.3.0 software on desktop i5 3.30GHz, 4 GB RAM, Intel HD Graphics 2000

Tab. 2: List of scenarios with shocks

Scenario	Shock
1 - Best estimate	No shock applied
2 - Mortality rate up	Mortality rate increased by 15%
3 - Mortality rate down	Mortality rate decreased by 20%
4 – Lapse rate up	Lapse rate increased by 50%
5 – Lapse rate down	Lapse rate decreased by 50%
6 – Interest rate up	Interest rate increased by 2pb
7 – Interest rate down	Interest rate decreased by 2pb

Source: the author's work

2.3. Cluster analysis

Cluster analysis is a common technique for statistical data analysis with the aim to reduce dimension. The task of cluster analysis in this application is to group model points with a similar development to one model point representing the whole group. Such reduction of the portfolio dimension may reduce the computation time of economic metrics.

2.3.1. Clustering algorithm

Cluster analyses is a tool with a variety of algorithms. A good choice of a clustering algorithm is one of the centroid model algorithms, where the center of each cluster is represented by medoids i.e. one concrete model point. The algorithm used in this paper is Clustering Large Applications algorithm formally known as CLARA (NG, 2002). Several clusters are created applying CLARA on the insurance portfolio, the center of each cluster being represented by a reference model point. This algorithm measures dissimilarities between each model point by Euclidean distance. Euclidean distance between *i*-th model point MP_i and *j*-th model point MP_j is defined as

$$d\left(MP_i, MP_j\right) = \sqrt{\sum_{k=1}^{K}\left(X_{k,i} - X_{k,j}\right)^2}\ , \qquad (2)$$

where $X_{k,i}$ and $X_{k,j}$ for $k = 1, 2, \ldots, K$ are values of clustered variables for *i*-th model point MP_i and *j*-th model point MP_j.

2.3.2. Variable selection

Clustering is performed under K variables. Metrics of economic profit described in the previous section (PVFC, PVPL, PVDE, PVPrem) and values of a cash flow projection are used as a clustering variables. The cash flows from the 11th year on are summed to one value. All the clustering variables are calculated on the assumption of the best estimate. To improve the clustering, the clustering variables need to be modified. If values as cash flows are clustered, the algorithm may group according to the nominal values instead of the development of individual cash flows. For example, two model points with similar development but different nominal values of cash flow may be considered very different, but in fact, they are quite similar because of the similar cash flow development. In order to solve this problem, a relative measure of the variable *k* and model point *i* $R_{i,k}$ need to be calculated by adjusting clustering variables $X_{i,k}$ by reference variable V_i of *i*-th model-point. PVFV has been selected as a reference variable V_i, therefore relative value is as follows:

$$R_{i,k} = \frac{100 \cdot \left(X_{i,k} - V_i\right)}{V_i} \tag{3}$$

After transformation to relative values, CLARA algorithm is applied to obtain homogenous[3] clusters and selects the representatives based on the smallest distance of trajectories. After the first step, the representative model points of each cluster are selected. These representatives are not necessarily the best choice in order to approximate full portfolio. Therefore, in the next iteration, a better choice of representative is searched in each cluster, if any. Afterward, the weight for each representative is determined as the sum of reference variable (PVFC) over each model points within the cluster.

3 Results

In order to determine an appropriate number of clusters for presented portfolio, 50, 100, 200, 300, 400 and 500 number of clusters were analyzed. The performance of clustering approach is measured as a percentage deviation of approximated portfolio from the best estimate projections. For each selection of clusters, the clustering procedure was repeated ten times because results of CLARA algorithm are random and the partition of portfolio into clusters can slightly differ.

After necessary clustering variables are obtained, the estimation of liability by clustering approach is done in two steps:

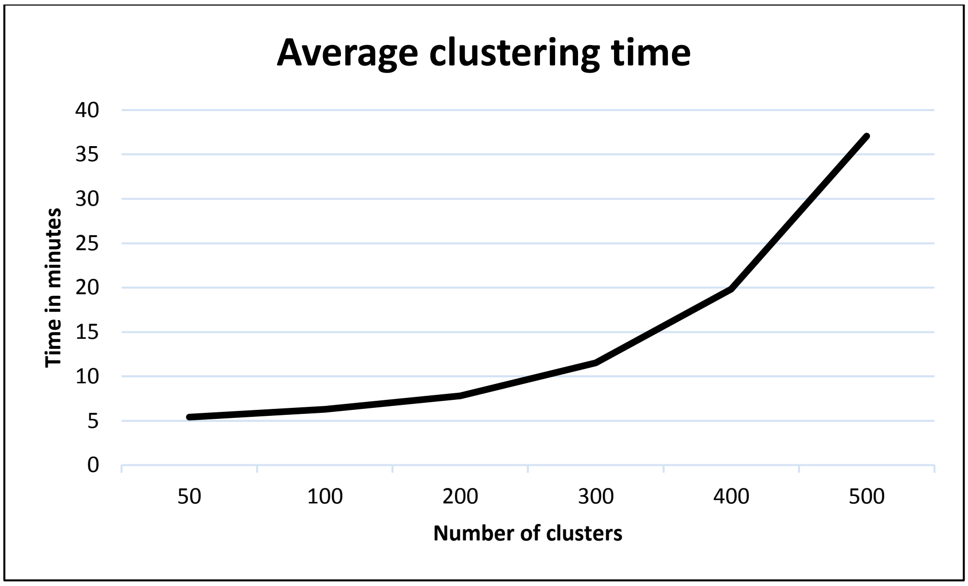
- clustering – selecting reference model points representing original portfolio;

[3] Within cluster variability is low – model point in one cluster are similar

- projection of cash flows and metrics of economic profit under a different scenario.

