PRECISION OF ASSET LIABILITY MODEL WITH FAST LIABILITIES ESTIMATION BY CLUSTER ANALYSIS

JAN FOJTÍK, JIŘÍ PROCHÁZKA, PAVEL ZIMMERMANN, SIMONA MACKOVA, MARKÉTA ŠVEHLÁKOVÁ,
University of Economics, Prague, Faculty of Informatics and Statistics,
Department of Statistics and Probability,
W. Churchill Sq. 4, Prague, Czech Republic
e-mails: xfojj00@vse.cz, xproj16@vse.cz, zimmerp@vse.cz, xsvem35@vse.cz,
simona.mackova@vse.cz

Abstract

Accurate modelling of insurance liabilities which is able to reflect the time value of financial options and guarantees (FOGs) is one of the essential actuarial tasks required by Solvency II and IFRS. Liability modelling of large insurance portfolio that requires running thousands of economic scenarios is very demanding on computational time. In our previous research, we proved that cluster analysis, which limits the number of the modelpoints to be modelled, is a good approximative method which can be used to speed up the liability estimation while preserving the accuracy. Using a faster method for the estimation of the liability value allows the actuaries to study more investment strategies and provide a more complex analysis or sensitivities to understand market risk or manage proper investment strategies by asset liability management. In this paper, we focus on the application of the dynamic asset liability management in life insurance business and the faster liability model based on cluster analysis. In the results, we present the answer to whether the faster liability model is suitable for two basic ALM methods – cash flow matching and duration.

Key words: *Life insurance, estimation of BEL, cluster analysis,*

JEL Codes: *C38, C63, G22,*

1. Introduction

New regulations and a stronger competition have increased the importance of responsible administration of assets and liabilities. Therefore, the asset liability management (ALM) is more and more important in the modern insurance business. The ALM is an ongoing process following certain competitive or profitability goals by managing investment strategies (Gilbert, 2016). The usual goal of ALM is to reach higher profit and manage overwhelming risk. Traditional application techniques of ALM models are based on methods such as cash flow matching (Xidonas, 2018), duration matching (Shang, 2018) or analysis of value at risk (Balestreri, 2011). It can be assumed that new techniques of ALM will be in high demand nowadays, especially with stronger regulations of Solvency II or IFRS (EIOPA, 2009).

The ALM models have high importance, especially in life insurance business where the insurance products include participation (Aas, 2018). This fact complicates whole life insurance modelling because the value of liability may depend on the value of assets and assets returns. A change of an investment strategy may also result in a change of the market value of liabilities and SCR due to the change in assets risk profile. This fact may have a simultaneously negative influence on own capital and final profit. The actuarial task is to prevent such an action by selecting a proper investment strategy. Actuaries usually calculate thousands of investment scenarios to obtain full information about the profitability and risk profile of the invested assets.

The first issue complicating ALM modelling in life insurance is the calculation time of liability model. The traditional liability models are based on cash flow projection of each contract (per-policy modelling), which is time demanding. Application of faster liability model is an important aspect of ALM modelling because it allows actuaries to test more scenarios, provide stochastic ALM model (Fernandez, 2018) and derive results in an acceptable time so that the results are not out-dated. Let us suppose the calculation of one ALM scenario on the mid-sized portfolio (300 000 model-points) lasts about 10 minutes. The simulation of thousand scenarios then lasts 10 000 minutes which is almost a week. Results derived with such a delay may not be actual and do not reflect current market position. The time aspect of liability modelling is crucial for the actuality of ALM models. Based on our previous research (Fojtík, 2017) and research based on Freedman (2008), we present cluster analysis as a good approximative method speeding-up liability modelling.

In this paper, we study whether the faster liability model can be used for the purpose of ALM modelling. In the first part, we introduce the benefits and time efficiency of faster liability model based on clustering approach and we compare it with the traditional per-policy model. Later we present how the income from assets is realized and how it can be used to optimize investment strategy in ALM. For this purpose, we present two ALM methods – cash flow matching and duration matching. The primary focus of this paper is to study whether the cash flows or liability duration calculated by the faster liability model are equal to the values from the traditional per-policy liability model.

2. Liability models

In this part, we introduce the main difference and time efficiency between traditional liability model based on per-policy projection and faster liability model based on cluster analysis.

2.1 Traditional approach for liability modeling

Traditional approach for liability modelling is based on cash flow projection of each policy. Calculating projection of each modelpoint on the monthly basis is highly time-demanding. Let us assume we have 100 000 modelpoints with prediction horizon 50 years (600 months). Such a calculation takes 60 000 000 interactions. Even with the newest software and hardware, the results will be calculated with significant delay. The traditional per-policy liability model can be represented by the following formula:

$$CF_t = Prem_t - Surr_t - Death_t - Mat_t - Comm_t - Exp_t + Intr_t \,, \tag{1}$$

where $Prem_t$ stands for expected premium at the beginning of the period t, $Surr_t$ stands for expected value of surrenders, $Death_t$ stands for expected death outgo in time t, Mat_t stands for expected value of maturities, $Comm_t$ stands for expected value of commissions, Exp_t stands for expected value of expenses at the end of the period t and $Intr_t$ stands for expected returns from investments at time t.

2.2 Liability modeling using cluster analysis

Liability model based on cluster analysis is an approximative method which enables insurance companies to calculate the liability value in significantly shorter time. The main principle consists in the reduction of the size of the original portfolio because modelling smaller amounts of modelpoints (contracts) saves time. Cluster analysis groups modelpoints with certain similarity pattern into clusters from which a limited number of modelpoints can be chosen to create a smaller representative portfolio. Such a portfolio should represent original portfolio with high precision.

In order to group modelpoints into clusters, it is necessary to define a set of clustering variables used as a measure of similarity between the modelpoints. One can use the attributes already available in the dataset such as the characteristics of the insured person or the properties of the policy. However, these variables might have an ambiguous impact on the cash flow development and they may not lead to accurate results. An alternative approach is to use metrics of economic profit such as present value of future cash flow (PVFC), present value of profit (PVPL), present value of premium (PVP) or individual values of cash flow projection. On the one hand, these variables cannot be directly obtained from the dataset and they need to be computed first. On the other hand, such variables can better characterize the development of the liabilities, which is why they are used in this paper. As the model is more concerned with the development of the variables than with their nominal values, all clustering variables need to be adjusted to their relative values using

$$R_{i,k} = \frac{100 \cdot \left(X_{i,k} - V_i\right)}{V_i}, \tag{2}$$

where $R_{i,k}$ represents the adjusted value of the *k-th* variable of the *i-th* model point, $X_{i,k}$ is the non-adjusted value and V_i is the reference variable.

There are various measures which can be used for quantifying the dissimilarities between each modelpoint. The measure used in this paper is Euclidian distance. The distance between the *i-th* and the *j-th* modelpoint is defined as:

$$d\left(MP_i, MP_j\right) = \sqrt{\sum_{k=1}^{K}\left(X_{k,i} - X_{k,j}\right)^2}, \tag{3}$$

where $X_{k,i}$ is the value of the *k-th* clustering variable.

The clustering is performed using CLARA algorithm (NG, 2002). This clustering algorithm is derived for large datasets and it utilizes the technique of sampling in order to reduce the computational time. It starts with selecting a random sample from the portfolio and finding k medoids in the sample. These medoids are then used for clustering the whole portfolio. The process is repeated a pre-specified number of times searching for a better choice for medoids. The final choice of medoids can be used to create a representative portfolio. Since each cluster comprises a different number of model points, the medoids need to be weighed. In this paper, the sums of PVFC over each modelpoint within the relevant cluster are used as weights. The calculation is done in statistical software R using package Cluster see (Maechler, 2017).

