



SESSIONAL PAPER

Long-term stochastic risk models: the sixth generation of modern actuarial models?

Bill Curry*

*Correspondence to: Bill Curry, LV= County Gates, Bournemouth, BH1 2NF, UK. E-mail: billcurry986@hotmail.com

Abstract

This paper discusses the use of modelling techniques for the purpose of risk management within life insurers. The key theme of the paper is that life insurance is long-term business and carries with it long-term risks, yet much of modern actuarial risk management is focussed on short-term modelling approaches. These typically include the use of copula simulation models within a 1-year Value-at-Risk (VaR) framework. The paper discusses the limitations inherent within the techniques currently used in the UK and discusses how the focus of the next generation of actuarial models may be on long-term stochastic projections. The scope of the paper includes a discussion of how existing techniques, together with new approaches, may be used to develop such models and the benefits this can bring. The paper concludes with a practical example of how a long-term stochastic risk model may be implemented.

Keywords: Solvency II; ORSA; Risk management; Stochastic modelling

Details

Section 1 provides an introduction with a brief history of modern actuarial models.

Section 2 describes the approaches and techniques currently widely used in the UK and how these fit in with the Solvency II regulatory framework.

Section 3 discusses in detail the limitations of the use of a 1-year Value-at-Risk framework.

Section 4 discusses how plan projections, which are carried out in a traditional deterministic manner may differ from the average (mean) or median stochastically produced results.

Section 5 considers the use of copula simulation models and how these are not necessarily suitable for long-term projections.

Section 6 introduces the use of a long-term stochastic projection model and describes the benefits this may bring.

Section 7 discusses in detail how the components of a long-term stochastic projection model may be constructed.

Section 8 introduces a demonstration example, and the results of which are set out in section 8.

Section 10 summarises the key conclusions of the paper.

This paper is written with a focus on UK life insurance firms under the Solvency II regime. A number of the techniques discussed within are likely to be applicable in a wider context.

This paper is intended for the UK or European Life actuaries who are interested in:

- Risk management.
- Modelling techniques.
- Business planning.
- ORSA.

It is expected that the reader will have a working knowledge of the key aspects of Solvency II.

Disclaimer

The views expressed in this paper are those of the author only. The information and expressions of opinion contained in this paper are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations.

1. Introduction

1.1. Purpose

- 1.1.1. This section gives a brief history of modern actuarial models. It then gives an outline of the regulatory requirements under Solvency II and the typical approaches used by the UK firms.

1.2. History of Modern Actuarial Models

- 1.2.1. Actuarial work primarily concerns the management of risk. The oldest form of models may be considered to date back to ancient times whereby societies were formed to meet the cost of funeral expenses¹. These societies could be described as using a basic form of the model.
- 1.2.2. An important development in actuarial work was the use of the life table (Graunt, 1662), which was later used by Edmund Halley to determine the appropriate price for an annuity. The Equitable Life was formed in 1762 and became the first company to use such techniques to set premium rates in a robust mathematical fashion.

1.3. Generation One – Commutation Function Valuation

- 1.3.1. By the middle of the 19th century², an established technique for modelling actuarial liabilities was the net premium valuation. The technique made use of commutation functions to give an approach that was for its time both practical and scientific.
- 1.3.2. Although itself a very old idea, the net premium approach with commutation functions remained as an established technique for many years and only started to become obsolete through the widespread use of computer technology within firms. For this reason, the use of commutation functions is may be considered to be the first generation of modern actuarial models.

1.4. Generation Two – Discounted Cashflow Models

- 1.4.1. Since the 1980s, the use of modern computational power has started to become widespread throughout the insurance industry. This has had a transformative effect on the work of actuaries, making manual calculation work obsolete and allowing for a more sophisticated generation of model.
- 1.4.2. The second generation of actuarial model is may be considered to be the discounted cashflow model. Such a model is based on the projection of all future cashflows for a contract in order that these can be discounted to the start date such that it can be valued. This form of the model is quite intuitive in concept, yet would have been prohibitively

¹Such societies existed in ancient Greece and Rome.

²The actual date or inventor of the technique is not clear.

onerous when carried out through manual calculation. By contrast, spreadsheets or simple coding can comprehensively overcome such problems.

- 1.4.3. The key benefits of the discounted cashflow models were that they permitted more sophisticated modelling features to be allowed for. They also provided actuaries with a greater insight into the underlying nature of the liabilities they worked with.
- 1.4.4. Unlike commutation function-based models, discounted cashflow models are very much in use today as they remain the best valuation approach for many products. Rather than replacing the discounted cashflow models, further generations of models have instead been used alongside them or within them.

1.5. Generation Three – Stochastic Simulation Risk-Neutral Model

- 1.5.1. In 2000, the original life insurance company – the Equitable Life, ran into severe problems and was forced to close to new business. There were a number of contributory reasons for its demise, but perhaps, the most significant was in the approach to the valuation of Guaranteed Annuity Options (GAOs).
- 1.5.2. In keeping with industry practice, the Equitable used a deterministic approach to the valuation of these guarantees. This approach meant that, while interest rates remained above the rates implicit within the guarantees, no value would be attributed to them and so no reserve was held³.
- 1.5.3. The Equitable Life case was a key factor in a step change in the UK regulatory reporting regime in 2004. This saw the introduction of “Realistic Balance Sheet” reporting under which options and guarantees were required to be valued using market-consistent techniques. The requirement to use such techniques came with the widespread use of the next generation of the actuarial model – the stochastic simulation risk-neutral model.
- 1.5.4. The market-consistent valuation of options and guarantees in practice normally requires the use of simulation models. The models use a large number of (e.g. 2,000) risk-neutral simulations in order to produce a liability valuation consistent with observed market prices. The move from a single projection to instead of thousands of projections represented a significant increase in the requirements of valuation systems. As a result, such models typically took a substantial time to run when first introduced.

1.6. Generation Four – Risk-Based Capital Modelling with the Correlation Matrix

- 1.6.1. Also in 2004, equally as important as the introduction of realistic balance sheets with market-consistent valuation of guarantees was the introduction of the Individual Capital Adequacy Standards (ICAS) framework. The ICAS framework brought with it the introduction of risk-based capital modelling techniques.
- 1.6.2. Under the ICAS framework, firms were required to calculate a capital requirement based on a 1-year Value-at-Risk (VaR) approach at the 99.5th percentile. In order to achieve this, for the first time, firms needed to be able to carry out stress valuations of their assets and liabilities and to aggregate the results in order to produce the required capital.
- 1.6.3. The approach used as a standard was that individual stress runs would be carried out for each risk at 1-in-200 level, and these would then be aggregated using a correlation matrix calculation (also known as a variance–covariance formula). Such calculations have a

³In fact, due to an artificial financial reinsurance arrangement, the Equitable didn't even hold reserves in respect of guarantees that were in the money.

number of limitations⁴. As a result, firms would typically apply an approximation technique (e.g. the “medium bang” or “Single Equivalent Scenario”) approach to mitigate these.

- 1.6.4. These correlation matrix calculation models represented a new generation of models. Rather than replacing existing models, they were implemented effectively above them such that the correlation matrix calculation models used stress valuations of the underlying discounted cashflow and risk-neutral simulation models.

1.7. Generation Five – Copula Simulation Models

- 1.7.1. The most recent step change in actuarial modelling occurred through the introduction of copula simulation modelling techniques around 2008. The use of such techniques allowed actuarial risk modelling to move beyond the limitation of the correlation matrix approach to more comprehensive risk-based capital calculations.
- 1.7.2. The approach typically used is to generate real-world simulations of an insurer’s risks over a 1-year time frame, estimate the losses within the simulations and use these to generate key results (such as the capital at 1-in-200 level). The models normally run very high numbers of simulations (e.g. 1 million) and as such require sophisticated computation power to complete. Even with such power, the full valuation of assets and liabilities for such a large number of simulations is prohibitively onerous, particularly where those liabilities may themselves normally be calculated using risk-neutral simulations. For this reason with the introduction of copula simulation models came the use of proxy models.
- 1.7.3. Proxy models are a simplified approximation to the use of full asset and liability models. They may take a number of forms (e.g. replicating formula, Least Squares Monte Carlo (LSMC)) but essentially the purpose of the proxy model is to provide a fast estimate of how assets and liabilities may respond to different risk events. Given the calibration of a suitable proxy function, a firm may perform high numbers of valuations in a reasonable time frame.

2. Current Regulatory Framework and Market Practice

2.1. Purpose

- 2.1.1. This section contains a discussion of the modelling requirements under the Solvency II regime and how these are reflected in the UK market practice.

2.2. Solvency II Requirements

- 2.2.1. Under the current Solvency II regime, the main modelling requirements may be considered separately under Pillar 1 and Pillar 2.
- 2.2.2. SII Pillar 1 concerns the amount of regulatory capital required to be held. This may be calculated using either a firm’s own Internal Model⁵ or through the Solvency II Standard Formula. In either case, the capital is calculated using a 1-year VaR measure at a

⁴The calculation is correct without approximation under the assumptions that:

- The joint distribution of risks is elliptical.
- Individual losses are a linear function of the risk movements.
- Joint risk losses are equal to the sum of individual risk losses.

⁵Subject to regulatory approval.

confidence level of 99.5% (1-in-200). Where the Standard Formula is used, the capital must be calculated using a correlation matrix calculation prescribed within the regulations.

- 2.2.3. Within the SII balance sheet, firms must also allow for a Risk Margin (included as a reduction in firms' Own Funds). The purpose of the Risk Margin is to make an allowance for the value of non-hedgeable risks. It is calculated using a cost of capital approach based on the projected capital of a firm's non-hedgeable risks.
- 2.2.4. A key part of SII Pillar 2 is the concept of the Own Risk and Solvency Assessment (ORSA). The ORSA is intended to form a firm's own view of the risks it faces. A firm's ORSA may typically contain business plan projections that could contain sensitivity analysis to show how plans may be affected by adverse future conditions.
- 2.2.5. More recently, there has been a focus in the UK on the concept of 1-in-X risk buffers used in setting a Risk Appetite. The buffer represents an amount of capital held by a firm in excess of its regulatory capital requirements. Whereas, SII Pillar 1 is focussed on the amount of capital required for policyholder protection, a risk buffer is instead intended to ensure that a firm's regulatory capital surplus is sufficiently robust and that the firm can therefore continue with its business plans.

2.3. Market Practice

- 2.3.1. In the design of their actuarial models, firms have two main considerations. These are:
 - to ensure they comply with SII and other regulations;
 - to be able to manage their business and future plans as effectively as possible.

In practice, these considerations have a significant overlap. For example, the effective management of the future of a firm's business (and regulatory solvency) would be expected to be an important focus with its ORSA.

- 2.3.2. With these considerations in mind, common modelling practice is for firms to use discounted cashflow and risk-neutral stochastic simulation models for liability valuation. Asset valuation may be carried out using market values as a base, together with the use of asset models to give stress valuations (e.g. in order to value a corporate bond following a spread stress). For the capital calculations under Pillar 1 (and frequently also within Pillar 2 in a different form), firms use an aggregation technique. Larger firms commonly use an Internal Model based on a copula simulation approach. Smaller firms more frequently use a correlation matrix calculation (mainly using the SII Standard Formula but possibly with their own Internal Model).
- 2.3.3. In addition to valuation and capital calculations, firms also need to be able to carry out longer term projections in order to support business planning. The purpose of these plans is to ensure that appropriate targets and strategies may be developed and that future problems, such as a threat to solvency, may be anticipated and mitigated. Projections would normally include balance sheet items such as a firm's SCR and Risk Margin in order that the future solvency position may be better understood.
- 2.3.4. For long-term planning purposes, firms typically use a deterministic projection approach taking into account real-world assumptions. The assumptions may include different scenarios designed to reflect an optimistic, best estimate or pessimistic view. Projections are not necessarily straightforward as the projection of future amounts, such as market-consistent guarantee costs or a firm's SCR, may be challenging. For this reason, approximations such as the use of risk drivers to project individual metrics are

commonly used. The risks associated with such projections are normally measured through the use of sensitivity analysis.

3. Limitations of the 1-Year VaR Approach

3.1. Purpose

- 3.1.1. This section contains a discussion of the limitations of the 1-year VaR approach widely used throughout the Solvency II regime.

3.2. History of VaR

- 3.2.1. Central to Solvency II is use of the 1-year VaR. The VaR is informally defined as the minimum expected loss on a portfolio for a given probability of outcome p . The VaR can then be used as the minimum amount of capital required to be held against such losses.
- 3.2.2. The use of VaR originated within investment banking in the 1990s and became a key part of the Basel II agreement first published in 2004. VaR became established as the preferred measure of market risk within the Basel II framework. Within banking, VaR is often used in a relatively short time frame (e.g. daily or 10-day VaR measures are widely used).
- 3.2.3. In 2004, VaR was introduced in the UK insurance industry within the new ICAS regime. Key differences compared to its use in banking were that it was used for a longer time frame (1 year) and that it was applied to insurance as well as market risks. The development of the Solvency II framework more recently has continued with the use of the 1-year VaR measure as the basis of the regulatory capital requirements.
- 3.2.4. VaR has the benefits of simplicity and ease of calculation compared to other metrics⁶; however, there are drawbacks to its use. More importantly, VaR is not a coherent measure of risk. The concept of coherent risk measures was developed (Artzner et al. (1999)) as a set of properties a risk measure should reasonably satisfy. A coherent measure of risk is, therefore, defined as a function that fulfils properties of monotonicity, subadditivity, homogeneity and transitional invariance.
- 3.2.5. VaR specifically does not satisfy the property of subadditivity. The property requires that the risk measure for the sum of two risks should not be greater than the sum of the risk's individual risk measures. This is the principle of diversification – that combining risks should be beneficial (at least not make things worse). That the property does not hold true for VaR can give undesirable results. For example, where regulatory capital requirements are based on VaR, it could be possible to carry out a form of regulatory arbitrage by splitting a firm into smaller firms in order to reduce capital requirements (despite the real loss of diversification).
- 3.2.6. A further criticism of VaR is that it focuses all attention on risks at the VaR probability level to the exclusion of those inside or outside it. This may be unsatisfactory as, for example, a failure event that leaves policyholders 5% short of the amounts they are due is not nearly as bad as an event that leaves them with nothing.

3.3. Use of VaR Under Solvency II

- 3.3.1. The design of the SII balance sheet based on the 1-year VaR is that a firm should have sufficient capital such that following a 1-in-200-year event, its business should still have

⁶For example, VaR may be considered easier to work with compared to other measures such as Tail Value at Risk (TVaR).

a positive transfer value so that it could be taken on by another firm without injection of further funds. The capital requirements are, therefore, based on VaR under a 1-year time frame with a probability of 0.5%.

- 3.3.2. In practical terms for insurers, VaR is typically implemented through instantaneous stresses to net assets. The levels of stresses are set to correspond to a 1-year time frame. An aggregation technique is used to derive the overall VaR from individual stresses.

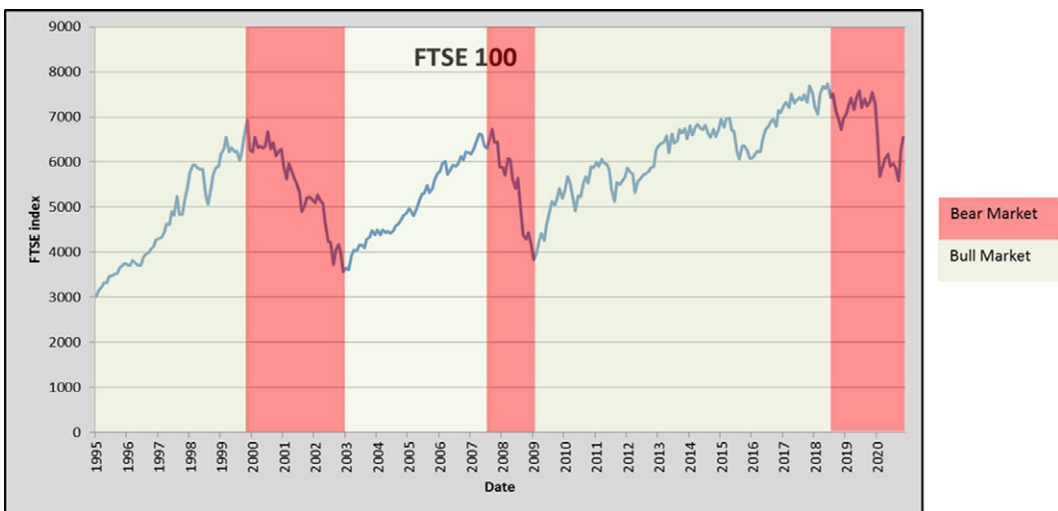
3.4. Time Frame

- 3.4.1. A clear limitation with the use and application of the 1-year VaR framework under Solvency II is that life insurance risks often materialise over the long term, and are therefore not well suited to the use of 1-year stress events. This is particularly relevant to a firm's long-term business planning. The examples in the following sections are used to demonstrate this point.

3.5. Equity Risk Example

- 3.5.1. An insight as to the long-term behaviour of financial markets can be gained from looking at the language used. Equity traders frequently use the terms "Bull" or "Bear" market referring to the tendency of the markets to go on long-term runs.

The following graph shows long-term values for the FTSE 100 from 1995 to 2020.



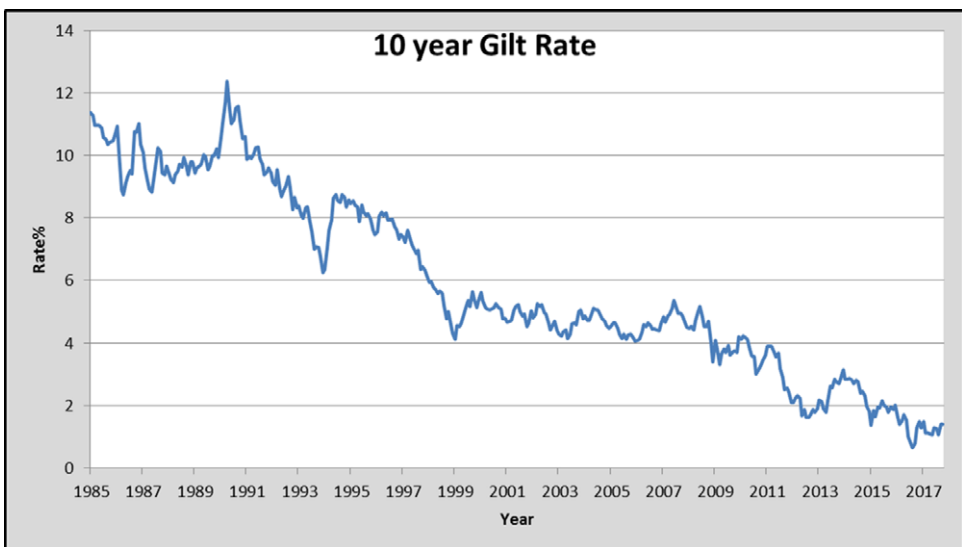
- 3.5.2. The graph shows how long-term runs characterise the market. Given the nature of such markets, a 1-year VaR focus may ensure a firm is well capitalised to survive a single bad year in the markets, but it may not be an adequate protection against a longer bear run.
- 3.5.3. Perhaps, the most severe equity market event originated at the end of the 1920s and preceded the Great Depression. The Dow Jones Industrial Average famously "crashed" in October 1929 with losses of around 12% on consecutive days (known as black

Monday and black Tuesday). However, this was only a part of much larger fall from a high of 381 on 3 September 1929 to just 41 on 8 July 1932 (a fall of nearly 90% over nearly 3 years).

- 3.5.4. This event helps to demonstrate the problems that could occur through the use of a 1-year time frame. A 1-in-200 equity calibration under a Solvency II Internal Model is normally considered to be a 40–50% fall over a 1-year time frame. Clearly, this may not be sufficient for a larger event that takes place over a longer time frame.

3.6. Interest Rate Risk Example

- 3.6.1. The following graph shows the history of the 10-year UK gilt rate from 1985 to 2017.



The graph shows a clear long-term downwards trend. Such a trend would not normally be captured within the 1-year VaR models used under Solvency II.

- 3.6.2. The above graph for interest rates (together with that for equity) illustrates the long-term nature of market events. By monitoring a firm's exposure only to short-term or instantaneous stresses, the significant exposures to longer term risks may be missed.
- 3.6.3. As an example, many UK firms wrote business with guaranteed annuity rates in the mid-1980s where the above graph shows interest rates were high (around 10% for the 10-year gilt rate). Had the 1-year VaR measure at 99.5% confidence been used at the time, this may have implied an interest rate 1-in-200 stress of 3–4% would have been considered reasonable. An amount of capital held may then have been sufficient to meet some of the shorter term movements that occurred, but would not have been able to meet the longer term trend of falling rates.
- 3.6.4. The example shows the importance of long-term risk considerations within business planning. Other forms of long-term market risk that could generate concerns could be a long bear run in the equity markets or a “stagflation” event such as occurred in Japan from around 1992 for many subsequent years.

3.7. Longevity Trend Risk Example

- 3.7.1. Longevity is amongst the most important long-term risks for a life insurer and amongst the least well represented by a 1-year VaR approach. Under the 1-year VaR framework, there are often considered to be three main aspects to the risk.
- Longevity volatility risk (variation in the number of deaths over 1 year).
 - Longevity level risk (variation in the level of expected future longevity).
 - Longevity trend risk (variation in the trend of longevity improvements).

Of these, longevity trend is typically the most significant.

- 3.7.2. Changes to the longevity trend under a 1-year VaR approach are normally modelled by consideration of how the basis used may be affected by data risk and event risk. Data risk is the risk that an additional year of live data may change the resulting fit of longevity models and therefore the assumed trend rate. Event risk is the risk that an event such as a cure for cancer could occur and would be reflected in a step change in the basis used.
- 3.7.3. The difficulty with the use of the 1-year VaR approach is the rate of improvement used in annuity models is unlikely to change significantly based on only 1 year.
- 3.7.4. With regards to the data risk, should a year of data indicate particularly low mortality, it is unlikely that actuaries would immediately fully include the new data in mortality models and mechanically set the basis to reflect new calibrations. Instead, actuaries would wish to be confident that this was a genuine change in trend rather than a single outlier year. Therefore, it is unlikely that a single year of data would materially change the basis. The COVID-19 pandemic shows an example of this. The UK mortality experience in 2020 was very significantly different to previous years (of a similar level to the 1-in-200 Standard Formula mortality catastrophe stress), yet longevity trend assumptions have not immediately changed to reflect this experience.
- 3.7.5. With regards to event risk, under normal conditions (rather than the exceptional times of the COVID-19 pandemic discussed below), it appears highly unlikely that any single event could cause a significant step change improvement in longevity over a 1-year timescale. A key example often used for event risk is a cure for cancer. In the past, there may have been a belief that a single drug could be found that would provide this. However, it is now more widely understood that cancer is an umbrella term for a large number of individual conditions. For some of these, a cure could be developed (or may have already been developed). However, it is considered highly unlikely that a single drug could cure all cancers.
- 3.7.6. Another key consideration for the cure of cancers or other diseases is that there may be a considerable time for them to be implemented. Within the COVID-19 pandemic, the development and approval of vaccines have taken place at unprecedented speed. However, for other conditions, there may be a considerably longer time for the benefits of any cure to be realised. For example, if a new drug were to be identified that reduced the long-term likelihood of heart attacks, it may require trials lasting years before the benefits can be fully realised.
- 3.7.7. In the exceptional circumstances of the COVID-19 pandemic, it may be possible that a single event could cause a step change. This could be, for example, the development of an effective vaccine or perhaps a mutation of the virus into a less severe form. This would only lead to a step change in the firm's liability valuation if the mortality basis used had been developed based on the assumption that mortality rates into the future continued the high levels seen in 2020. This assumption would seem unlikely in practice as the likelihood of a vaccine has been discussed since the beginning of the pandemic and to exclude this from the mortality basis for an annuity product could be considered to be unrealistic and not prudent.

- 3.7.8. In summary, longevity trend is a risk that can be expected to materialise very slowly over the long term. It is, therefore, not something that can be easily modelled through a 1-year VaR approach.

3.8. Other Risks

- 3.8.1. In addition to market risks and longevity risk, there are many other examples of risks that is not well suited to a 1-year VaR approach. Some key examples are as follows.
- 3.8.2. Persistency risks may include risks such as to changes in long-term lapse rates or GAO take up. These risks normally become apparent through gradual increases over many years rather than an immediate stress event.
- 3.8.3. Operational risks may take many years to fully crystallise. For example, under a mis-selling scenario. There may be many years between the initial identification of the event occurring and the cost of settling all claims.
- 3.8.4. Climate change risk is a clear example of a long-term risk. Its effects would be expected to be small over a 1-year VaR framework, but over the long term could give risk to significant problems for a firm if not appropriately considered.

3.9. Implications of Company Failure Under a Long-Term Event

- 3.9.1. Should a firm be unable to meet the costs associated with a long-term event, there are two main effects to consider:
- The effects on policyholders.
 - The effects on the firm's shareholders or members.
- 3.9.2. In theory, the 1-year VaR framework for SII is intended to ensure policyholders are protected. SII requires that firms are able to ensure they are able to cover their capital requirements at all times. Therefore, in the event of failure, it should have been possible to monitor the Solvency II surplus position (perhaps as markets fell), and as the solvency position approached zero to sell the business to another firm with a stronger capital base.
- 3.9.3. In practice, there could be practical challenges with the transfer of a firm's business under difficult conditions. It may take time to find a suitable buyer (particularly if other firms are affected by the adverse market conditions). Where a buyer is found, they may not be prepared to take on the business for the value of the Own Funds.⁷
- 3.9.4. The design of the Solvency II framework is such that it is assumed that a firm's business could be transferred for the value of the Own Funds. The Own Funds allows for the Risk Margin. The Risk Margin is calculated in respect of non-hedgeable risk only. Therefore, it is implicitly assumed that should a firm fail and its business be transferred, all risks that are hedgeable (normally all the market risks) could be de-risked. Therefore, should a firm fail, policyholders' basic liabilities would be protected but they could then be in a fund in run-off taking no material market risk. This may not be expected to give a good long-term outcome for any with-profits policyholders.
- 3.9.5. The effects on a firm's members or shareholders of failure to meet the costs of a long-term event would likely be severe. Under a breach of its capital requirements, it would likely be subject to regulatory intervention and would be expected to close to new business. Even an event that does not cause a breach of SCR could have very significant consequences. For example, if a firm's solvency position began to approach zero then

⁷The Own Funds are made up of the assets and liabilities adjusted by the Risk Margin. The Risk Margin is an approximate allowance for the cost a firm may require to take on the costs of non-hedgeable risk.

it would be likely to consider drastic actions such a closure to new business or sale of individual business units.

- 3.9.6. For the examples on market and longevity risks, there is a risk that the use of a 1-year VaR framework gives rise to a misleading assessment of the nature of the risk.

4. Mean, Median or Risk Free?

4.1. What Does a Plan Projection Represent?

- 4.1.1. An interesting question when considering traditional (deterministic) plan projections is, what does the projection represent?

It could reasonably be considered to be any of the following:

- An average (mean) projection.
- A median projection.
- A risk-free projection (i.e. there is no stochastic variation).

4.1.2. It may initially be viewed as being something determined by the nature of the input parameters. For example, if inputs such as equity returns are viewed as being an average return, the resulting projection is also an average. However, this doesn't necessarily hold true – setting all the input assumptions to represent their average value does not typically give rise to a projection that is the average of all different outcomes.

4.1.3. Where demographic assumptions are specified on a “best estimate” basis, what does this mean in practice? The Solvency II BEL gives a definition of this which is considered in section 4.2 below. However, for plan projections, firms are free to choose whatever methodology is most appropriate. Although this may not actually be specifically defined, assumptions are perhaps chosen such that the true experience may be equally likely to be better or worse (the median assumption). As for using averages based on the mean, using median assumptions as the inputs does not typically give rise to a median overall projection.

4.1.4. In effect, if we made the simplifying assumption that risk distributions were all symmetric, the average (mean) and median assumptions would be equal to the risk-free assumptions. This would result in an overall risk-free projection of key outputs such as Own Funds and CCR. Of course, it is not realistic to consider all risks as being symmetrically distributed, particularly for market or operational risks.

4.1.5. Therefore, in summary, traditional deterministic plan projections don't provide the mean or median outcomes. It is perhaps best to consider these as a broad representation of what may occur but to understand the limitations of a projection without stochastic risk modelling. Scenario testing may be used to help understand the risks. Where such plan results are presented and communicated, it should be clear that the results do not represent an “average” outcome and that there is not an equal chance of actual results outperforming or underperforming the plan.

4.2. Solvency II BEL

- 4.2.1. The BEL is defined in the Solvency II directive as follows:

The best estimate shall correspond to the probability-weighted average of future cash-flows, taking account of the time value of money (expected present value of future cashflows), using the relevant risk-free interest rate term structure.

- 4.2.2. Therefore, the BEL is defined to be the average (mean) present value of future cashflows, or equivalently, the present value of average cashflows. The question is, therefore, how should the assumptions used in the BEL be determined? There are specific requirements for the market risks (to be modelled on a market-consistent basis). The key question, is therefore, how non-market risk assumptions should be specified.
- 4.2.3. In order to align with the above definition, in theory this would indicate that BEL be calculated through the use of a long-term stochastic projection model. This could be used to produce projected cashflows allowing for stochastic variation in the underlying risks. The results could be discounted using the risk-free curve and the average taken to be the BEL.
- 4.2.4. This isn't the practice commonly used by firms. Instead, demographic assumptions are normally set by firms using approaches that broadly correspond to a median assumption for each risk. Firms use this approach for practical purposes. It is relatively simple to produce individual risk assumption estimates on the basis that these are equally likely to be over or underestimates of the actual experience (corresponding to the median). Tests are often carried out in order to validate the use of such assumptions. However, these may be challenging to carry out accurately without long-term stochastic projections. The following examples are used to illustrate this point.
- 4.2.5. Lapse risk

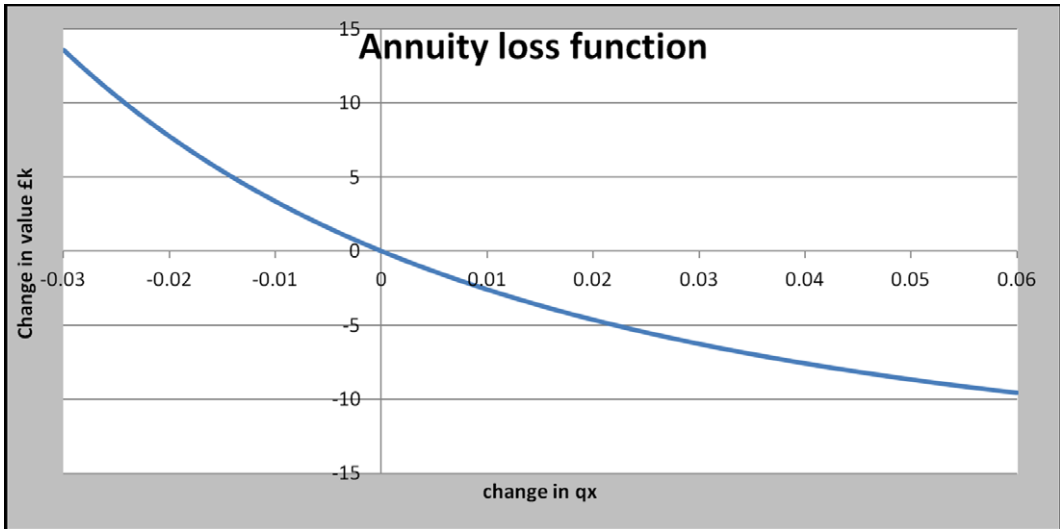
The use of individual median assumptions and an approximation will tend to work better for symmetric risk distributions. A difficulty is that many key risks are not symmetric. An example is lapse risk. A feature of lapse risk is the possibility of a mass lapse event. Such an event could consist of a very large increase in lapses under an extreme event. The risk is very non-symmetrical as there is no possibility of a similar-sized mass fall in lapses. For example, under a mass lapse event, lapses could rise from 5% to 40% but there is no possibility they could fall below zero.

- 4.2.6. Similar arguments apply to mortality risk in that event such as a pandemic or natural disaster could cause a catastrophic mortality event, but that an equivalent opposite event may not be possible. (e.g. a pandemic could conceivably cause a 0.5% addition to all q_x but nothing could cause a 0.5% decrease to all q_x) as the rates have a lower bound of 0%.
- 4.2.7. In the lapse and mortality examples, the non-symmetric nature of the risks means the calculated BEL is likely to be understated compared to the SII definition.
- 4.2.8. Bias through non-linearity

Where assumptions are specified for mortality or lapse risks, it is highly likely that the assumptions are more sensitive to a decrease than an increase. As an example, consider a basic annuity policy of £1,000 per annum in arrears for 50 years as a maximum, with spot rates of 2% for all durations. The following results may be derived:

q_x (all ages)	Value (£k)	Change compared to base (£k)
Base (3%)	17.8	-
2%	21.2	3.4
4%	15.2	-2.6

- 4.2.9. The results are clearly more sensitive to a 1% decrease in mortality than they are to a 1% increase in mortality (note that similar conclusions apply if the stress is applied by multiplying by $(1 + x)$ for stress x rather than these additive stresses).
- 4.2.10. The reason for these results is that mortality and lapse assumptions act in a compound manner over the course of a projection. For example, the probability of an annuitant in the above example surviving for 10 years is the product of surviving each of those 10 years (given they have reached that point). The following graph shows the non-linear nature of the exposure within this example.



- 4.2.11. The above effect is likely to introduce a systematic bias such that the BEL is understated according to the specified distribution. This is because the probability-weighted average value is below the value based on a median assumption.
- 4.2.12. Bias through risk interactions

Consider a product with a GAO, for which the key demographic risks are longevity and the take-up rate of the GAO. This is an example in which the combined effects of the risks are likely to exceed their individual sum (another form of non-linearity). Such risk interactions are again likely to mean that the BEL could be understated through the use of median individual assumptions. However, this does depend on the nature of the interactions.

- 4.2.13. What makes the above effect more significant is the effects of dependencies. In the above example, there would be expected to be a correlation between the two risks such that the probability of their joint occurrence is greater than it would be for independent risks.
- 4.2.14. The effects of risk interactions again may mean that the BEL is likely to be understated through the use of median individual assumptions. However, in this case, this does depend on the nature of the relationship between risks including the dependency between them.

4.2.15. Summary

In conclusion, the Solvency II BEL may not be particularly well approximated by the approaches currently used to set the demographic assumptions within the industry. What is more, the suitability of the approximation cannot be reliably estimated with the existing models.

- 4.2.16. This is not to suggest that setting demographic assumptions to reflect an approximate median for each risk is poor actuarial practice. There are many benefits to such an approach. Perhaps most importantly, it is intuitive and can be practically applied. The difficulty is that such an approach isn't consistent with the definition of the BEL under Solvency II.
- 4.2.17. It would be possible for a number of firms to develop long-term stochastic model as described in this paper in order to compute the BEL. However, this is unlikely to be a practical approach for all firms in the industry given the required frequency of reporting and the modelling sophistication required. Note that the SII SCR and RM are also calculated through stress valuations of the BEL. Therefore, these would be needed to be calculated using stressed calculations of the stochastically calculated BEL.
- 4.2.18. The analysis in this section demonstrates the challenges associated with the definition of the BEL under Solvency II. The key challenge is perhaps that it is impractical to calculate it as per the definition or to even reliably assess the accuracy of approximations used without the use of long-term stochastic models.
- 4.2.19. Within Solvency II, the sum of the BEL and Risk Margin (the technical provisions) are intended to represent a transfer value of the liabilities. These can be broadly considered, respectively, to be the average costs, together with the additional cost required to compensate for the associated risk that costs are different to the average. Perhaps, a more practical definition would be to define the technical provisions to be the sum of:
- Risk-free costs;
 - Value of options and guarantees;
 - Value of non-market risks;

where the value of non-market risks represents the value of variation of the non-market risks from the risk-free level. An alternative approach could be to instead define the technical provisions through the use of utility functions applied to the distribution of possible outcomes. These could be used to allow for the risk-averse preference of investors. However, this would still require the use of long-term stochastic risk modelling.

5. Limitations of Copula Simulation Models

5.1. Purpose

- 5.1.1. This section contains a discussion of the limitations within copula simulation models commonly used within the UK Internal Model firms.

5.2. Background

- 5.2.1. Typical current practice amongst most UK Internal Model firms is to use copula simulation models for Pillar 1 capital calculations under Solvency II. Essentially, the approach used is that a large number of 1year real-world risk simulations are generated, before

proxy functions are used to estimate the losses within the simulations. The capital requirements can then be derived by taking the 0.5th percentile of the simulated losses.

- 5.2.2. Such copula simulation models are examples of single-period models. They are based on a single time step (of 1-year length).
- 5.2.3. Risk simulation within the models is carried out using defined algorithms taking into account the copula and individual risk distributions used. The copula and individual risk distributions together define the joint distribution of risks. This joint distribution is a probability distribution that describes risk movements over the single period of 1 year. That is, it describes only the distribution of differences in the risk values at the end of the year compared to the values at the start of the year.

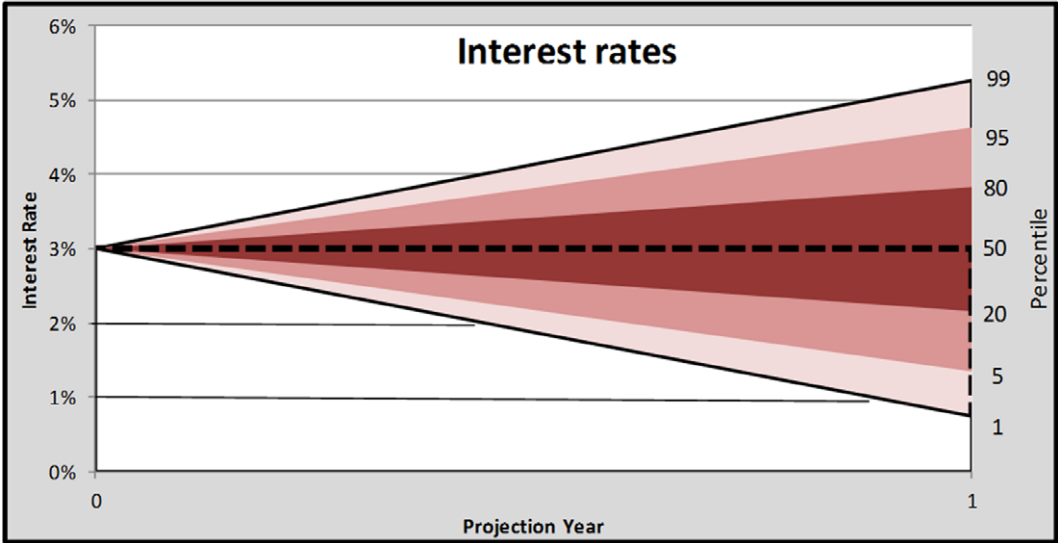
5.3. Application to Other Time Steps or Periods

- 5.3.1. There are many reasons that actuaries may wish to apply their risk models to steps or periods other than 1 year. Most obviously, this could be in long-term business planning. However, there are also other applications. For example, firms may wish to design a product smoothing process that looks at returns over 3 months. A further long-term example is that firms may wish to consider what returns a policyholder may have at the end of their savings policy.
- 5.3.2. A key difficulty with the use of copula simulation models is that they lack the flexibility to use a different time step for reasons discussed in the examples below.

5.4. Individual Risk Distributions

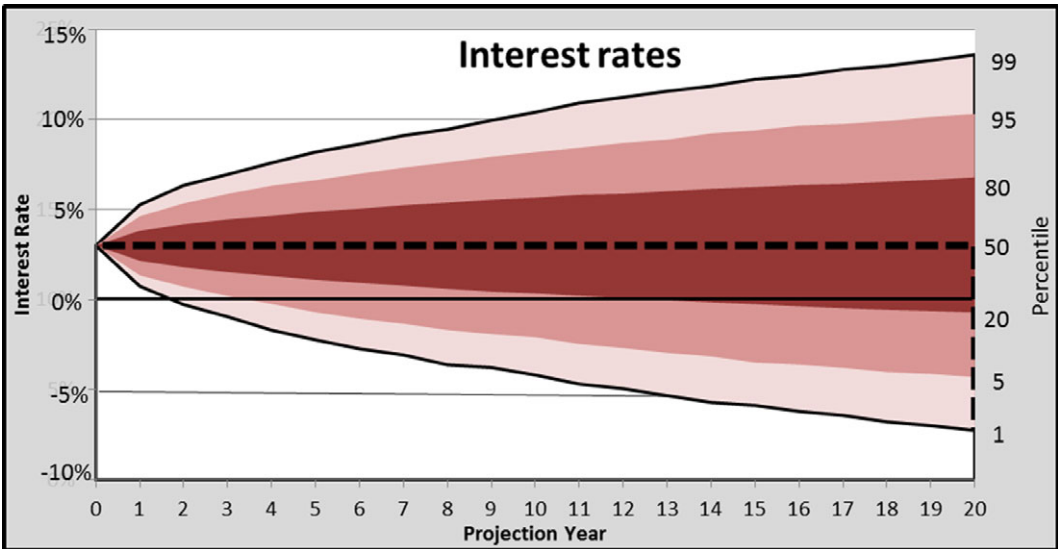
- 5.4.1. Suppose a firm needs to be able to generate simulations representing risk movements over a 3-month period. Actuaries may consider how the 1-year joint risk distribution used could be adjusted for this purpose. The first consideration may be individual risk distributions. Where a very simple risk model is used, this may be done through scaling. For example, under a basic lognormal equity model, scaling the variance may be used. However, in practice, firms typically use much more sophisticated forms of model allowing for features such as dynamic volatility and mean reversion. The use of simple scaling does not work effectively for these features.
- 5.4.2. If the firm instead needs to look at long-term projections (beyond 1 year), the way different years interact needs to be considered so that simulations can be generated to represent risk movements in all future years. The only practical way this can be achieved is to assume the years are independent and identically distributed (i.i.d.). This approach may be used to generate simulations, but they would normally be of no useful benefit. The following examples help to illustrate why this is the case.
- 5.4.3. Interest rate risk

Say that a firm uses a risk distribution such that 1-year changes in interest rates are distributed normally with zero mean and standard deviation of 1%. If the starting rates are 3% this gives the following 1-year projection.



5.4.4. The results look reasonable. This would be expected as a firm would normally have carried out a significant amount of analysis (including backtesting) to ensure the 1-year distribution is suitable.

If however, we use the same distribution for 20 years, under the assumption that the distributions in each year are i.i.d., this gives the following:

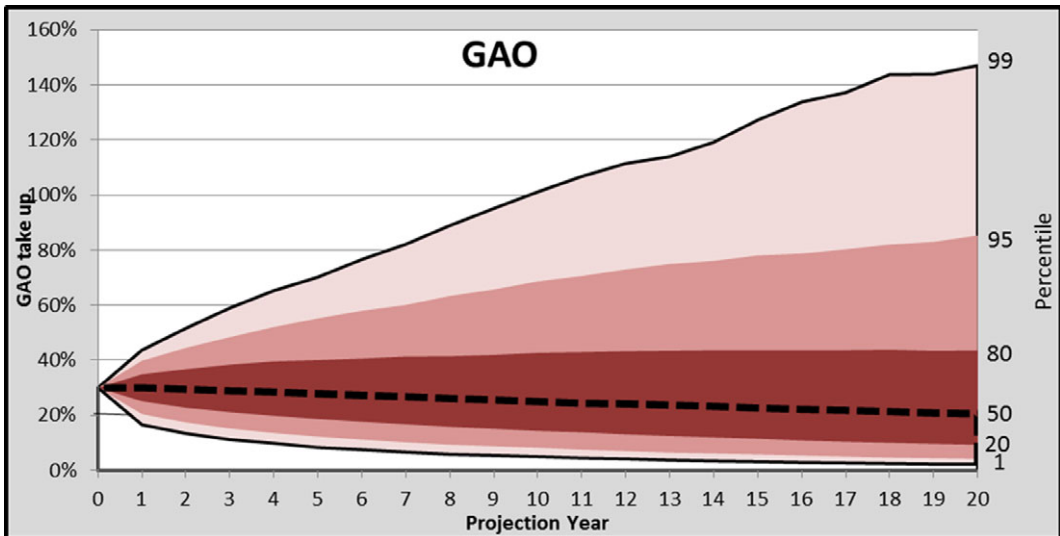


5.4.5. It can now be seen that after 20 years, rates as low as -5% are seen within the simulations. While negative rates are now an accepted possibility, it may seem unrealistic that rates as low as this should be incorporated. The graph shows the effects of using a model with i.i.d. risk distributions for different years. The key problem here is the lack of any mean reversion that is inherent within a model based on i.i.d. distributions.

5.4.6. GAO take-up risk

This example discusses GAO take-up risk, but could equally be applied to similar risks such as lapse risk. Say in this case a firm has annuity options for which the current take-up assumption used in the reserving basis is 30% . The form of risk would normally take a multiplicative stress. For example, the rate in 1 year would be $1 + X\%$ times the current rate.

5.4.7. If a 1-in-200 stress was considered to be 50% , based on a zero mean normal distribution the risk would be distributed $N(0, 19.4\%^2)$ (the standard deviation of 19.4% is 50% divided by the inverse standard normal distribution function for $199/200$). By assuming i.i.d. risk distributions, the following long-term graph of rates can be seen:



5.4.8. In this case, the obvious problem is that the take-up rates start to go above 100% . A simple approach to address this would be to cap them at 100% , but the problem is that the results would then not be realistic. For risks such as GAO take-up or lapse, the rates are bounded by 0% and 100% and are often unlikely to closely approach either of these. For example, there are likely to be some policyholders who will take cash over an annuity option if they have a short-term need for funds.

5.4.9. Another problem with extending this nature of risks over multiple time periods is that the calibrations are often driven by judgement taking into account various event risks. For example, a GAO take-up calibration may take into account the possibility that regulators require firms include a clear message on policyholder statements that they have a valuable option they should consider using. This kind of event may be considered as a one-off in that once it has occurred the event risk is no longer present. For this reason, assuming the risk distributions are i.i.d. over different years may not be appropriate.

5.4.10. Operational risk

Similar to the event risk example on GAOs above, operational and counterparty default risks are often event based. Operational risk is typically modelled through scenario analysis. Such a scenario could include, for example, a mis-selling incident on a specific product resulting in the need to compensate policyholders. Such an incident is likely to occur only once on a particular product (although it could possibly occur on other products). Another point to consider is that the amount of operational risk is unlikely to remain fixed over time. For example, it would in reality be expected to vary with the complexity and volume of business in the future.

5.4.11. Risk Aggregation

The joint risk distribution used in a copula simulation model is made up of the individual risk distributions, together with a copula. The copula is the means of combining the individual risk distributions in order to get a joint risk distribution. The copula most commonly used in current industry practice is the Gaussian copula. This is parameterised through a correlation matrix.

5.4.12. In addition to the problems described above for individual risk distributions, the correlations (or more generally the copula) provide a key difficulty when looking at time steps other than 1 year.

5.4.13. Where a shorter than 1-year time step is required, the difficulty is that correlations can vary substantially depending on the time step used. For example, considering the correlation between the UK FTSE100 and the US S&P500 from 1995 to 2019, the correlation using annual steps is 91% yet the monthly correlation is only 80%. This result is not unusual and perhaps reflects that short-term market fluctuations may be largely uncorrelated between the different markets, but longer term trend factors may be much more closely related.

5.4.14. Where multiple time steps are required to construct a long-term projection, the difficulty this time is that using the same correlations in each step is not likely to be realistic. One problem is that correlations tend to vary depending on market conditions such that the correlation of market risks may be much higher in a bear market than a bull market. Another problem is the relationship between risks varies over time. For example, lapse rates may be highly negatively correlated with equity returns in the early years of a unit-linked savings product (if, e.g. there is a guaranteed surrender amount). In the later years, the unit fund may be well in excess of the guarantees and the correlation could be less significant.

5.5. Risk Losses

5.5.1. The key components of a copula simulation model are the joint risk distribution and the loss model (represented by a proxy function). If a copula simulation model is used on a short time frame, the proxy model is likely to be suitable for use. If longer term projections are required, the proxy function is not expected to be realistic.

5.5.2. A key challenge is in respect of the volume and maturity of business. This is likely to mean that, for example, a 20% equity or lapse stress costs a materially different amount in 10 years compared to that at time zero. A scaling approach for business volumes could mitigate these problems but this does not account for how risk exposures change through a policy lifetime. For example, a longevity stress has a much lower cost to an 80-year-old annuitant than a 70 year old.

5.6. Summary

- 5.6.1. The above examples demonstrate that the 1-year copula simulation models commonly used in industry cannot be effectively used outside of the 1-year time frame. The reason for this is in the inherent design of the models. This is that they:
- Use a single-step approach.
 - Are purely statistical models.
- 5.6.2. The use of a single-step approach is discussed in detail above. This gives the models a fundamental weakness that means they can never be entirely suitable for other time periods even with the use of different approximations or extensions.
- 5.6.3. The second point above is that the copula simulation models are purely statistical risk models. The only risk information used within the models is the probability distribution of each risk together with the copula. They do not use the information on economic theory or the underlying drivers of different risk events or how these may affect other risks, instead, they are simply statistical models.
- 5.6.4. As an example on the limitations of purely statistical model, we consider two forms of interest rate risk model.
- 5.6.5. Model 1 – Purely statistical model

Annual change in Interest rates \sim Normal(a, b^2) for mean a and standard deviation b

- 5.6.6. Model 2 – Stochastic differential equation model (the Cox–Ingersoll–Ross (1985) or CIR model)

$$dr_t = a(b - r_t)dt + \alpha r_t^{0.5}dW_t$$

where r_t is the short rate of interest at time t
 a, b, α are parameters
 W_t is a Wiener process

- 5.6.7. Unlike the purely statistical model, the CIR model incorporates modelling assumptions on the true behaviour of interest rates. For example, it incorporates a mean reversion term, $a(b - rt)dt$ and that the amount of random variation of rates is proportional to the square root of the current rate.
- 5.6.8. It is only by using a representative mathematical model of interest rates that results over a long time frame may be derived. The same is true for other risks. If we wish to understand the long (or very short)-term behaviour of the risks, we need to use a time series model incorporating the realistic features of each risk.
- 5.6.9. The above example uses a single risk only. Another key consideration is how risks are combined to form a joint distribution. Copula simulation models of course achieve this through the use of a copula. Similar to the discussion on interest rates above, a copula simulation model represents a purely statistical approach to the formation of the joint risk distribution.
- 5.6.10. The copula, (for which the key input is normally a correlation matrix) simply models how likely it is that different risk events will occur at the same time. Therefore, a copula simulation model does not take into account how movements in one risk may give risk to events in another. Instead, it simply models the likelihood that events occur together. This is an approach that may be appropriate under the single-step approach used within copula simulation models. However, over a longer time frame, it is important that a hierarchical form of the model is used. This is discussed in detail within the next section.

6. A Long-Term Stochastic Modelling Framework

6.1. Purpose

- 6.1.1. This section sets out how the structure of a long-term stochastic modelling framework could be developed.

6.2. Introduction

- 6.2.1. In order to be able to achieve the aims of monitoring and managing risk over the long term, a model should have the following design points:
- It should encompass all material risks to an insurer that can be directly modelled (this may exclude some exceptions such as regulatory risks).
 - For these material risks, the model should be capable of assigning the probability of any single or combined risk event over any specified time period. This is the **Risk model**.
 - For these risk events, the effect on key business metrics should be modelled. This is the **Loss model**.

6.3. Modelled Risks

- 6.3.1. The modelled risks would normally include risks such as:
- Market risks.
 - Demographic risks.
 - Expense risks.
 - Counterparty risks.
 - Operational risks.
 - New business risks.

The risks would need to be fully defined with an appropriate level of granularity. It may also be possible to model liquidity risks depending on the design of the model.

- 6.3.2. The following examples of risks would not normally be included from a model:
- Strategic risks.
 - Risk of regulatory change.
 - Group risks.

These risks cannot be practically modelled through the use of probability distributions. However, they may be separately considered through the use of scenario analysis.

6.4. Risk Model

- 6.4.1. As stated above, the risk model should be capable of assigning the probability of any single or combined risk event over any specified time period. This is not dissimilar to the copula models described in section 5 above whereby the probability is determined through the use of a copula and individual risk distributions. However, the key difference here is that the copula models are **single-period models**. They may therefore be used to describe probabilities over these single periods (normally 1 year), but are unable to be used for the more challenging requirement of describing probabilities over **any** time frame.
- 6.4.2. As an example, a copula model could be used to estimate the probability of equities falling 20% and spreads increasing 100bps over a 1-year time frame (assuming the copula

model is based on 1-year steps). What such a model could not do is answer questions such as:

- What is the probability of a 20% equity fall over a month?
- What is the probability of a 100bps spread increase over 5 years?
- If equities fall 20% over the next year, how long would a full recovery take on average?
- What is the highest we expect to see spreads over the next 5 years (to 99% confidence)

6.4.3. In order to address questions such as these, it is necessary to use a time series model. This may take a discrete or continuous form. Discrete time series models include simple Moving Average (MA) models through to more complex forms such as Generalised Auto Regressive with Conditional Heteroskedastity (GARCH). Continuous time series is defined through the use of stochastic differential equations to describe variable changes. An example is the Vasicek (1977) model of interest rates:

$$dr_t = a(b - r_t)dt + \alpha dW_t$$

where r_t is the instantaneous rate of interest at time t , a , b , α are constants and W_t is a Weiner process.

6.4.4. The discrete and continuous time series may both be used to achieve the aim of long-term projections. Where a discrete approach is used, this of course means the modelling may only be applied to those discrete steps (e.g. projections may use monthly steps). While it is normally possible to specify a process through either a discrete or continuous approach, a continuous approach is typically preferred for the greater flexibility this brings.

6.4.5. As an example, a simple equity model could take the following form (e.g. the Geometric Brownian Motion that underlies the Black–Scholes model):

$$dS_t = \mu S_t dt + \alpha S_t dW_t$$

where S_t is the equity value, μ and α are constants and W_t is a Weiner process.

This model has the solution that the equity value S_t is lognormally distributed with expected value and variance given by

$$E(S_t) = S_0 e^{\mu t}$$

$$Var(S_t) = S_0^2 e^{2\mu t} (e^{\alpha^2 t} - 1)$$

These results can be used to provide the answers to questions such as those above (section 6.4.2).

6.4.6. For a more complex form of model, simulation modelling is required to derive the required results. For example, under this model, we could have generated a high number of stochastic simulation paths and used these to find the proportion that gave a 30% fall over the month.

6.4.7. The above examples have been based on single risks only. It is important that the model is also able to be used for joint risk events. This can be achieved through the use of multivariate time series incorporating multiple risks. An example of this is the multivariate version of the above Geometric Brownian motion model.

$$dS_t^i = \mu S_t^i dt + \alpha S_t^i dW_t^i$$

where S_t^i represents the value of equity index I at time t .

Using this model it is possible to answer questions about both combined and individual risk events. This is an example under which all risks take the same form (Geometric Brownian Motion). In reality, it is likely to be more suitable to apply a number of different forms of the risk model.

6.5. Loss Model

6.5.1. Within firms' existing copula simulation models (single-period models), the loss model consists of a proxy function used to estimate losses in Net Asset Value (NAV) as a function of movements in each of the risks.

Change in NAV = $f(\text{risk1, risk2, risk3, risk } n)$ for n risks.

This function is applied to an array representing the risk movements for each simulation.

6.5.2. Under a multiperiod model, there are two key differences in the requirement.

- The model needs to use a time series of risk inputs per simulation rather than just a risk movement.
- The model needs to be capable of producing time series output data.

6.5.3. The time series data output needs to include at least NAV losses per simulation at each time step. A more sophisticated model may also be able to estimate balance sheet items such as the SCR and Risk Margin and use these to construct an estimated solvency position.

6.5.4. In practice, these requirements are beyond the scope of traditional proxy models and it is, therefore, only through discounted cashflow models that results may be derived.

6.6. Output Key Business Metrics

6.6.1. The key metrics that may be required from a long-term stochastic model could include the following:

- Surplus generation (changes in NAV).
- Solvency II balance sheet (assets, liabilities, SCR, Risk Margin, TMTP, MCR, surplus).
- Economic Capital balance sheet.
- Liquidity.
- Policyholder returns.
- Value of new business.

6.6.2. These metrics should be broken down where necessary into contributing elements. For example, by product or fund. Surplus generation should be broken down into different elements such as variances by risk and product.

7. Practical Implementation of a Long-Term Stochastic Model

7.1. Purpose

7.1.1. This section discusses the practical steps required in the construction of a long-term stochastic model.

7.2. Model Overview

7.2.1. The model design should be considered as a series of individual modules. The modules should be constructed as independent elements but with a consistent approach to design and implementation.

7.3. Data Store

7.3.1. A key element of the design is the flow and storage of data. The main principle is that all data items should be stored consistently as simulated time series. This is expected to generate a high volume of data. Where necessary, it may be practical to only store

(for the long term) simulation data for key results rather than all series used in the calculation.

7.4. Market Risk Module

7.4.1. The purpose of the market risk module is to produce simulation results that consist of a set of time series representing economic indices. These would typically include the following **asset cashflow** series:

- Equity dividends.
- Bond coupons and redemption values.
- Property rental income.
- Equity Release Mortgage income.
- Derivative payments.
- Cash income.

They would also include the following **value** series:

- Equity values;
- Bond values and ratings;
- Property values;
- Equity Release Mortgage values;
- Derivative values;
- Cash value;

and the following series of **economic indices**:

- Swap and gilt-based yield curves.
- Corporate bond spreads.
- Current and expected inflation.
- Current and expected implied equity and interest rate volatility.
- Currency exchange rates.

7.4.2. The above items are typically available through real-world Economic Scenario Generator (ESG) models widely used through the industry. The ESGs incorporate detailed time series models of each of the key elements above, together with an allowance for the dependencies between different indices.

7.4.3. For this reason, the most practical approach to the market risk modules of a long-term stochastic model is simply to use existing ESG capability. Where this is not practical, it would be possible to construct a relatively simple ESG using more basic elements such as a lognormal equity model. However, such an approach is unlikely to come close to the level of sophistication within commercially available real-world ESGs.

7.4.4. The ESG model used will need to be calibrated appropriately using established techniques. However, the market risk module would not necessarily require the use of any non-market time series as inputs. This implicitly uses the assumption that economic movements are not caused by movements in non-market risks. This is not necessarily true in practice. For example, a longevity event could cause firms increased costs in the support of defined benefit pension schemes with the result that bond and equity markets fall. However, economic volatility resulting from such sources is expected to be a relatively small element.

7.5. Mortality/longevity Risk Module

- 7.5.1. In the same way that simulated time series data for market risks may be sourced from existing systems, simulated time series data for mortality risks may be sourced from commercially available software or from open source software such as R. The software may be used to apply a number of existing stochastic mortality models (e.g. the Lee–Carter (1992) and Cairns–Blake–Dowd (2006) models).
- 7.5.2. These models may be used to generate simulated time series of the q_x probabilities of death by age and sex or other key indices.
- 7.5.3. Mortality would not normally be expected to be materially affected by other insurance risks. In the long term, trends in mortality are affected to an extent by economic prosperity. For example, higher GDP may lead to higher government and personal health-care spending. Such effects are not typically allowed for in common stochastic mortality models. For this reason, it may be practical not to include an allowance for the relationship between market risks and mortality. An alternative would be to use a causal mortality model, within which contributory factors to mortality, including economics/healthcare spending could be included.

7.6. Lapse Risk Module

- 7.6.1. Unlike market risks and mortality risks, lapse risk simulation models are not widely used within the UK. For this reason, where lapses are a material risk, it may be more appropriate to develop a stochastic lapse model. There are many ways such a model may be constructed. The following approach shows an illustrative example based on a savings policy (other policies such as on protection business may be affected differently).
- 7.6.2. The first step in the construction of a lapse model may be to consider the **drivers** that could cause a policyholder to wish to lapse. The following factors are likely to be relevant:
 - Long-term plans.
 - Unplanned cash needs.
 - Loss of confidence in the company.
 - Investment switching (e.g. to get a cheaper product elsewhere).

In turn, there may be reasons that would reduce policyholder's desire to lapse. There could include:

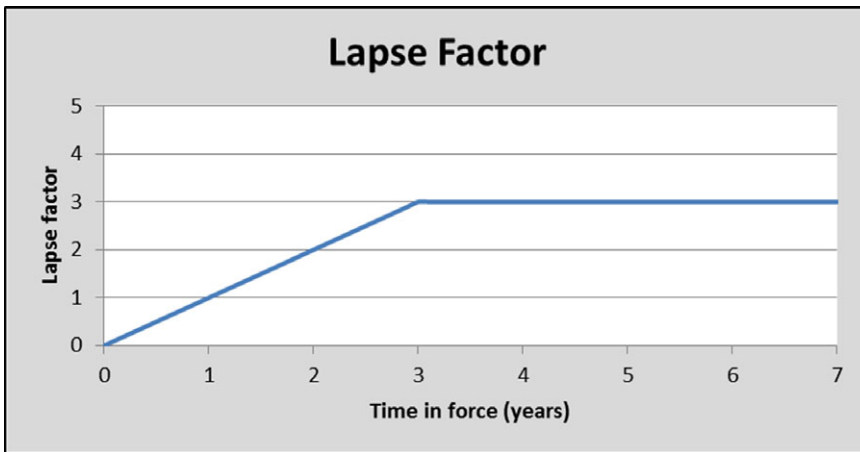
- Surrender penalties.
 - Guarantees (e.g. if the policy gave a valuable guaranteed surrender benefit at 5 years, the policyholder is unlikely to surrender shortly before this).
- 7.6.3. Consider the propensity to lapse as a score with zero being the point where no lapses occur.

If we can consider how each of these contributes to an overall score, we can then estimate a function that converts this score to a lapse probability.

- 7.6.4. This could be done at an individual policy level. However, in order that results can be used for simulation modelling, it is more appropriate that model points are used to group policies into key categories such as start year and time remaining. Grouping could also consider factors such as age or fund size. However, the grouping should not result in a very small amount of data in different groups. Here are some example parameters that could be used:

7.6.5. Long-term plans

Long-term plans for lapse rates represent lapses that are expected to occur irrespective of other factors. They could include, for example, funds set aside for a deposit on a new home or a round the world trip. As these plans don't depend on other factors, they don't need to be considered as a function of other indices. They may however depend on the time from outset of the policy. For example, for a long-term savings product, it is unlikely that there will be many lapses in the first year that was due to policyholders' long-term plans. The following graph shows how the lapse factor could be considered to vary over time.



7.6.6. Unplanned cash needs

Unplanned cash needs may by their nature be unrelated to the time in force. Examples of cash needs could be

- An unexpected bill (e.g. housing repair, tax, medical treatment).
- Unplanned spending (new car, new conservatory).
- Loss of income (e.g. due to redundancy).

These items are effectively a further breakdown of the key lapse drivers set out above. With regards to the first two items here, there are always likely to be an element of these within groups of policies and they are not considered to be dependent on other factors or time.

Example parameters for these could be

Unexpected bills lapse factor = 1 (constant).

Unplanned spending lapse factor = 2 (constant)⁸.

7.6.7. With regards to loss of income, rather than being unrelated to other factors, this could instead be expected to be correlated with economic factors. A key economic variable here could be national rates of unemployment. Loss of income-related lapses could then be expressed as a function of national unemployment. However, to do so would need a

⁸A more detailed approach here could set the unplanned spending lapse factor could vary with factors such as the age of the policyholder, policy size or economic conditions.

risk model that may estimate this. An alternative would be to use movements in equity markets as a proxy for economic prosperity and therefore employment.

As an example, we could use a function of the following form:

Lapse factor due to loss of income = $0.5 + \text{If (equities have fallen 20% over 6 months, 1, 0)}$

This allows for increased lapses under challenging economic conditions.

7.6.8. Loss of confidence in company

The level of trust in a company is likely to be a significant factor affecting lapse risk. Perhaps the most significant concern here is that a single event (possibly an operational risk event) could result in widespread reputational damage and give rise to a mass lapse. There are two main possible ways in which confidence-based lapses could be modelled:

- Model an index of company/brand value and use this to derive a lapse loss function,
- Assume the main driver of reputational damage would be operational/conduct risk and therefore use this to derive a lapse loss function.

7.6.9. Using the first of these two approaches, a firm's existing operational risk model could be used to understand how different operational risk scenarios not only have a financial cost, but also give a cost to reputation. For example, reputation could be modelled using again a scale between 0 (worst) and 10 (best), reputation could be assumed to be reduced under specific operational risk scenarios and to benefit as time passes since events have occurred. Given a reputational index, a reputational lapse factor could be expressed as a function of reputation.

7.6.10. Under the second approach, reputation-based lapses could be derived from a function of operational risk losses.

7.6.11. Of these two approaches, the first is likely to be stronger as it could explicitly allow for modelled operational risk scenarios to give rise to reputational damage and a possible mass lapse event. An example function could be

$$\begin{aligned} \text{reputational lapses} &= 0 \text{ where reputation index} > 5 \\ &= 5 - \text{reputation index where } 1 < \text{reputation index} < 5 \\ &= 10 \text{ where reputation index} < 1 \end{aligned}$$

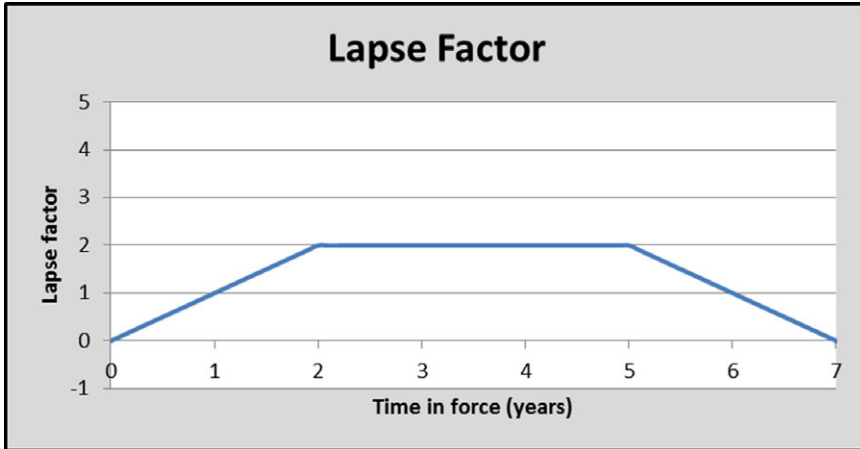
The design of the function is such that a mass lapse event may occur should an operational event cause severe reputational damage.

7.6.12. Investment switching

Lapses due to a switch in investment could take many forms. They could include policyholders wishing to:

- Switch to a similar product offered by a competitor (e.g. to take advantage of lower charges).
- Switch to a different kind of investment (e.g. direct equity investments).
- Switch away from equity, bond exposure to just hold cash.

The first two types here are likely to occur regularly to some extent. However, they may be affected by the time from the start of the policy and possibly the time remaining. Such switches may be unlikely near the start or end of a policy. Therefore, a function such as shown in the graph below could be used (assuming a product that matures after 7 years).



With regards to switches out of risky investments into cash, these depend on the nature of the product offered. However, it may be considered that increased numbers of such lapses occur in volatile markets (or following market falls). Therefore, a function that expresses the lapse factors in terms of an equity volatility index or recent equity falls may be appropriate.

7.6.13. Surrender penalties and guarantees

Surrender penalties and guarantees vary significantly over different products. Where surrenders affect commission, this would also have a significant effect.

A function that takes these factors into account would need to be carefully designed for specific product features. Examples that could be allowed for are that:

- Where a surrender penalty applies for an initial period – lapse factor is reduced by 3 within that period.
- Where a policy has a guarantee that is more than 10% “in the money” – lapse factor is reduced by 3.

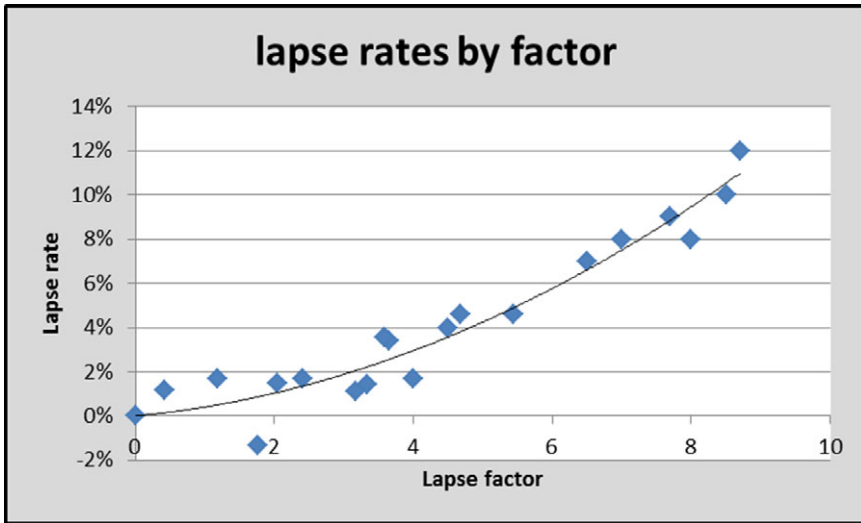
7.6.14. Random variation

Having allowed for the key features affecting lapse rates, it is important to recognise that actual experienced lapse rates will also be affected by other factors beyond the scope of the model. For this reason, the lapse factor should also include a stochastic element. For example, the lapse factor could be assumed to include a term of $\text{Normal}(0, s^2)$. This allowance for uncertainty is to recognise that, even where factors such as market conditions and operational risk events are fixed, actual lapses observed are still subject to uncertainty.

7.6.15. Mapping from the lapse factor to actual lapses

After deriving an approach to calculate a lapse factor for each model point, the final stage is to design a function to convert the lapse factor into an actual lapse rate. The lapse rate could be expressed as a probability over a specified time frame or instead as an instantaneous “force of lapse”.

To map from the lapse factors onto lapse rates, lapse factors can be calculated for model points using actual company data. The results can be used to produce a scatterplot of data to which a function may be fitted as illustrated below:



7.6.16. Fitting the parameters

The above example is intended to show at a high level how a causal lapse model could be developed. In practice, the process used to fit parameters could be a blend of judgement (e.g. to allow for reputational effects), together with statistical analysis. The intention of the modelling would be to be able to predict the lapse rates as accurately as possible from the lapse factors. This should be on an “Out Of Sample” (OOS) basis in that testing of the function should use points that were not included in the fitting process.

The development of such a lapse model could form an annual cycle whereby, the performance of the existing model is assessed, if the results are not considered to be strong enough then investigations could take place in order to understand what factors have not been sufficiently modelled so that improvements can be made.

7.6.17. Lapse model summary

At the end of the above process, the result is a causal lapse model in which lapses are analysed and modelled through consideration of the factors that cause their occurrence. This may be a stronger approach than traditional techniques whereby lapse rates are simply fitted to data by duration. The table in the next page discusses the differences in the approaches.

7.7. GAO Take-up Module

As with lapse risk, causal models of GAO take-up risk are not widely available from commercial providers. Therefore, it may be appropriate that causal models are instead developed internally. As with lapse risk, it is important to consider the underlying causes that a policyholder may choose to take up an annuity option. There are two ways such an annuity option may be considered:

- A policy that provides an option to take a guaranteed annuity.
- An annuity policy that provides an option to instead take a cash sum.

	Traditional Approach	Causal Approach	Comments
Best estimate	Lapse rates are estimated by duration. These may represent a median expected outcome.	Best estimate lapse rates are derived as the probability-weighted average rates derived from simulations in the causal model.	Causal model requires more work to calibrate but may give greater insights. It may respond better to new information such as how lapses increase under stressed market conditions.
Risk distribution	Probability distribution fitted to historical changes in lapse rates.	Simulations from the causal lapse model directly determine the risk distribution.	Distribution fitting is compromised by data limitations (not enough years). Causal model may potentially give much greater accuracy.
Relationship with other risks	Normally approximated using correlations incorporated within a Gaussian or T copula.	Simulations from the risk model directly determine the relationship between risks.	The causal model directly considers the relationship between risks rather than simply using estimated correlations (which are themselves very approximate).
Flexibility	Can be used only for the 1-year period normally applied within Solvency II.	Can be used for any time period (longer or shorter term). Can be used for what-if analysis, for example, what would happen to lapse rates under market falls.	As the causal model directly considers what factors give rise to lapses it may be far more flexible and adaptable to different purposes.

The approaches are of course equivalent. The most suitable to use is likely to be aligned with the wording used in the policy. For the purposes of this paper, the first of the two methods is used.

7.7.1. Types of GAO policyholders

Consider three groups of policyholders:

- Those who have a need for an annuity – they will take the GAO if it is “in the money” at maturity.
- Those who have no wish for an annuity and so will always take cash (some products include an annuity option element that many policyholders have no interest in).
- Those who are happy to take the cash option, but could consider taking the annuity if the guarantee makes it particularly attractive.

The presence of the first two groups here means that GAO take-up rates will always be greater than zero and will never reach 100%. Here, the GAO take-up rate for a group of policyholders is measured as the policyholders taking the annuity divided by the number of GAOs in the money.

If the proportions of policyholders in each of these groups can be estimated, this may be used to develop a GAO model. The most accurate way of estimating the proportions could be to ask the policyholders through a limited survey. This could be either policyholder recently matured or instead policyholders currently inforce. If this is not practical then the proportions could be estimated by subject matter experts familiar with the details of the product. Proportions could potentially be split by the cohort of business.

7.7.2. Drivers of GAO take up

For the policyholders in the first two groups, they will simply take the annuity (if in the money) or the cash. For the third group, as with lapse risk, we can consider the propensity to take up a GAO as a score with zero representing the point at which no policyholder would take up the option.

The key drivers of GAO take up may be:

- The value of the option (annuity relative to market rates).
- Company reputation.
- Visibility of the option.

Of these, the value of the option is likely to be most important.

7.7.3. Value of option

Where the annuity is guaranteed as a fixed rate, the value of the option is determined by market annuity rates. These are largely driven by market interest rates and mortality assumptions. An example function based on the value of the option would be to say that

$$\text{GAO factor} = \text{constant} \times (\text{GAO annuity} - \text{market annuity}) / \text{market annuity}$$

This factor can readily be estimated for different model points using simulated interest rates and annuitant mortality.

7.7.4. Company reputation

In a similar way to lapse risk, if a company's reputation has suffered, it is possible that this could reduce annuity take up as policyholders may no longer trust a company and may wish therefore to have no further contact with them. A function could be derived with a similar form to that used for lapse risk.

7.7.5. Visibility of the option

For some products, the presence of a valuable GAO may not have always been clear to policyholders. Typically, visibility of such options has improved over time as firms improve the level of communication. The visibility of the option is not something suitable for stochastic modelling. However, it could be included in the GAO factor as a function of the year. For example, if it is known that the communications were changed in 2010 to make an option much clearer, a function could be derived to take this into account.

Option visibility on its own is unlikely to have an effect on take up (e.g. if the GAO value is only 1% higher than market rates, take-up rates (amongst this policyholder group 3) are unlikely to be material even if visibility is crystal clear. Similarly, option value on its own may make very little difference if visibility is so poor that very few policyholders are even aware of the option. For this reason, the GAO factor may be better expressed as the product of the GAO value factor and visibility factor rather than the sum.

7.7.6. GAO model summary

Having derived a model of GAO factors, it remains to parameterise this and use it to derive a function that maps from the GAO factors to actual take-up rates. This may be done using a similar approach as for lapse risk. Also as for lapse rates, it is also important to include an allowance for randomness.

Having carried out these required steps, we will have constructed a causal GAO risk model. As for lapse risk, this may be a much stronger and more flexible tool than a model constructed by traditional techniques. A key example here is how GAO take up risk interacts with interest rate and mortality risks. Existing aggregation models simply use a correlation to estimate how much GAO take up, interest rates and longevity vary together. A causal model by contrast uses the actual value of GAOs rather than just changing interest and mortality rates. It also directly models how

the interest rates and mortality rates drive take-up rates, rather than simply using correlations to estimate the likelihood of them occurring together.

7.8. