MOODY'S ANALYTICS

B&H RESEARCH

MARCH 2013

One-year projection of run-off conditional tail expectation (CTE) reserves

Steven Morrison PhD
Laura Tadrowski PhD
Craig Turnbull FIA
Moody's Analytics Research

Contact Us

Americas
+1.212.553.1658
clientservices@moodys.com

Europe
+44.20.7772.5454
clientservices.emea@moodys.com

Asia (Excluding Japan)
+85 2 2916 1121
clientservices.asia@moodys.co

Japan
+81 3 5408 4100
clientservices.japan@moodys.com

Overview

This paper discusses whether the quantitative techniques that have been successfully applied to the *nested stochastic* challenge arising in 1-year VaR in insurance economic capital can also be applied to another nested stochastic problem: that of making a one-year projection of run-off CTE reserve requirements. This approach to reserving and capital requirements has been a popular alternative to market-consistent VaR in North America and Asia-Pacific. But the development of capital projection techniques has not, to our knowledge, kept apace with the work done in recent years in the projection of market-consistent liability values. This paper aims to go to some way to redressing that balance.

The analysis produced in this paper strongly suggests that the most sophisticated technique that has been applied to the 1-year VaR nested stochastic problem – the *Least Squares Monte Carlo* method – can produce similarly powerful results in the projection of a run-off CTE reserve measure. Interestingly, the paper's analysis also suggests that the 'traditional' curve fitting methods whose limitations have been strongly highlighted in 1-year VaR implementation will be found to be similarly limited in their effectiveness in the context of run-off CTE reserve projection.

Contents

1. Introduction

Insurance firms' risk and capital assessment frameworks increasingly have computational needs that theoretically require *nested* stochastic simulations (sometimes referred to as *stochastic-on-stochastic*). These requirements can arise in a variety of contexts:

1. The calculation of a 1-year Value-at-Risk (VaR) capital requirement that is based on the changes in a market-consistent balance sheet. (In theory this requires a set of market-consistent scenarios for each of thousands of one-year real-world scenarios.)
2. The forward projection of the above 1-year VaR. (Note a stochastic projection of the 1-year VaR capital requirements could in theory require a 'double nesting' or stochastic-on-stochastic-on-stochastic approach—each real-world projection scenario would require a set of one-year real-world scenarios and each one-year real-world scenario would require a set of market-consistent scenarios)
3. The calculation of a run-off reserve or capital requirement that incorporates future management actions such as dynamic hedging, the modelling of which requires market-consistent valuations and sensitivity estimates. (Each multi-year real-world scenario would require several market-consistent scenario sets at several future real-world time-steps.)
4. The forward projection of a run-off reserve or capital requirement. Again, the projection of the capital requirement could demand a stochastic-on-stochastic-on-stochastic approach.

It is clear from the above examples that the computational demands associated with the estimation and projection of capital requirements are exploding. This is particularly challenging when we consider that these exercises are unlikely to be 'fire-and-forget' processes that run once a year. In a regulatory context, supervisors are likely to require various forms of sensitivity testing and investigation. Similarly, management use of the capital model should inevitably give rise to a multitude of frequent 'what-if' capital analyses that require timely responses.

In general, global insurance firms have opted to try to avoid the computational burden of the nested stochastic approach by developing quantitative solutions that can robustly estimate some of the above risk and capital metrics more efficiently. This work has been most developed in the context of 1. above, both in the context of the economic capital models that have been implemented globally by insurance groups, and also in the regulatory context of the development of Solvency II Internal Models[1]. The multi-year projection of capital requirements (i.e. numbers 2 and 4 above) is a topic of active interest in the context of ORSA solvency capital projection, and will be the subject of further research reports to be published in the near future.

In this paper, we focus on developing a robust approach to the one-year projection of run-off CTE (Conditional Tail Expectation) reserves that avoids the need for nested stochastic processes. This is a topic that may be of particular interest to North American and Asian insurance firms, where the use of run-off CTE reserve measures is quite prevalent. In particular, the paper aims to address the following:

» Can the quantitative methods that have been developed as efficient alternatives to nested simulation in the context of estimation of the 1-year VaR of a market-consistent balance sheet also be used in the stochastic projection of run-off CTE reserves?

» If so, how should these quantitative methods be adapted in order to provide a robust solution in this different context? And can it be sufficiently effective?

» Demonstrate answers to the above questions using an illustrative case study.

The paper is structured as follows:

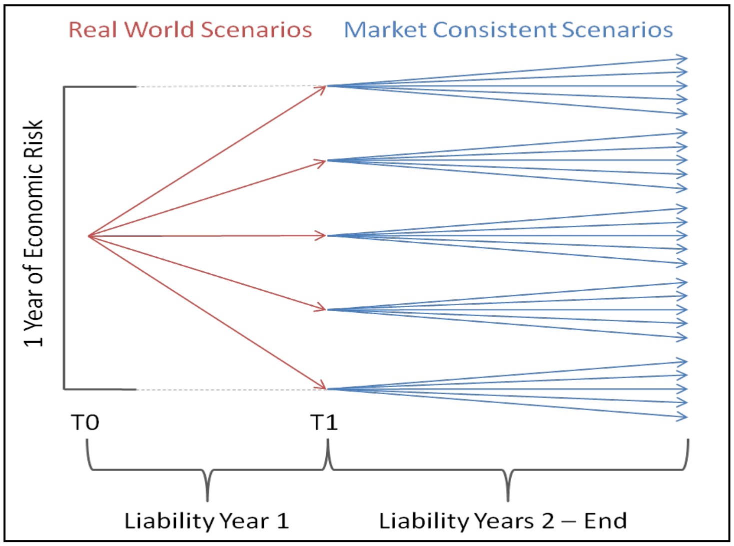
Section 2 provides a brief overview of the quantitative approaches used in the efficient implementation of 1-year VaR models for insurance firms' economic capital;
Section 3 introduces the illustrative case study that will be used to apply these methods in run-off reserve projection;
Section 4 presents the technical fitting process, fitting results and validation results for the run-off projection function;
Section 5 shows how the quantitative method can deliver an estimate of the end-year run-off reserve requirement, and how this can be used to calculate a 1-year VaR economic capital requirement where the capital requirement is defined as a VaR of the run-off reserve.
Section 6 sets out our conclusions.

[1] See KPMG's "Economic Capital Modeling in the Insurance Industry", August 2012.

2. A brief overview of proxy function fitting

Over the last few years, a significant amount of quantitative work has been done on the efficient implementation of 1-year VaR of market-consistent balance sheets (i.e. an implementation that avoids a nested stochastic calculation). Figure 1 illustrates the 1-year VaR nested stochastic problem.

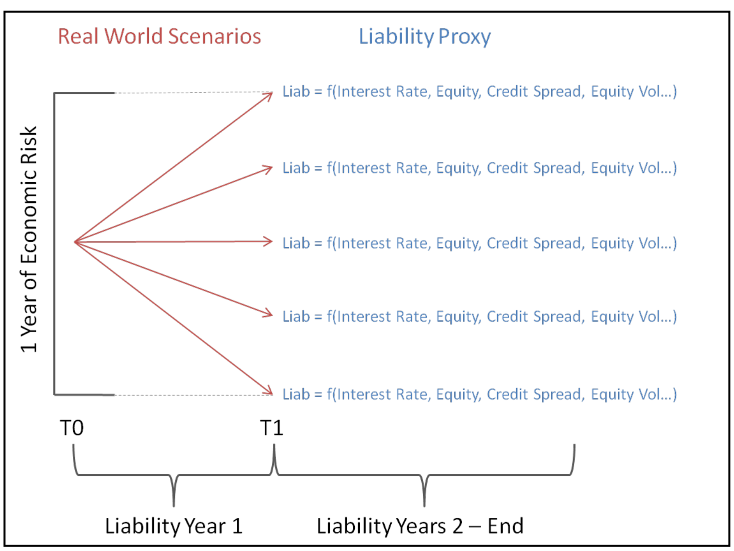
Figure 1 1-year VaR of a market-consistent balance sheet: a nested stochastic problem



Note that this nested stochastic requirement only arises for liabilities that require Monte-Carlo simulations to be used in their market-consistent valuation. Generally, this need arises for liabilities with complex, path-dependent guarantees such as variable annuities or European with-profits business.

Virtually all firms have chosen to implement the above VaR calculation using a quantitative short-cut to the full nested stochastic approach. The most common strategy has been to develop what are known as *liability proxy functions*. These are formulas that estimate the change in the market-consistent value of the liabilities that arises over the 1-year projection horizon as a function of 1-year risk factor outcomes. This approach is summarised diagrammatically in Figure 2 below.

Figure 2 1-year VaR of a market-consistent balance sheet: liability proxy function solution



Broadly speaking, these liability proxy functions can take one of two forms:

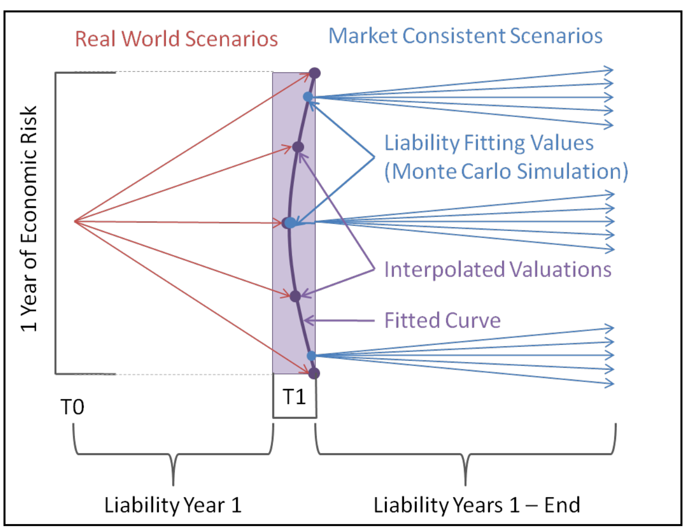
» The market value of a portfolio of assets that has been constructed to produce cash flow behavior that replicates that of the liabilities. This is known as the *Replicating Portfolio* approach.

» A more general polynomial function of the liability values' risk factors, where the parameters of the function are fitted to a number of stressed liability valuation results. This is known as the *curve fitting* approach.

The Replicating Portfolio (RP) approach is highly appealing as it has an immediate 'real-life' interpretation, and it can provide intuitive insight into the economic behavior of liabilities and how they can be priced and hedged. However, implementation of the RP approach at insurance firms has highlighted that it has some inherent and important limitations, particularly in the context of replicating the complex, path-dependent guarantees that create the nested stochastic problem that we are trying to solve. This is because the RP method is most useful when the assets that are included in the replicating portfolio are instruments for which a traded market price can be observed. And it is not usually possible to find market prices for assets that have the path-dependency and complexity of the liabilities that create the nested stochastic requirement in the first place. This is particularly problematic in the context of Solvency II Internal Models, where high standards of model quality and validation are demanded. As a result of these limitations, there has been a steady shift away from the use of the replicating portfolio technique for 1-year VaR, and we anticipate this trend will continue. By 2012, more than two-thirds of firms implementing a liability proxy function framework were using a curve fitting approach instead of a replicating portfolio approach to their 1-year VaR economic capital modelling implementation[2].

Curve fitting describes a very general approach based on fitting a polynomial function of the liabilities' risk factors to some stressed liability valuation results. This is an area where best practice implementation methods are evolving. In 'traditional' curve fitting, the modeler specifies a curve of the form of a general polynomial function with *n* parameters, and then *n* accurate re-valuations of the liability are run using re-calibrated market-consistent simulation sets. This simulation process is represented below in Figure 3.

Figure 3 'Traditional' curve fitting



'Traditional' curve fitting suffers from a number of significant limitations:

» The modeller must decide which parameters are to feature in the polynomial, and, by extension, which parameters of the general polynomial function can safely be assumed to be zero. This is problematic as there is a limit to how reliable the modeller's intuition can be in determining if, say, the co-efficient of the interest rate cubic term is material to the one-year change in the market-consistent liability valuation, or if the quadratic credit spread / interest rate cross-term is likely to be

[2] Page 16 of KPMG's "Economic Capital Modeling in the Insurance Industry", August 2012.

» important. So, for complex liabilities, the modeller will usually first obtain a function that does not perform well in out-of-sample validation testing, and will then have to guess at which parameters should be added to the function.

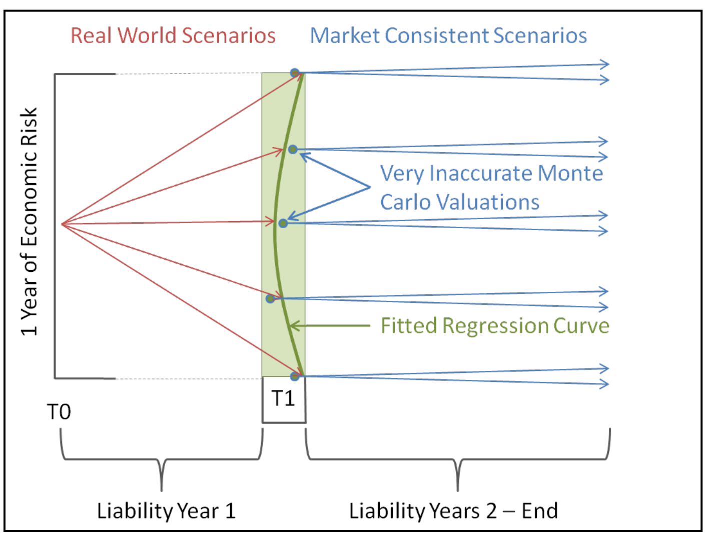
» This approach is based on making some 'exact' liability-revaluations. But even when several thousand simulations are used in the stressed liability valuations, some statistical errors will remain. These estimation errors can be amplified when the fitting points are then used to interpolate and extrapolate across the risk factor range.

» Simulation efficiency does not scale well with the number of risk factors and possible relevant parameters. For example, even when fitting only up to quadratic terms, the number of parameters that need to be fitted using full liability re-valuations increases from 3 to 15 when the number of risk factors increases from 1 to 4. Firms will typically use between 4 and 12 risk factors to describe changes in most liabilities so it is easy for the number of full stochastic runs required by curve fitting to start resembling the computational requirements of implementing the full nested stochastic approach.

Curve fitting can be implemented using a more sophisticated fitting method that overcomes these three key limitations. In our experience, the most successful such method uses the Least Squares Monte Carlo method (LSMC). LSMC turns the logic of traditional curve fitting on its head: instead of performing the smallest number of re-valuations possible and performing them with the maximum degree of accuracy permitted by the total scenario budget, the LSMC method performs a much larger number of re-valuations, each with a very low degree of accuracy. Note 'accuracy' here refers to sampling error from the valuation estimate that has been produced using simulation. LSMC works because these sampling errors are independent across the thousands of re-valuations that can be run, and even a 1-simulation estimate of a market-consistent liability valuation is unbiased.

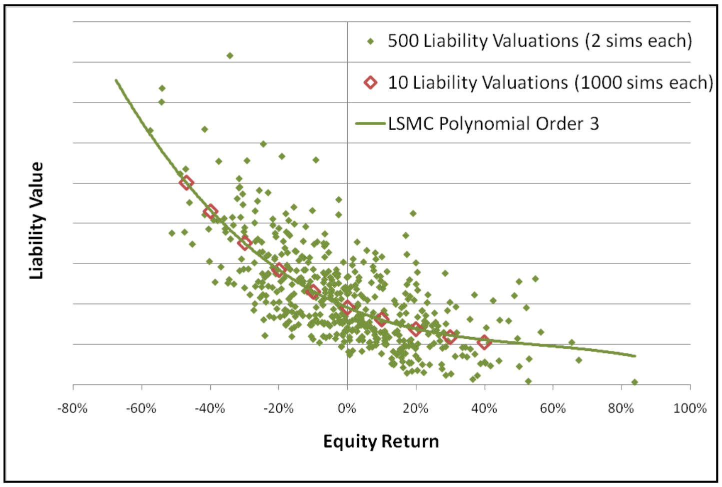
So, for example, instead of running 20 re-valuations using 5,000 market-consistent simulations in each re-valuation, we could run 50,000 re-valuations, with only 2 market-consistent simulations being used in each re-valuation. In both cases, a total of 100,000 *fitting scenarios* have been used. This simulation process is set out in Figure 4.

Figure 4 Curve fitting using Least Squares Monte Carlo



Clearly, any of the re-valuations produced in the LSMC fitting process is useless in isolation. But, taken together, the independence of their errors means that they can be 'cancelled out' using standard regression techniques. This is the key insight to why LSMC works better: we cut corners in the 'inner' simulations because the independence of the many errors can be regressed away; this allows us to devote more of the simulation budget to the 'outer' simulations. And because we can now use a huge number of outer scenarios, we no longer need to make guesses about which parameters are the important ones – step-wise regression techniques allow us to systematically explore the entire space of candidate models, adding terms only when they are judged to have a statistically significant impact on the quality of fit. Instead of amplifying the impact of relatively small valuation errors (traditional curve fitting), this technique smooths away (several thousand independent, unbiased) large valuation errors. The regression process is illustrated below.

Figure 5 LSMC regression fitting: Illustrative example of an equity put option



The illustrative example shown in Figure 5 considers a simple equity put option example where the only risk factor driving the one-year change in valuation is the one-year equity return. 500 liability re-valuations have been made over a wide range of one-year equity returns, and 2 market-consistent simulations (one antithetic pair) have been used to value the liability in each of these 500 stresses. The green dots show the 500 re-valuations that have been produced by these 500 2-simulation valuations. Naturally, there is a large amount of sampling error in each of these valuations, and this is illustrated by the wide dispersion of the green dots. Standard regression techniques have then been used to derive a polynomial function from these 500 valuation observations. In this example, a cubic function has been fitted. 10 out-of-sample validation runs have then been implemented, each with 1000 simulations being used in the valuations. These are plotted in the red diamonds. You can see that the fitted function corresponds closely to the validation valuation results.

The particular example above could also have been treated well using replicating portfolios or traditional curve fitting – the intention of the example is simply to illustrate the LSMC implementation process. The key benefit of LSMC arises in the context of the complex, path-dependent guarantees that are found in the many forms of long-term savings products that are provided by the global life insurance sector. We introduce the analysis of such a product in the next section.

3. Introducing the case study

This section introduces the case study that will be used to test the application of curve fitting techniques in the projection of run-off measures of reserves or capital requirements. The case study will be loosely based on the guarantees that are found in Universal Life products in North America and Asia Pacific. In summary, it will take the following form:

» Assume an annual return of max (Fund return – 1.5%, 2%) is credited to the policy account.

» The underlying investment return is invested in a diversified portfolio of US corporate bonds. The bonds are assumed to be invested with a credit mix of 70% A-rated and 30% BBB-rated, and with a term of 8 years. The bonds' credit rating and term is assumed to be re-balanced annually.

» The policyholder is assumed to exit the policy after ten years, and will receive the value of the credited account at that point.

» No allowance is made for tax, mortality, expenses or lapses.

Before we consider estimating the one-year change in the run-off capital requirement that can occur for the above product, we should first calculate its starting run-off capital requirement. This can be calculated using a 10-year real-world scenario set from the B&H Economic Scenario Generator (ESG). This analysis uses our standard US real-world calibration as at end-2012 and a 5,000 simulation scenario set. Figure 6 below shows the cumulative probability distribution for the 10-year annualised total fund return.

Figure 6 Cumulative probability distribution for 10-year annualised investment fund return

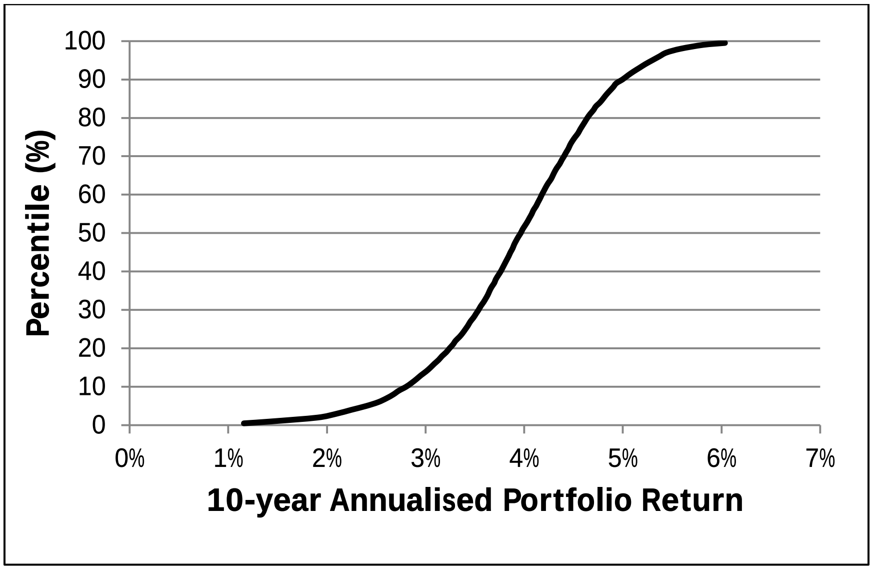
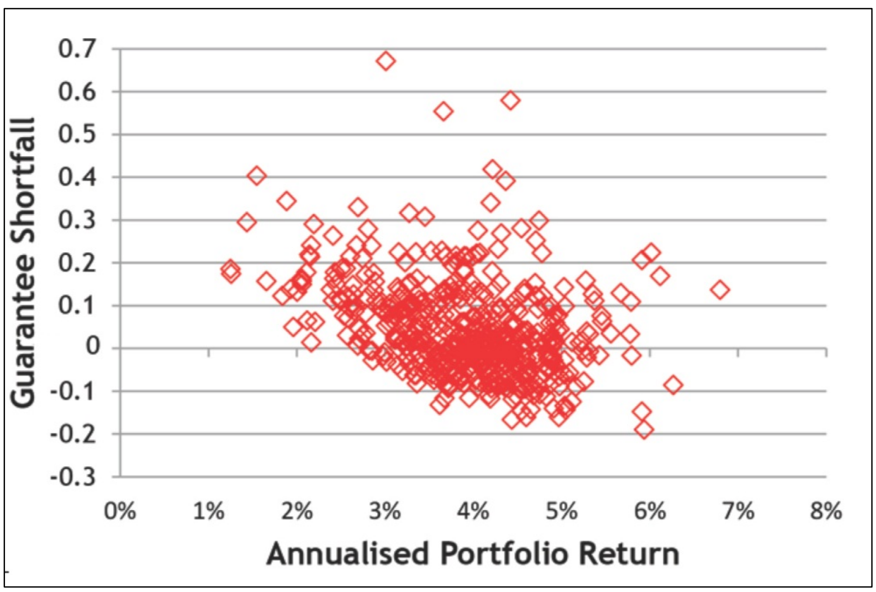


Figure 7 plots the simulated guarantee shortfalls together with the associated simulated 10-year return. Here, the guarantee shortfall is defined as the time-10 credited account value less the time-10 fund value.

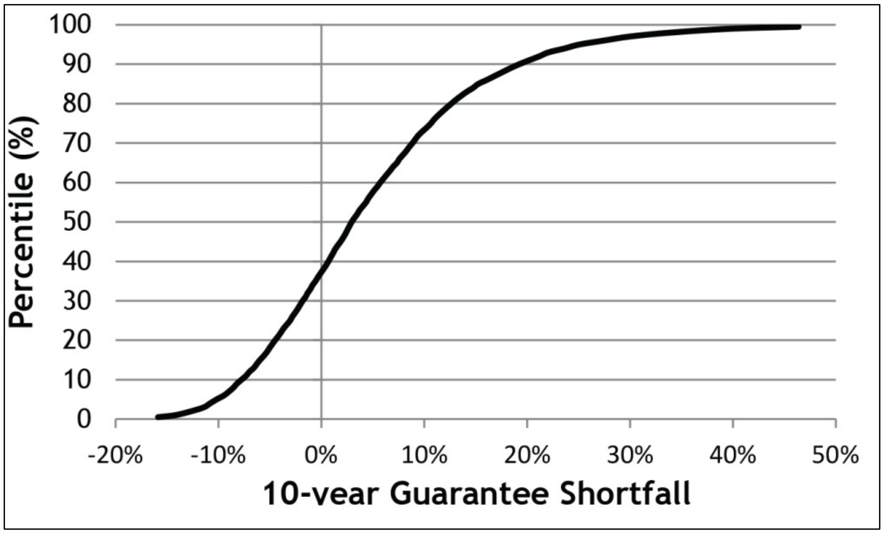
Figure 7 10-year annualised fund return and the guarantee shortfall



A negative relationship between the portfolio return earned over the 10-year period and the resultant guarantee shortfall can be observed in the above chart. This negative relationship is intuitive – the 2% annual minimum investment return will clearly be more costly during periods of poorly performing underlying investment returns. However, the relationship between the annualised 10-yerar return and the guarantee shortfall is not direct. This is because the policy credited rate is applied annually. So the difference between the policy account and the fund value after 10 years is a function of the set of 10 annual returns and doesn't only depend on the 10-year total return. The chart highlights that this guarantee is highly path-dependent.

Finally, Figure 8 shows the cumulative probability distribution for the 10-year guarantee shortfall.

Figure 8 Cumulative probability distribution for the 10-year guarantee shortfall



The run-off capital requirement can now be obtained by calculating the present value of the guarantee shortfall (the underlying 10-year fund return is used as the discount factor) and then calculating the required Conditional Tail Expectation. This produces a CTE (70) and CTE (90) of **12.1%** and **18.5%** of the starting fund value respectively.

4. Fitting the run-off reserve proxy function

Having obtained the current run-off capital requirements of the product, we now consider how the proxy methods outlined in section 2 can be used to produce a function to estimate how the run-off capital requirement can change over a one-year horizon.

In the case study there are three risk factors in the model that will drive changes in the run-off reserve: two of which drive changes in the risk-free yield curve, and one that drives changes in credit spreads. So the challenge now is to develop a polynomial function that can provide an estimate of size of the CTE (70) run-off reserve that will be required after one year as a function of these three risk factors.

The Replicating Portfolio (RP) method is specifically geared to providing a market-consistent valuation estimate –knowing the RP for the liabilities is of no direct use in providing an estimate of how the CTE (70) run-off capital requirement behaves. But the curve fitting approach is more general: the concept of fitting a general polynomial function to a series of stressed observations can apply to run-off CTE reserves as much as it does to market-consistent valuations. This section develops proxy functions for the end-year CTE (70) using different curve fitting approaches:

- Section 4.1 uses 'traditional' curve fitting method as described in section 2;
- Section 4.2 uses Least Squares Monte Carlo-based approaches; and
- Section 4.3 provides some out-of-sample validation test results for each of the curve fitting methods that are implemented in 4.1 and 4.2.

4.1. Fitting the function using traditional curve fitting

In the traditional curve fitting approach a polynomial function of *n* parameters is specified; *n* re-calculations of the metric of interest are then computed using *n* large simulation runs; and the polynomial function's parameters are then fitted to these results.

For the case study described in section 3, we consider a 20-parameter polynomial function. 20 sensitivity tests are therefore run and the CTE (70) run-off reserve is calculated using 5,000 simulations in each test. So a total of 20 x 5,000 = 100,000 scenarios are used in producing the fitting data.

The polynomial function is assumed to include all terms of up to a cubic power, including cross-terms. Mathematically, the function can be written as:

$$\widetilde{X} = f(dz_1, \dots dz_3) = a + \sum_{i_1,\dots i_3=1}^{3} a_{i_1 i_2 i_3} (dz_1)^{i_1} (dz_2)^{i_2} (dz_3)^{i_3}$$

This function's twenty parameters can be summarised as:

» Three terms that are a function of the first interest rate factor only (a linear, a quadratic and a cubic term)

» Three terms that are a function of the second interest rate factor only (a linear, a quadratic and a cubic term)

» Three terms that are a function of the credit spread factor only (a linear, a quadratic and a cubic term)

» Three terms that are coefficients of cross-terms between the first and second interest rate factors

» Three terms that are coefficients of cross-terms between the first interest rate factor and the credit spread factor

» Three terms that are coefficients of cross-terms between the second interest rate factor and the credit spread factor

» One term that is the coefficient of the cross-term between the combined behavior of all three risk factors

» One term that is the function's intercept value, i.e. the estimate we obtain for the end-year CTE (70) run-off reserve when all risk factors take their expected end-year values.

The 20 stress tests that are used to fit these 20 parameters are produced using the following process:

» We define the *range* of one-year risk factor values over which the function is to be fitted. There is a trade-off here— the range should be wide enough to capture all the feasible values that the risk factors can take over the next year; but if the range is too big, we will be fitting to results that are outliers to the area in which we want to estimate the reserve's behavior. In this case study, we select the range for each risk factor based on estimated extreme percentiles of the real world distribution.

» We then specify 20 stresses within the three-dimensional hyper-cube that is defined by the three risk factors and their specified ranges. These 20 risk factor stresses are uniformly distributed within the hyper-cube. (Technically, this is done using SOBOL numbers[3]).

A CTE (70) run-off reserve is then produced using 5,000 simulations under each of the 20 sensitivity tests. These results are then used to fit the 20-parameter function. Note that under this method the function will fit exactly to each of the 20 results produced using simulation. So the primary approach to determining the quality of fit of the function will be by out-of-sample validation tests. We will discuss this later in section 4.3.

The two charts below provide some insight into the behavior of the fitted function. The data in the two charts has been produced by running a set of 1-year real-world scenarios for interest rate and credit spread risks and then using the fitted function to re-calculate the end-year CTE (70) run-off reserve arising in each real-world scenario. Figure 9 plots the estimated run-off reserves against the simulated end-year 10-year risk-free spot rates, and Figure 10 plots the reserves against the simulated end-year 10-year A-rated credit spread.

[3] See Section 2 in "Low Discrepancy Numbers and Their Use within the ESG", David Redfern, Feb2010 http://www.barrhibb.com/knowledge_base/article/low_discrepancy_numbers_and_their_use_within_the_esg/

Figure 9 Figure 9: Simulated CTE (70) run-off reserves and 10-year risk-free interest rates (Traditional curve fitting)

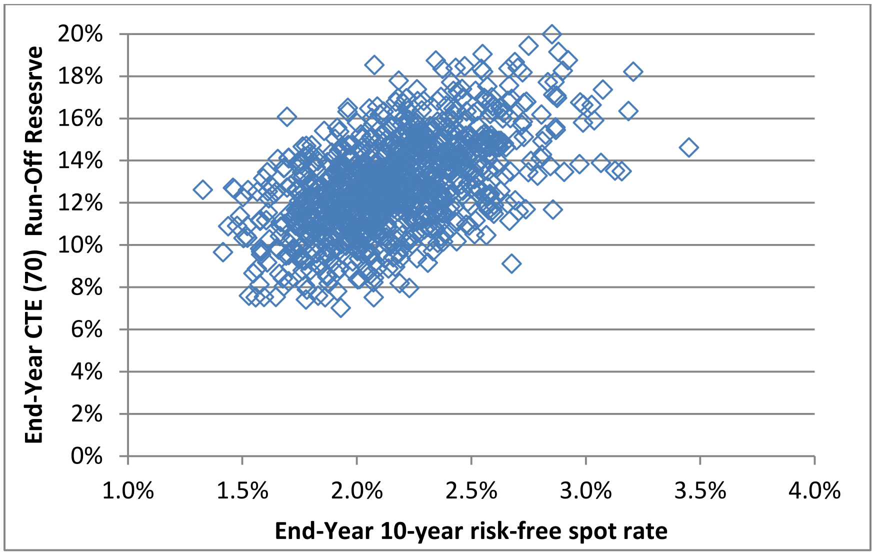
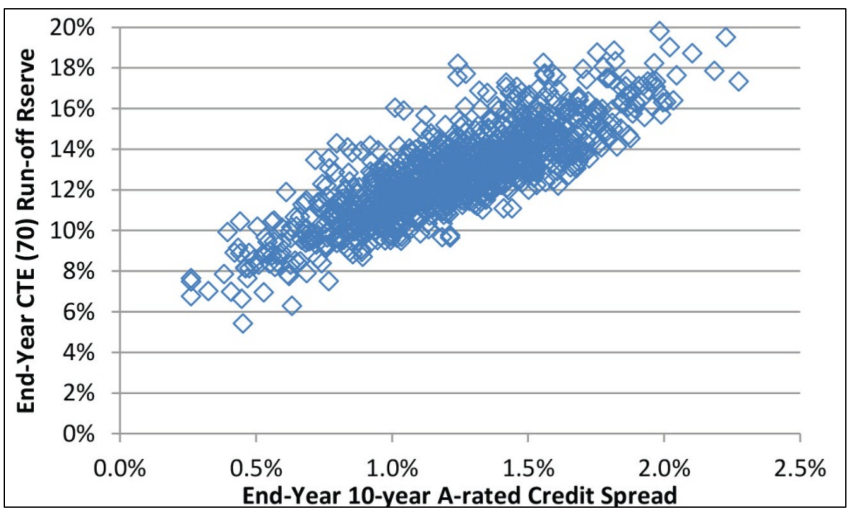


Figure 10 Simulated CTE (70) run-off reserves and 10-year A-rated credit spreads (Traditional curve fitting)



These charts give some insight into the behavior of the reserve that is implied by the sensitivity test results and the fitting process. In particular, the proxy function suggests that increases in credit spreads and increases in interest rates will result in a higher run-off reserve being required at the end of the year. This general dependency is fairly intuitive as these are scenarios in which the fund value has earned less than the 2% guaranteed minimum policy rate, hence moving the guarantee 'in-the-money'.

4.2. Fitting the function using LSMC-based curve fitting methods

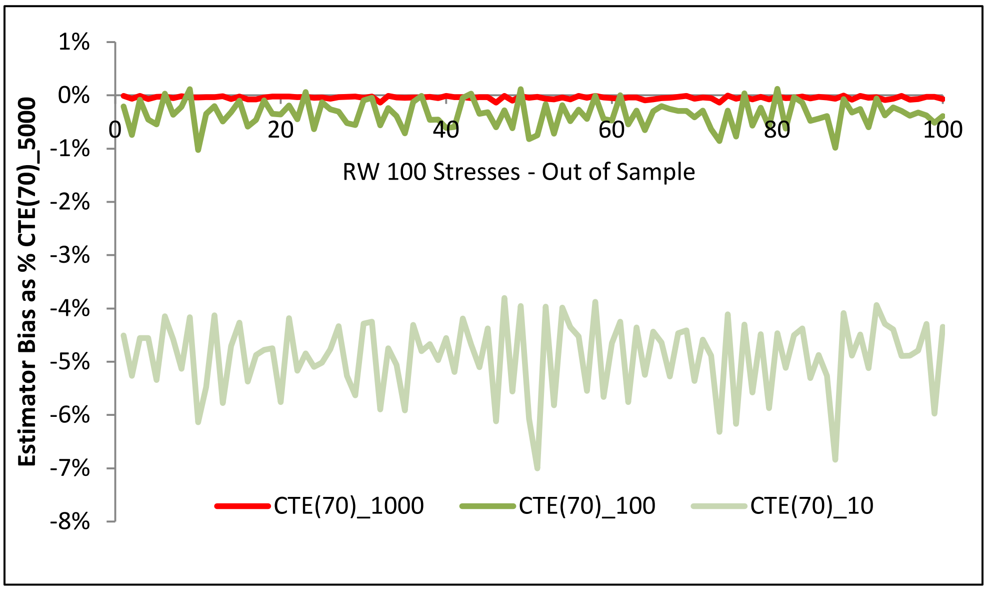
We now consider how the Least Squares Monte Carlo approach to curve fitting as described in section 2 can be applied to fitting a proxy function for the case study's end-year CTE (70) reserve. To analyse the improvement in fitting accuracy and modelling efficiency that can be obtained from LSMC methods, we retain the total fitting scenario budget of 100,000 scenarios, and consider different outer / inner scenario combinations that use this budget.

When using LSMC in proxy functions for market-consistent valuations, we generally find that the accuracy of the fitted function is maximized by maximizing the number of outer simulations and hence running only a single simulation in each the inner simulation (or a single pair of simulations when using antithetic variables)[4]. The application of LSMC techniques to the fitting of a CTE (70) proxy function will be more complicated – it is intuitively clear that a CTE (70) estimate cannot be obtained from one or two simulations. It is also known that small sample sizes produce a *biased* estimate for a conditional tail expectation[5]. As per the discussion of section 2, the LSMC technique requires each estimate of the metric in question to be *unbiased*. So the implementation of LSMC techniques in CTE proxy functions must address two related questions:

1. What is the minimum number of simulations that can be used to reliably produce an unbiased estimate of the CTE? (Note this number may vary for different types of distribution and different CTE points.)
2. Does fitting with this number of inner simulations still result in accuracy and efficiency benefits relative to the traditional curve fitting approach?

To analyse the first of these questions, we considered 100 1-year real-world scenarios and calculated the CTE (70) run-off reserve in each of these 100 scenarios with 5,000 simulations for each reserve calculation. We then re-calculated each of these 100 CTE (70) reserves using 1000 simulations; and then again with 100 simulations; and finally once more using only 10 simulations. We then considered the proportional estimation error that is produced from the reduction in the number of simulations used in the CTE (70) calculation. Figure 11 plots the estimation errors produced for the reserves in each of the 100 real-world one-year scenarios as a function of the different numbers of simulations used in the CTE (70) reserve calculation.

Figure 11 CTE (70) estimation errors observed in 100 different scenarios



Remember that in the context of LSMC-based approaches, we are not concerned about the 'accuracy' of any single CTE (70) estimate (as we will smooth out random noise in estimation errors using regression); we are only concerned with whether there is any systematic *bias* in the estimates.

Figure 11 strongly suggests there is no material systematic bias in the CTE estimates that are produced using 1000 simulations. It also strongly suggests that there *is* significant systematic bias in the CTE (70) estimates when they are estimated using 10 simulations. So this suggests that well-performing proxy functions for the CTE reserve will not be obtained if as few as 10 inner simulations are used in the fitting process. How many more inner simulations should be used? The 100-simulation results in Figure 11 suggest there is some bias in the CTE estimate when 100 simulations are used, but the bias is around -0.35% of the reserve. To put this in context, the average starting CTE (70) reserve was 12.7%, and 0.35% of 12.7% is 0.045% of the starting fund value. It could be argued that this bias is of an immaterial size, particularly in the context of the other assumptions that are made in the broader modelling process. So, we anticipate that using 100 inner scenarios will provide a well-behaved proxy function. But, of course, there is only one way to know for sure: fit the function and then perform out-of-sample validations.

[4] "Monte Carlo Simulation Approaches to the Valuation & Risk Management of Unit-Linked Insurance Products with Unit Guarantees", M.J.Cathcart PhD Thesis, University of Heriot-Watt & Barrie & Hibbert, July 2012

[5] "Variance of the CTE Estimator", Manistre & Hancock, North American Actuarial Journal, 2005

We now produce a number of new proxy function fits, using the following outer/inner simulation mixes:

» 100 outers x 1000 inners

» 1000 outers x 100 inners

» 10,000 outers x 10 inners. (Whilst the above analysis implies this configuration will not produce a good proxy function, it is included for comparison).

In each of these three cases, the selection of the 'outer' scenarios that will be used in the fitting process are produced using the same algorithm as in the traditional curve fitting case described in section 4.1. That is, the scenarios are uniformly spread across the 3-dimensional risk factor hyper-cube.

In the following section, the performance of these proxy functions will be compared with the one produced in section 4.1 using the traditional curve fitting approach (recall this used 20 outers and 5,000 inners). Note that in all cases a total of 100,000 fitting scenarios are used.

4.3. Out-of-sample validation tests

Sections 4.1 and 4.2 generated four different proxy functions for the end-year CTE (70) reserve. Each used 100,000 fitting scenarios, but essentially differed in terms of how these scenarios were distributed within the risk factor fitting space. In this section, the quality of fit of each of these proxy functions is assessed using out-of-sample validation tests.

The out-of-sample validation testing process can be summarised as follows:

1. 125 1-year scenarios for the behavior of the three risk factors are generated. The scenarios we use here have been selected because they are adverse and stressful for the product as this is the area of the distribution that we are most interested in validating from a risk and capital measurement perspective[6]. But any scenarios (other than the scenarios directly used in the fitting process) could be used in validation – they could be a random sample from a real-world model or they could be 'hand-selected'.

2. The end-year CTE (70) run-off reserve that arises in each of these 1-year scenarios is calculated by running 5,000 simulations under each of the 125 scenarios. This gives the 125 CTE (70) run-off results that are used as the 'right answers' in the validation analysis.

3. The four proxy functions are then used to generate four different run-off reserve estimates in each of the 125 validation scenarios.

4. The run-off reserve estimates produced by the proxy functions are compared with the validation results produced using 5,000 simulations. This comparison can take a few forms – visual comparison using scatterplots, and some summary statistics of the average difference between the proxy function results and the 5,000-simulation results. Both are shown below.

Figures 12, 13, 14 and 15 present the scatterplots of the proxy functions' reserve estimates versus the 5,000-simulation results for each of the four proxy functions in turn.

[6] Identification of these "stressful" scenarios involved generating 10,000 real world scenarios for the risk drivers, estimating the CTE (70)s using the proxy functions and ordering these scenarios from worst to best in all four scenario allocations. The 99th percentile was taken from each scenario allocation and these were combined into a 400 scenario stress superset. Our proxy functions produced a strong agreement on the location of this tail and the 400-element stress superset was reduced to a 125 element stress superset once degeneracies were removed.

Figure 12: Scatterplot of end-year CTE (70) reserve in 125 validation scenarios: 'actual reserve' versus proxy function (20 outers and 5,000 inners)

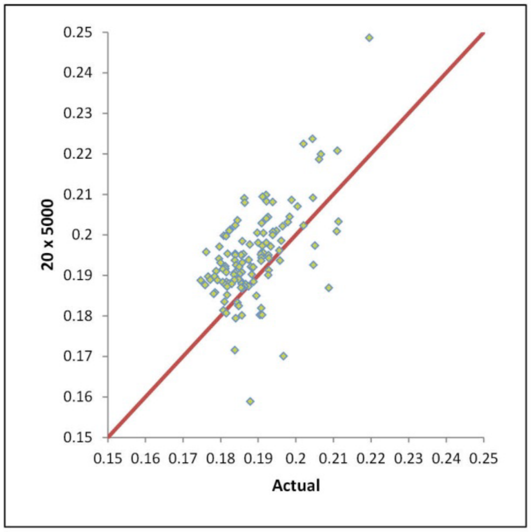


Figure 13: Scatterplot of end-year CTE (70) reserve in 125 validation scenarios: 'actual reserve' versus proxy function (100 outers and 1,000 inners)

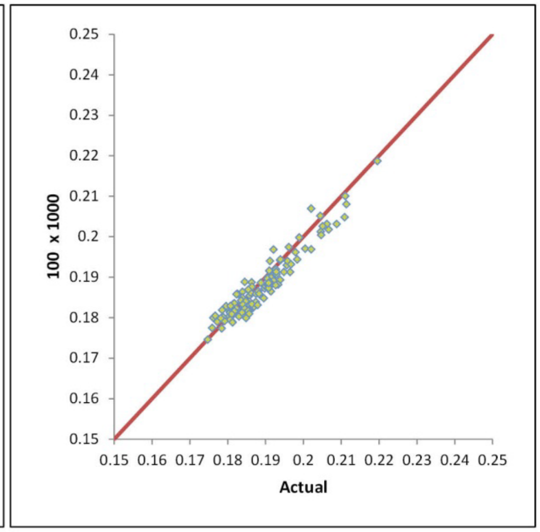


Figure 14: Scatterplot of end-year CTE (70) reserve in 125 validation scenarios: 'actual reserve' versus proxy function (1,000 outers and 100 inners)

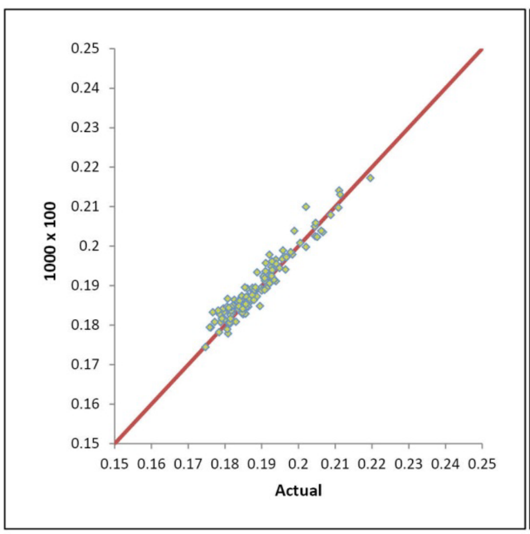
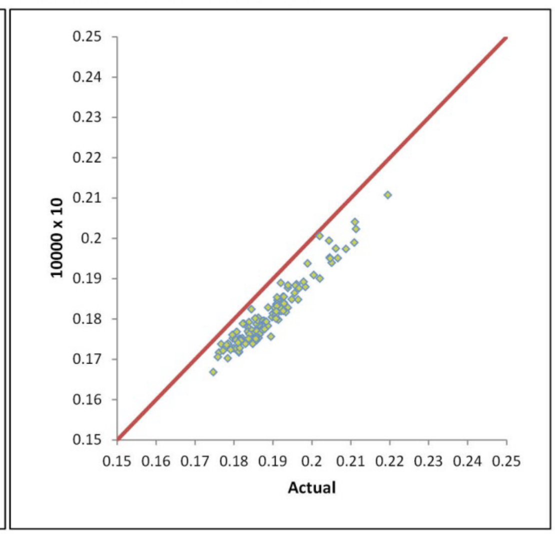


Figure 15: Scatterplot of end-year CTE (70) reserve in 125 validation scenarios: 'actual reserve' versus proxy function (10,000 outers and 10 inners)

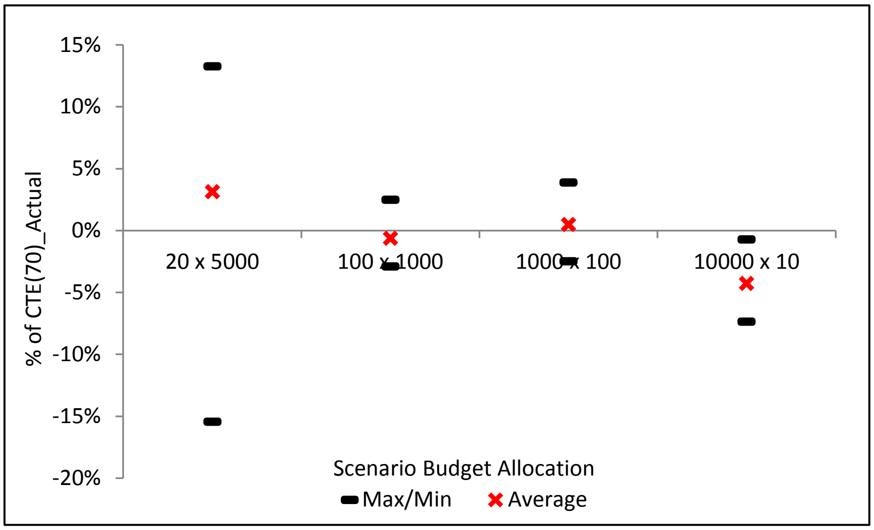


We will shortly present some summary statistics of the information in the above four charts, but before doing so there are already two key points that can be observed by the naked eye:

1. The (100 outer x 1000 inner) and (1000 outer x 100 inner) LSMC cases appear to produce significantly more accurate proxy function fits than that produced by the traditional curve fitting approach. This conforms to our intuition and experience from applying LSMC techniques to the fitting of proxy functions for market-consistent liability valuations.
2. However, the (10,000 outer x 10 inner) LSMC case produces a fitted function that exhibits significant bias in the run-off reserve estimate. This is consistent with our expectations based on the analysis produced in section 4.2 (particularly Figure 11).

Figure 16 summarises the results of validation of the four functions under the 125 validation scenarios.

Figure 16 Summary of validation results



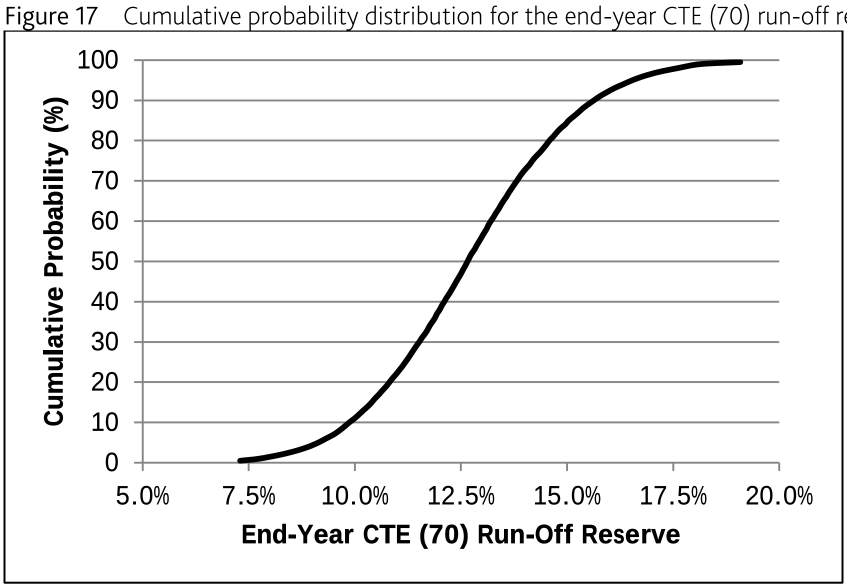
The above chart again highlights that the LSMC approach produces significantly more accurate and reliable proxy function results than traditional curve fitting with the same total fitting scenario budget. The magnitude of the estimation errors produced by the traditional curve fitting approach highlight why firms have found its implementation in Solvency II Internal Models to be problematic for complex, path-dependent guarantees. The analysis suggests that running with the 100 inner simulation size produces the optimal fitting results for the case study.

5. Using the proxy function to calculate 1-year VaR capital

Having obtained a validated proxy function, we can consider the probability distribution for the end-year CTE (70) run-off reserve that is implied by it. This requires us to also make some assumptions about the 1-year real-world probability distributions that drive the case study's three risk factors (two interest rate factors and one credit risk factor).

In this example, the standard B&H end-2012 real-world calibration is used to make the joint one-year interest rate and credit projection. 5,000 one-year simulations were produced and the proxy function was used to produce the run-off reserve estimate in each of these simulations. The cumulative probability distribution for the end-year CTE (70) run-off reserve is shown below in Figure 17.

Figure 17 Cumulative probability distribution for the end-year CTE (70) run-off reserve



The above probability distribution may be useful for a range of applications. It may be provide important insights for capital planning, especially under stress testing. It could also be used to provide an estimate of current economic capital. For example, an insurance firm could define its economic capital requirement as a 1-year 99.5% Value-at-Risk (VaR) of the CTE (70) run-off reserve.

This modelling estimates the 1-year 99.5% VaR of the CTE (70) run-off reserve at **19.1%** of the starting fund value. You may recall from section 2 that the current CTE (70) run-off reserve was estimated at 12.1%. The 1-year 99.5% VaR of the CTE (70) is estimated to be very similar to the current CTE (90) – in section 2 this was calculated to be 18.5% of the starting fund value. It is interesting to note that these two alternative economic capital definitions (the 1 -year 99.5% VaR of CTE (70) and the 'time-0' CTE (90)) provide such similar results, though we should resist making a sweeping generalisation of this result based on the observation from a single example.

6. Conclusions

In theory, the evaluation and projection of the measures of capital requirements arising in global insurance increasingly require nested stochastic simulation methods. In recent years insurance firms implementing 1-year VaR capital measures for economic capital assessment or in Solvency II Internal Models have usually opted to side-step these nested stochastic demands by applying quantitative techniques that can accurately estimate the capital requirement more efficiently. This has generally been done by developing liability proxy functions that provide formulaic estimates for how market-consistent liability valuations will change over the one-year projection horizon. Experience in this area has been mixed and the best practice quantitative methodologies have evolved significantly in recent years in order to produce robust and accurate functions that can pass rigorous model validation criteria.

This paper has discussed whether the quantitative techniques that have been successfully applied to the 1-year VaR nested stochastic problem can also be applied to another nested stochastic problem: that of making a one-year projection of run-off CTE reserve requirements. This approach to reserving and capital requirements has been a popular alternative to market-consistent VaR in North America and Asia-Pacific. But the development of capital projection techniques has not, to our knowledge, kept apace with the work done in recent years in the projection of market-consistent liability values. This paper aims to go to some way to redressing that balance.

The analysis produced in the paper strongly suggests that the most sophisticated technique that has been applied to the 1-year VaR nested stochastic problem – the Least Squares Monte Carlo method – can produce similarly powerful results in the projection of a run-off CTE reserve measure. Naturally, there are some specific implementation considerations that differ in the two cases, but these can be considered simply as configuration choices within an identical fitting and validation framework. Interestingly, the paper's analysis also suggests that the 'traditional' curve fitting methods whose limitations have been strongly highlighted in 1-year VaR implementation will be found to be similarly limited in their effectiveness in the context of run-off CTE reserve projection.