Searching for the right proxy approaches to life

Care has to be taken in using replicating portfolio techniques, least-squares Monte Carlo approaches and curve fitting for estimating the risk capital of a life insurer, as Tigran Kalberer and Zeljko Strkalj explain

"Although this may seem a paradox, all exact science is dominated by the idea of approximation."
Bertrand Russell

The problems and issues discussed in this article are not academic but highly relevant to the insurance industry. While individual insurance companies and consultants will inevitably have their own preferred approaches, it is not in our interests to be openly critical of any of the methods in general because we believe it is not the method, as such, which determines whether a suitable approach has been taken, but the specific application of the method. This article is therefore primarily written to give impartial advice on the choice of a proxy approach and how to use it to the best effect.

With the emergence of Solvency II in Europe and impending Solvency II-style supervision in other parts of the world, insurance companies find themselves in need of more powerful analytical tools than ever before.

An example that illustrates this need is the modelling requirement dictated by Solvency II. For the more complex business, like traditional with-profits products, this essentially requires life insurance companies to apply a nested stochastic approach to calculate the solvency capital requirement (SCR, see figure 1).

For most members of the industry, these types of calculations are complex and computer intensive, which means that actuarial and risk departments are finding it extremely challenging to obtain results with the required precision and within the required timelines.

THE APPROXIMATION PROBLEM

In recent years, a number of different proxy approaches have been introduced to make SCR calculations more manageable. Proxy approaches for SCR calculation purposes are generally based on finding (simpler) functions which approximate a value function[1] and minimise the sum of the squared differences to the given value function over a set of so-called calibration scenarios, potentially under certain constraints. The best-known approaches applied in the insurance industry are:

- Replicating portfolio techniques (RPT)
- Least-squares Monte Carlo (LSMC)
- Curve fitting

The common idea underlying all these approaches is that the valuation of the liabilities is not performed directly but an approximation for the values is used (see figure 2 overleaf). Although the calibration techniques underlying the proxy approaches are different, the application is similar in the sense that they assume that on the set of calibration scenarios the approximating function is a linear combination of "basis functions." While RPT and LSMC essentially ask to solve an optimisation problem and thus find a good fit of basis functions[2], curve fitting looks to determine a basis function[3] which fits selected sensitivities.

Experience has shown that an insurer's decision on which proxy approach it uses is driven by a number of factors, including:

- Nature of the business, guarantees and management actions to be modelled
- Time, capacity and resource constraints
- Supplementary information requirements (e.g. movement analysis)

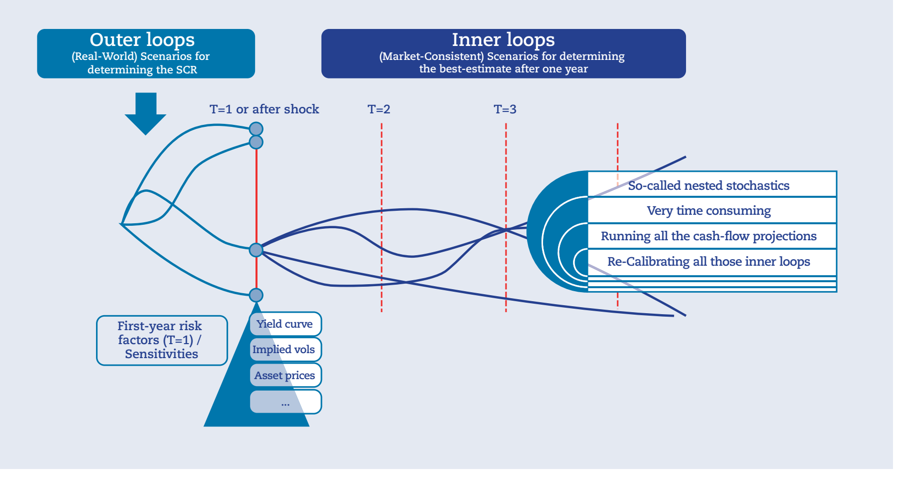


Figure 1: Valuation framework based on the nested stochastic approach

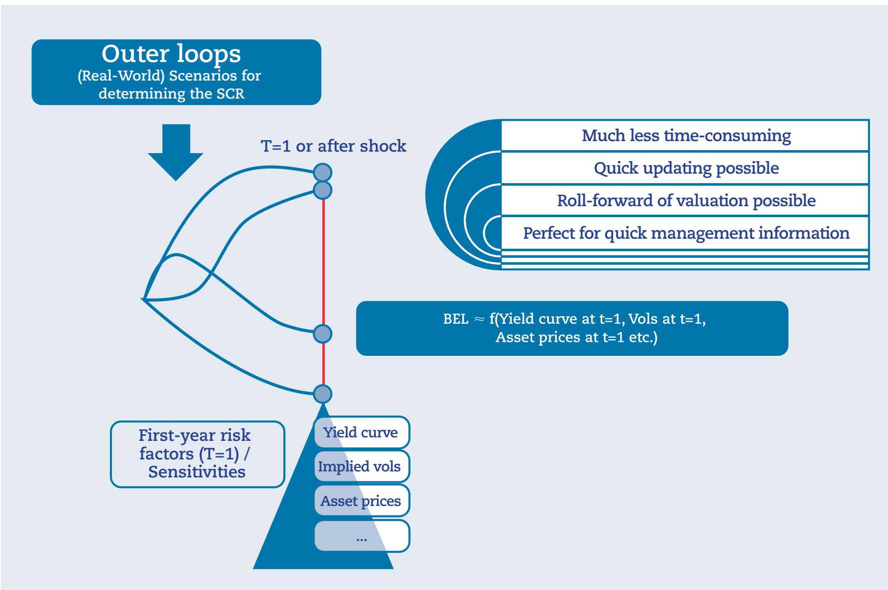


Figure 2: Valuation framework using a proxy approach

- Comprehensibility for senior management
- Ability to satisfy regulatory requirements (e.g. the use test)
- Scope to validate the results (e.g. in terms of confidence intervals for the SCR).

Clearly, a reliable proxy approach involving relatively little effort is what companies are aiming for. However, enhanced operational and governance as well as regulatory requirements have recently put more emphasis on documentation (i.e. to describe each step of the entire calibration process in an auditable way) and validation (to demonstrate thoroughly the adequacy of a proxy), and actuarial and risk departments are challenged to ensure overall compliance with those.

In the last couple of years, we have observed several of these approaches being applied by various insurers, and in quite a few cases these approaches turned out not to be robust in an insurance context and were applied without sufficient accuracy. This led, for example, to SCR results which fluctuated over time for no apparent reason – a situation which is certainly not acceptable. Regulators have also noticed these issues and are thus highlighting a general problem of how to ensure that the proxy approach applied is really adequate.

DON'T HIDE BEHIND THE COMPLEXITIES

There are various issues we have observed in applying the above-mentioned approaches. A mathematically rigorous description of the approaches, which then allows one to identify the issues and solutions, can be found in a series of papers by Tigran Kalberer, published in *Der Aktuar*, the quarterly magazine of the German Actuarial Association. This article summarises the main results of this series in less technical language.

The three most important issues emerging from the proxy approaches mentioned above show the dilemma over why the approaches used in the industry sometimes are not as robust as they should be.

1. Make sure the requirements for proper approximation are met

(a) The coverage issue – don't mess with important risks

In general, a simple-calibration scenario generator has preference over a more complicated one unless this leads to material distortions of the results ("as sophisticated as necessary" and "as simple as possible").

When calibration scenarios are considered and their adequacy analysed it is important to address their coverage. It is obvious that the proxy approach cannot lead to a reliable result if substantial risk factors are not reflected within the calibration scenarios. In such cases, one can potentially get a good fit of the approximation, but this does not mean anything if significant risks are neglected. Here are some examples of risk factors which tend to be addressed in a much less than perfect way:

- The yield curve shape risk is often left out, as most economic scenario generators are based on a low number of interest rate risk factors and thus do not reflect the possibility of yield curve shape changes adequately.
- Credit spread, migration and default risks are often neglected in economic scenarios or are reflected in a way which is too crude. This is highly questionable, as such scenarios do not allow one to properly include the time value of options and guarantees. Additionally, such an approach can overstate required capital substantially, as the risk-absorbing capacity of policyholder participation is not reflected.
- Calibration scenarios should also cover non-economic risk factors, such as longevity or other risks which are material to an insurer's business.

(b) The co-linearity issue – get the maths right

The issue of coverage leads us to a broader area that needs to be properly addressed when approximations via linear combinations of basis functions are considered. This area is also very much related to the choice of calibration scenarios, as the realisations of basis functions are determined by the calibration scenarios.

Our observations have shown that realisations of basis functions used by insurance companies are often highly co-linear[4]. This can lead to the following problems:

- The basis functions can produce null vectors. In insurance terms, this means that linear combinations of the basis functions exist, which, evaluated on the set of calibration scenarios, produce near zero cash flows. Examples show that these null vectors could be significantly different from zero in some areas not well covered by the calibration simulations.

This can potentially lead to SCR estimates which deviate significantly from the correct results.

- The SCR results can vary wildly between two different optimisations, e.g. between two different times when the optimisation is performed (quarterly reporting).

We have observed that the RPT approach in particular, if not applied very carefully, suffers from these severe shortcomings[5]. Ultimately, applying an approach with highly co-linear basis functions is by no means robust (both in terms of time as well as in the tails).

If you have a portfolio of basis functions (or candidate assets) which adds up to a near-zero cash flow over all calibration scenarios (this is what high co-linearity means), then obviously large amounts can be added to or subtracted from this portfolio without changing the optimisation result much. This means that in some cases, just by numerical coincidence, an arbitrarily high amount of such a portfolio will be added, or not, giving rise to non-robust proxies.

What such an approach also implies is that the results are not only useless for asset-liability management purposes but even dangerous from a SCR calculation point of view.

To overcome this issue and to ensure a robust process, co-linear functions need to be avoided and the entire approach should be built upon linearly independent basis functions. This implies that the calibration scenarios should have a specific structure. They should consist of independent "post-shock" risk-factor realisations, followed by (few) market-consistent continuations (more on that in the next section).

(c) The portfolio issue - nobody told you to lie back and do nothing

Usually basis functions used for RPT applied in the industry are functions depending on one risk factor only. However, the cashflows which need to be approximated are, in general, functions of whole sets of risk factors. The cashflows can, for example, take the form of put options on a portfolio of assets.

Now, such cashflows are typically not very well approximated by a linear combination of put options on each asset individually. In order to reflect such cash flows, it is necessary to include basis functions which are functions of a whole set of risk factors. Experience shows that it is quite difficult to find out on which exact sets (of risk factors) these basis functions should depend.

A good way to investigate such phenomena is to perform a LSMC fit using polynomials with cross-terms (in the risk factors). If the coefficients of the mixed terms are high, then this is a clear sign of a portfolio issue.

What such an approach also implies is that the results are not only useless for asset-liability management purposes but even dangerous from a SCR calculation point of view.

2. Focus on what you want to approximate

Although the actual goal of any approximation is to find a simpler function (and thus to evaluate faster) which is close to the (real) function[6], surprisingly this goal in many cases is not appropriately defined.

While insurance companies use their proxies to calculate SCR results, i.e. calculate value-at-risk (VaR), Tail-VaR or similar types of tail results, they effectively only ensure a good approximation on average, potentially well around the median. Real cases from insurance portfolios show that this does not inform the quality of the approximation in a given quantile, which is ultimately of interest. This issue is addressed in some cases by adding adverse scenarios, which requires management judgement.

It is thus important that the appropriateness and quality of the SCR estimation coming from the approximation is ensured.

3. Have an efficient and stable approximation process in place

SCR calculation processes are generally highly complex in nature and include various sub-process dependencies (e.g. asset model, cashflow model, assumptions, scenarios, aggregation, etc.) that determine how fast and reliable the process ultimately can be.

Integrating the proxy approaches discussed so far within a SCR calculation framework offers a variety of potential benefits to improve and make the SCR calculations more manageable. But it also increases the complexity of the entire framework significantly. Usually such processes have been developed organically and piecemeal, and thus can result in additional complexities, slowness and lack of robustness of the overall process. Finally sub-processes underlying proxy approaches are also often highly manual and combined with "expert" judgement[7].

It is therefore of utmost importance to make the approximation process as efficient and stable as possible, which can be quite challenging. We are convinced that this challenge can only be addressed by a full industrialisation of the existing cash-flow-modelling processes within a company and by potentially enhancing the current scope of risks modelled (e.g. including credit risk modelling in the cash flow models).

WHAT NEEDS TO BE DONE?

1. Address the issues and make the (unavoidable) process adjustments

In case you were hoping that we would now magically produce a wonderful idea that solves all the issues discussed so far, then dream on. Let's face it, addressing the issues around the coverage, co-linearity and portfolio issues is required in order to be in a position to use adequate proxies. This essentially means:

- Appropriate coverage of first-year risk factors
- Basis functions with sufficient non-co-linearity (on the set of calibration scenarios), preferably orthogonal
- Approximation functions with sufficient dependency on the risk-factors at t=1

But this is only the bare minimum. In addition to what has been stated so far, we believe that certain aspects of current proxy approaches have to be adjusted substantially, regardless of the proxy approach used. We mention the major ones in order to present the full scope of potentially unavoidable changes.

(a) Specific calibration scenarios

The appropriate approximation process will require the production of a special set of calibration scenarios which is entirely different from the set of scenarios typically used. Recent investigations have shown that sufficient accuracy for the approximation can be achieved by using a high number of outer scenarios[8] ("outer loops") and a rather low number of inner scenarios ("inner loops").

The calibration scenarios will in particular require the re-calibration of the economic scenario generator for each outer loop, i.e. one essentially needs market-consistent calibration scenarios for a large number of "shocks" or sensitivities. In this context it is important that the risk factors are independent.

Experience shows that for a not-too-complex portfolio, typically a scenario budget of 5,000 outer scenarios and 10 inner scenarios is sufficient. Sometimes for very important outer scenarios (e.g. for the most interesting 10% of outer scenarios) one needs 100 inner scenarios instead of 10.

This seems a lot, but there is now no need to determine sensitivities and thus (real) nested stochastic with all its advantages regarding robustness and simplicity becomes feasible. But of course the number of scenarios required depends on the circumstances, like the number of risk factors, the complexity of the portfolio, etc.

(b) Automated sensitivity generation

The calibration scenarios need to be run through the cashflow model of the insurer. In this context it is of utmost importance that the current process of producing sensitivities (i.e. outer loops), typically involving manual adjustments, is automated as the number of runs required will be large. Fortunately, if the calibration scenarios and the production process fulfil the requirements mentioned so far, the approximation approach works with very few "inner" valuation scenarios ("inner loops"), i.e. the amount of runs required will still be massive but will not explode.

2. A recipe for the optimisation approach

We have so far focused the discussion on making sure that the requirements for proper approximation are met. Obviously this is a very important question. But what happens now after we have chosen the (correct) calibration scenarios and we have ensured that we have the appropriate basis functions for the approximation? Which linear combination of basis function is the "best" one for the approximation? Here is a recipe for an optimisation algorithm:

1) Run the calibration scenarios through your cash-flow model.
2) Output all relevant variables you want to fit (e.g. liability cash flows).
3) Discount the cash flows to t=1 (resulting in present values for each outer loop x_s).[9]
4) Evaluate for each outer loop the value of each basis function B_i.
5) Use the L2 norm (on the set of outer loops) and determine the linear combination of basis functions ("proxy") which best approximate $PV(x_s)$, i.e. we need to choose a set of coefficients P_i such that the least square error of the residuals is minimised.

The resulting proxy (e.g. for the liability value) is the one we are looking for. We can now use this proxy and evaluate it on real-world outer loops[10] and consequently determine the SCR in a quick way.

3. The approximation is adequate!

We have mentioned that proxy approaches tend to avoid ensuring the validity of the proxy approach applied. In this context it is important to emphasise that a proxy might be perfect within the universe of basis functions/candidate assets but that the SCR might be still mis-estimated.

Using another set of scenarios, the "error-estimation scenarios," which look like the calibration scenarios described above but where the distribution of the outer-loop risk factors is the one used for the determination of the SCR, actually allows estimation of the error incurred in terms of SCR, when a proxy approach is used.

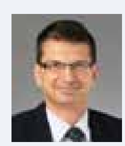
Thus, the quality of the SCR estimation using the proxy can be measured. It can be shown that the estimation error implied by using the proxy can be split into two parts:

(a) A component which is due to the asymmetry of the errors (i.e. the differences between the proxy and the value of the liabilities). This component is zero if the errors are symmetric.

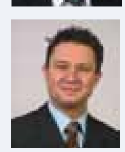
(b) A component due to the slope, or steepness, of the proxy around the SCR. This component is small if the distribution of the errors is sufficiently narrow or the slope is sufficiently flat.

Now, having the estimation error will be key for validation, i.e. in convincing regulators and internal governance functions that the approximation is not just a "good-to-have" tool to speed up the process of generating results but actually an "important-to-have" instrument that ensures a fast and robust reporting process.

And, in the end, that's exactly what life insurers are looking for.



Tigran Kalberer
Principal
Milliman (Zurich)
tigran.kalberer@milliman.com



Zeljko Strkalj
Senior Consultant
Milliman (Zurich)
zeljko.strkalj@milliman.com

[1] Value function in this context is the value of assets minus liabilities (at time 1 or 0) for each possible scenario realisation.

[2] In the context of RPT, basis functions are typically assets, while for LSMC the basis functions are polynomials in the first-year risk factors.

[3] Curve fitting also uses basis functions which are polynomials in the risk factors.

[4] In mathematical terms, up to a point where the matrix of their correlations has (a) a rank lower than the dimension of this matrix or (b) has very small eigenvalues.

[5] The RPT approach suffers typically also from a lack of path-dependency at t=1. To fix this situation we need an approximation function with sufficient dependency on the risk factors at t=1.

[6] Here the word "close" means in a pre-defined and sensible way. For calibration purposes, the Euclidean norm L2 is shown to be appropriate.

[7] E.g. (a) which candidate assets for RPT or (b) which sensitivities for curve fitting or (c) what to do in case of overfitting, etc.

[8] Depending on the approach chosen, these are effectively sensitivity scenarios or first-year scenarios.

[9] The discounting to t=0 is required if the approach is based on sensitivity scenarios instead of first-year scenarios.

[10] Creating outer loops corresponding to the real-world distribution that a company assumes to govern the behaviour of the risk factors is a topic of its own and is far from easy.