As Figure 1 presents, clustering time grows exponentially with the number of clusters.

Fig. 1: Dependence of clustering time on number of selected clusters



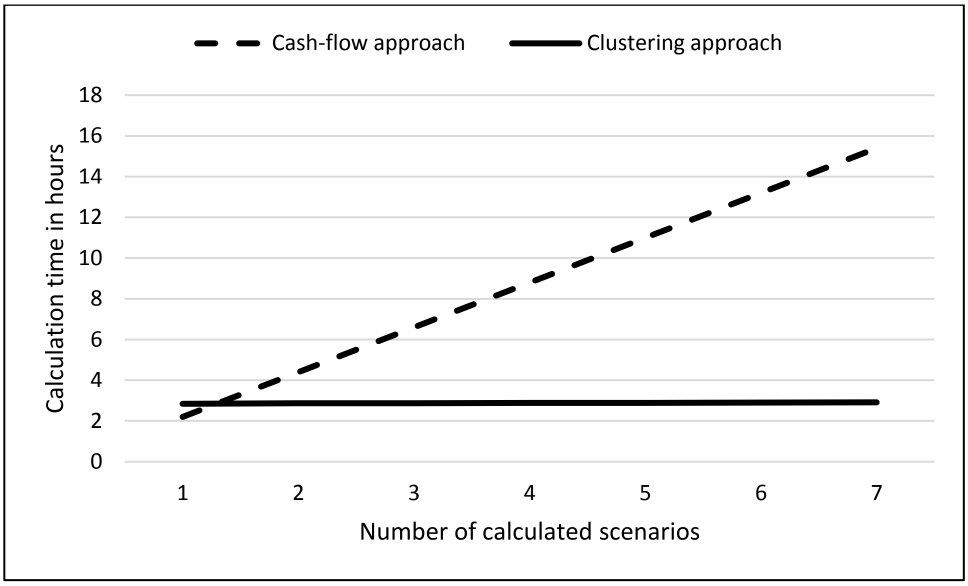
Source: the author's work

The calculation time of cash flows and economic profit projection also depends on the number of selected clusters, the time is derived by:

$$T \cdot \frac{N_{\text{clusters}}}{N_{\text{modelpoints}}}, \qquad (4)$$

where T is calculation time of one scenario by the cash flow approach, N_{clusters} represents the number of clusters and $N_{\text{modelpoints}}$ is the number of model points in the original portfolio. In the case of the original portfolio with 106 524 model points, the calculation of one scenario takes about 2.2 hours, but the calculation of one scenario using clustering approach with 500 clusters (representative model points) takes only 1.2 minutes. Figure 2 shows time efficiency of clustering approach compared to the cash flow model. Results are presented in seven different scenarios. Note that clustering approach includes time of scenario calculation (1.2 minutes per one scenario), clustering time (40 minutes) and time used for calculating clustering variables by the cash flow model (2.2 hours).

Fig. 2: Comparisons of cash flow and clustering approach with 500 clusters



Source: the author's work

Table 3 presents results of the clustering approach with three different error measurements, namely average error, median error and maximal error. It is obvious that increase of the number of clusters leads to lower estimation error but to higher computational time**.**

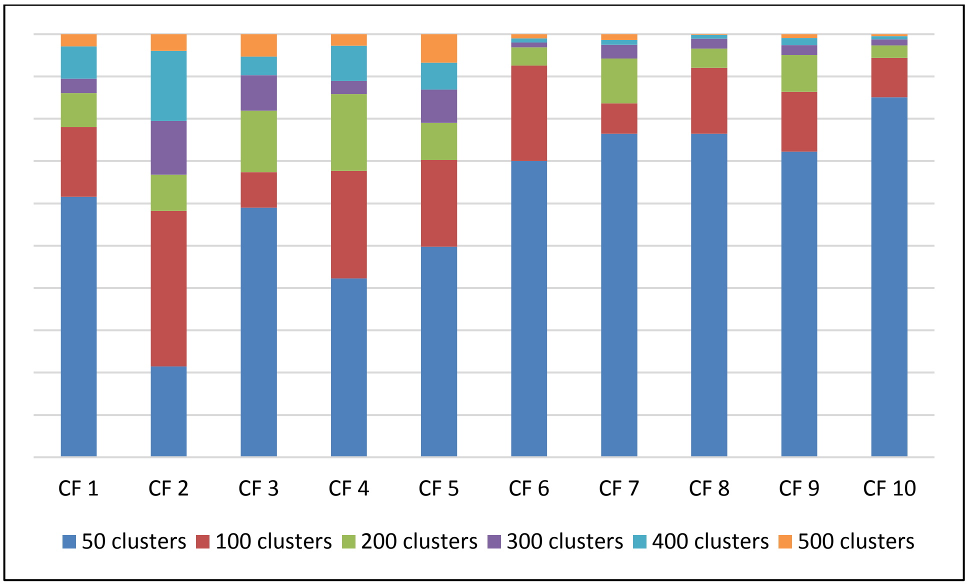
Tab. 3: Error of clustering algorithm over different number of clusters

Product type	Average error (%)	Median error (%)	Maximal error (%)	Computational time (minutes)
50 clusters	0.283	0.058	4.376	4.835
100 clusters	0.101	0.015	1.154	5.222
200 clusters	0.058	0.006	2.289	7.129
300 clusters	0.047	0.004	3.143	11.406
400 clusters	0.031	0.002	1.138	20.799
500 clusters	0.019	0.002	0.686	40.952

Source: the author's work

The proportion of error with respect to the number of clusters can be seen in Figure 4 for selected cash flows. Both cases suggest that the more clusters are used, the better estimation of individual metrics are obtained.

Fig. 4: Proportion of error on cash flows (median error)



Source: the author's work

Conclusion

Proper use of the cluster analysis approach requires selecting several optional attributes such as the number of clusters or clustering variables. The number of selected clusters has a direct impact on calculation time and the precision of the clustering approach. A large number of clusters increase the precision of clustering approach but also the clustering time. On the other hand, a low number of clusters reduce the computational time but it may lead to poor precision of clustering approach. Clustering approach is divided into two steps. In the first step, the proper number of clusters and reference model points are selected. In the second step, the individual scenarios are calculated using the reference model points. Dividing clustering approach into two steps is the main benefit compared to the ordinary cash flow model. Calculating different scenarios using the ordinary cash flow is slow because the whole portfolio must be recalculated for each scenario. In the clustering approach, the whole portfolio is calculated only once to obtain the clustering variables and reference model points and only the reference model points are used for scenario results. As the results suggest, the cluster can bring a significant reduction in the computational time.

Acknowledgment

The support of grant scheme Methods for fast estimation of life insurance liabilities with respect to different investment strategie IG410017 is gladly acknowledged.

References

Cipra, T. (2014). *Financial and insurance formulas*. Place of publication not identified: Physica Springer.

EIOPA 2009: Solvency II Directive: DIRECTIVE 2009/138/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, https://eiopa.europa.eu/.

Freedman, A., Reynold, C., W. 2008. Cluster analysis: A spatial approach to actuarial modeling. http://www.milliman.com/uploadedFiles/insight/research/life-rr/cluster-analysis-a-spatial-rr08-01-08.pdf.

Giamouridis, D., Sakkas, A., & Tessaromatis, N. (2016). Dynamic Asset Allocation with Liabilities. *European Financial Management, 23*(2), 254-291. doi:10.1111/eufm.12097

Jensen, N. R. (2016). Scenario-based life insurance prognoses in a multi-state Markov model. *European Actuarial Journal, 6*(2), 307-330. doi:10.1007/s13385-016-0137-8

Kaucic, M., & Daris, R. (2015). Multi-Objective Stochastic Optimization Programs for a Non-Life Insurance Company under Solvency Constraints. *Risks, 3*(3), 390-419. doi:10.3390/risks3030390

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2017). cluster: Cluster Analysis Basics and Extensions. R package version 2.0.6.

Mohammed, M., Youssef, B., & Taoufiq, G. (2016). Time-Saving Approach for Optimal Mining of Association Rules. *International Journal of Advanced Computer Science and Applications, 7*(10). doi:10.14569/ijacsa.2016.071031

Ng, R., & Han, J. (2002). CLARANS: a method for clustering objects for spatial data mining. *IEEE Transactions on Knowledge and Data Engineering, 14*(5), 1003-1016. doi:10.1109/tkde.2002.1033770

Selimovic, J. (2010). Actuarial Estimation of Technical Provisions' Adequacy in Life Insurance Companies. In Interdisciplinary Management Research-Interdisziplinare Managementforschung *6, 523-533*.

Contact

Jan Fojtík
University of Economics, Prague
Czech Republic W. Churchill Sq. 1938/4
130 67 Prague 3 - Žižkov
xfojj00@vse.cz

Jiří Procházka
University of Economics, Prague
Czech Republic W. Churchill Sq. 1938/4
130 67 Prague 3 - Žižkov
xproj16@vse.cz

Pavel Zimmermann
University of Economics, Prague
Czech Republic W. Churchill Sq. 1938/4
130 67 Prague 3 - Žižkov
zimmerp@vse.cz

Markéta Švehláková
University of Economics, Prague
Czech Republic W. Churchill Sq. 1938/4
130 67 Prague 3 - Žižkov
xsvem35@vse.cz

Simona Macková
University of Economics, Prague
Czech Republic W. Churchill Sq. 1938/4
130 67 Prague 3 - Žižkov
macs03@vse.cz