Our previous research has demonstrated that using the reduced portfolio instead of the original one enables one to obtain relatively accurate results at a much lower computational time. It has been shown that both computational time and accuracy increase with the increasing number of clusters. The accurate computation based on the traditional cash flow analysis takes about 2.3 hours[1]. Using 10 clusters, one can obtain the results in 0.12 hours

[1] Computer used for this calculation has following performance: i5-6500 CPU @ 3.20 GHz, 4 GB RAM, Windows10

with the average accuracy of 80% while the computation based on 200 clusters takes about 0.73 hours and average accuracy is 99.6%. It could be concluded that cluster analysis is a useful tool for decreasing the computational time of valuation of life insurance portfolios.

2.3 Time comparisons of both approaches

The clustering approach is designed to speed-up liability estimation. The benefits of faster liability modelling are noticeable especially when testing a high number of scenarios. The formula for calculation time of traditional (per-policy) approach for $N_{\text{scenarios}}$ is given by:

$$T = T_{\text{CF}} \cdot N_{\text{scenarios}}\,, \tag{4}$$

where T_{CF} is calculation time of one scenario. The formula for calculation time of faster clustering approach for $N_{\text{scenarios}}$ is given by:

$$T = T^{*}_{CF} + T^{*}_{CL} + T_{CF} \cdot \frac{N_{\text{cluster}}}{N_{\text{Modelpoint}}} \cdot N_{\text{scenarios}}\,, \tag{5}$$

where N_{cluster} is the number of selected clusters and T_{CL} is clustering time. Note that the first part of this equation T^{*}_{CF} is the time required for the calculation of clustering variables. In part 2.2 we have presented that an eligible selection of clustering variables consists of metrics of economic profit such as variables based on individual cash flows or present values of profit or premium. Usually, the insurance portfolio includes only basic information about the contracts (age, sex, premium, …) and the metrics of economic profit need to be calculated at initialization.

When computing a high number of scenarios, the clustering time T_{CL} and time required to obtain clustering variables T^{*}_{CF} is negligible. The time efficiency of clustering approach can be described by the ratio:

$$\frac{N_{\text{cluster}}}{N_{\text{Modelpoint}}}\,. \tag{6}$$

3. Assets models

In this part, we present a basic asset model of the insurance company and we show how the assets model generates its cash flows. Commonly used assets in the insurance business are low-risk assets such as government bonds, corporate bonds or stocks. The risk profile of each asset is an important driver for insurance companies, because high risk may increase the value of solvency capital requirement. Therefore, a proper management of assets and liabilities takes high importance in life insurance business. In next sub-chapters, we introduce the valuation of fixed bonds and the way asset model consisted of fixed bonds generates the cash flows.

3.1. Bond valuation

Traditional bond valuation techniques are based on discounting future cash flows. The value of a bond is defined as a present value of future cash flows (Cipra, 2010). The general formula for bond valuation has the following form:

$$P = \sum_{t=1}^{T} \frac{CF_t}{(1+r)^t} + \frac{F}{(1+r)^T}, \tag{7}$$

where r is a spot rate with a tenor in time t. CF_t is a coupon at time t and F is a face value at maturity ($t = T$). The total value of asset portfolio at time t is given by the sum of present values of all bonds in time t.

3.2. Assets cash flow and total return

In general, the calculation of assets return depends on several factors such as the type of the asset or accounting scheme. The assets income is given by the sum of cash flows (coupons and face values) from all assets in the portfolio. The total return is then given by the total income divided by the current market value of the assets portfolio.

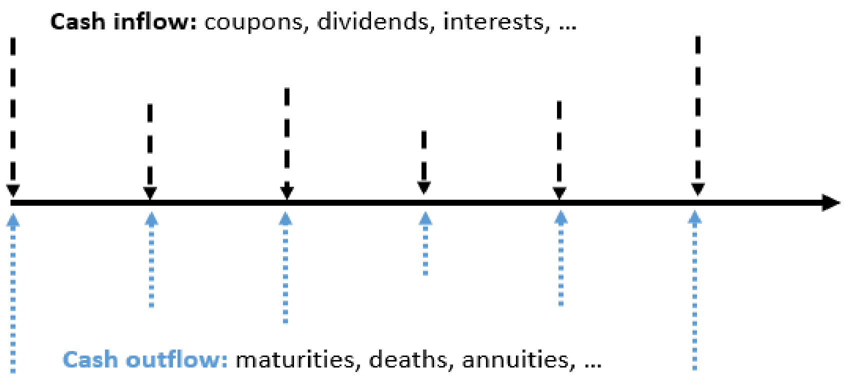
4. Optimization methods

The standard methods used in ALM to optimize investment strategies can be a method based on cash flow matching or duration matching. In this sub-chapter, we introduce the basic idea of optimization methods and their application and implementation into faster liability model based on clustering analysis.

4.1. Cash flow matching

The cash flow matching approach can be defined as a process of hedging in which an insurance company matches its cash outflow with cash inflow. As an outflow, we may understand the liability cash flow and as an inflow, we may understand the assets cash flow. The liability cash flow is defined in section 2.1 as the difference between income (premium) and outgoes (benefits, expenses and commissions). If the liability cash flow is negative (income is lower than outgo) then the insurance company needs to find a source to finance this loss. It can be done by setting investment strategies with same or higher income as the liability cashflow. This approach is called cash flow matching and its main principle can be seen in figure 1.

Figure 1: Principle of cash flow matching in life insurance



Source: Authors' work.

The advantage of cash flow matching principle is its simplicity. The insurance company can project its expected liability development and choose an investments strategy to cover each liability cash flow. The disadvantage of this approach is the fact that it does not reflect the change in the yield curves. When discount rates change the selected strategy may not be optimal and its income may not cover cash flow from liabilities.

4.2. Duration matching

Another commonly used optimization method in ALM is an approach based on duration matching. This approach is designed to reflect changes in yield curves because in classical bond theory, the terminology of duration represents the sensitivity of market value to the change in the yield curve. The insurance company tries to match its assets duration with the duration of its liabilities.

Duration of assets

The assets duration is a sensitivity of market value to the change in the yield curve. The formula for assets duration (Tsai, 2009) is following:

$$MD = -\frac{MV\left(YC+\Delta i\right) - MV\left(YC-\Delta i\right)}{MV\left(YC\right)} \cdot \frac{1+\Delta i}{2\Delta i}, \tag{8}$$

where:

$MV\left(YC+\Delta i\right)$ is the market value of asset calculated based on yield curve increased by Δi.

$MV\left(YC-\Delta i\right)$ is the market value of asset calculated based on yield curve decreased by Δi.

Δi is equal to 0.0001

Duration of liabilities

Analogically to the assets duration, we can define the duration of liability. Liability duration can be defined as a sensitivity of liability value (present value of future cash flows) to the change in the yield curve. The formula for liability duration has the following form:

$$MD = -\frac{PVFC\left(YC+\Delta i\right) - PVFC\left(YC-\Delta i\right)}{PVFC\left(YC\right)} \cdot \frac{1+\Delta i}{2\Delta i}, \tag{9}$$

where:

$PVFC\left(YC+\Delta i\right)$ is liability value calculated based on yield curve increased by Δi.

$PVFC\left(YC-\Delta i\right)$ is liability value calculated based on yield curve decreased by Δi.

Δi is equal to 0.0001

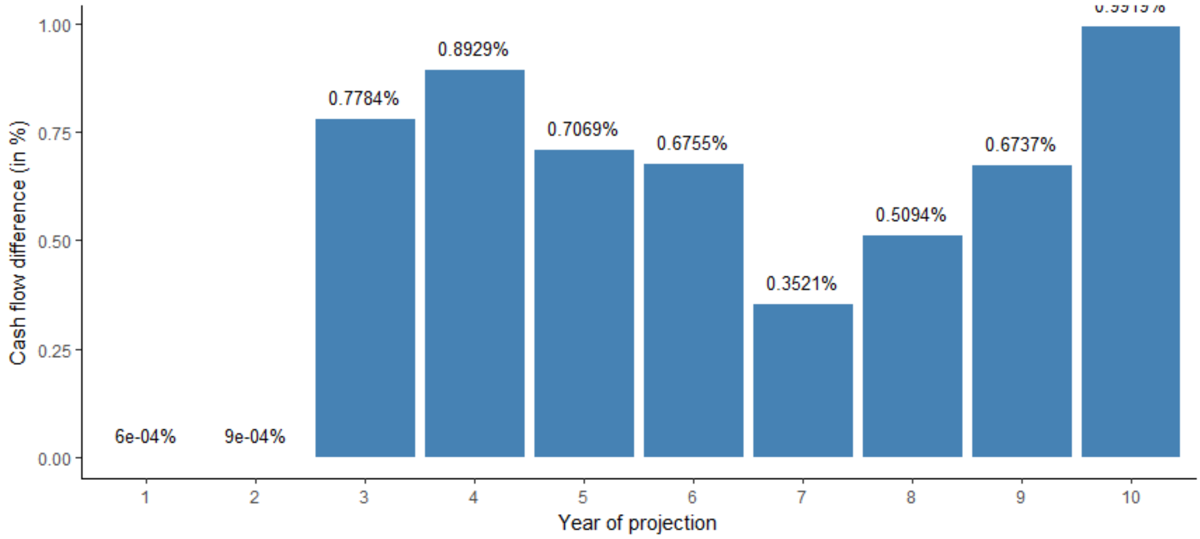
5. Results

In this part, we show whether the faster liability model based on clustering approach replicates the cash flows and duration property of traditional per-policy liability model. This property is important for optimization methods and cooperation with asset model.

The original portfolio used in this paper contains 106 524 modelpoints. Using cluster analyses, we reduced this size to 500 reference modelpoints. The accuracy of the model remains very high with the estimation error lower than 1% but the calculation time significantly decreased. Liability model is now more than 200 times faster. As clustering variables, we used PVFC, PVP, PVPL and individual cash flow up to the 10th year.

The first analysis focuses on whether liability cash flows realized from original portfolio match the cash flows realized from the reference portfolio. In figure 2, we can see that the individual cash flows of the reference portfolio almost perfectly match the cash flows of the original portfolio. The average difference is about 0.5%. The results are as expected because the individual cash flows are used as clustering variables.

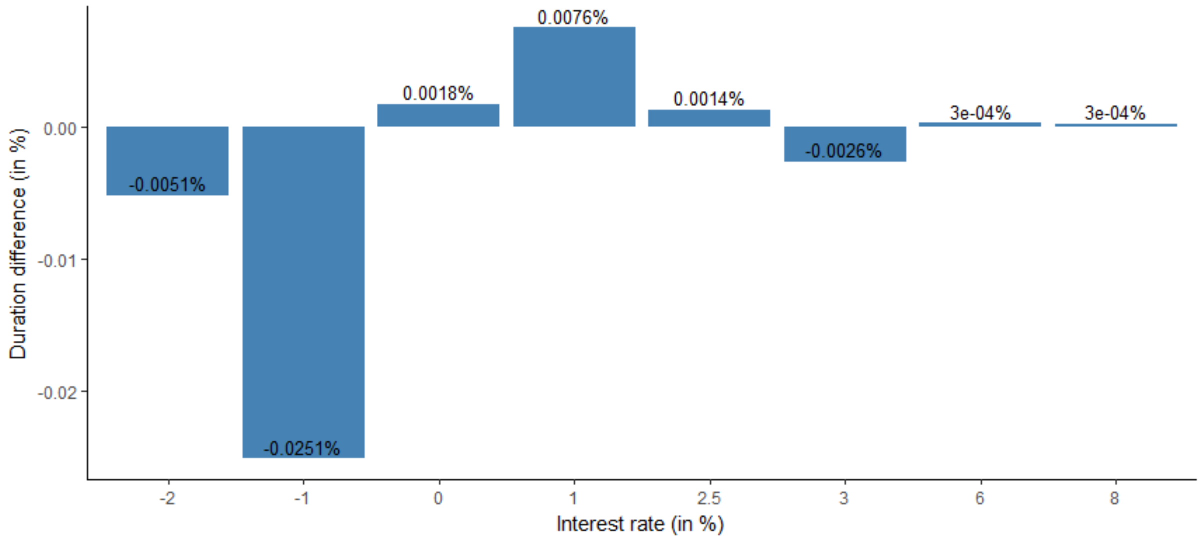
Figure 2: Comparisons of cash flow difference between faster clustering liability model and traditional per-policy model



Source: Authors' work.

The second analysis focuses on duration comparisons of both liability models. The difference between duration comparison is presented in figure 3. The duration of clustering model fits the duration of the traditional per-policy model with very high precision. The error representing average difference is still lower than 0.003%. To provide more complex information, we present results on eight different strategies. Each strategy differs in the expected interest rate of the fund

Figure 3: Comparisons of liability duration difference between faster clustering liability model and traditional per-policy model



Source: Authors' work.

6. Conclusion

Application of cluster analysis as an approximative liability model in life insurance has proven to be a good choice saving time and preserving a high precision of liability estimates. In our previous research, we have presented how to build a clustering approach and tested the stability for certain shocks (parallel change in interest rates or mortality and lapse shock). In this article, we extend this research by basic techniques of asset liability management. We studied whether the faster liability model based on clustering analysis is also a good alternative for the purpose of optimizing investment strategies as cash flow matching or duration matching. In both cases, the clustering model has proven proper for ALM optimization. In the first analysis – cash flow matching, the cash flows from reference portfolio perfectly fitted the cash flows from the original portfolio. The same conclusion was reached with the second analysis of duration comparisons.

The clustering approach keeps preserving high accuracy and significantly reduces the computational time. This property allows testing more investment strategies in a shorter time. Therefore, the clustering approach should be studied in a higher detail in the area of dynamic ALM modelling.

Acknowledgements

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

References

[1] Aas, K., Neef, L., R., Williams, L., Raabe, D. 2018. Interest rate model comparisons for participating products under Solvency II. SCANDINAVIAN ACTUARIAL JOURNAL. iss. 3. pp. 203-224

[2] Balestreri, A., Kent, J. 2011. Dynamic Asset Liability Management A Method for Optimising Investment Strategy. European Actuarial Journal, 7(16).

[3] Cipra, T. 2010. Financial and Insurance Formulas (1st ed.). Berlin Heidelberg: Physica-Verlag.

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

[5] Fernandez, J., L., Ferreiro-Ferreiro, A., M., Garcia-Rodriguez, J., A. 2018. GPU parallel implementation for asset-liability management in insurance companies. Journal of Computational Science. vol. 24. pp. 232-254.

[6] Fojtík, J., Procházka, J., Zimmermann, P. 2017. Method for fast estimation of life insurance liabilities with respect to different investment strategies. The 12th International Days of Statistics and Economics, Prague.

[7] 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.

[8] Gilbert, C. 2016. Asset Liability Management. In IAA Risk Book. Retrieved from https://www.actuaries.org/LIBRARY/Papers/RiskBookChapters/Ch13_Asset_Liability_Management_24Oct2016.pdf

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

[10] Ng, R., Han, J. 2002. CLARANS: a method for clustering objects for spatial data mining. IEEE Transactions on Knowledge and Data Engineering, 14(5), pp.1003-1016.

[11] Shang, D., J., Kuzmenko. V., Uryasev, S. 2018. Cash flow matching with risks controlled by buffered probability of exceedance and conditional value-at-risk. Annals of Operations Research. vol. 260, iss. 1-2. pp. 501-514.

[12] Tsai, C. 2009. The Term Structure of Reserve Durations and the Duration of Aggregate Reserves. Journal of Risk and Insurance. vol. 76, iss. 2, pp. 419 – 441.

[13] Xidonas, P., Hassapis, C., Bouzianis, G., Staikouras, C. 2018. An Integrated Matching-Immunization Model for Bond Portfolio Optimization. Computational Economics. vol. 51. iss. 3. pp. 595-605.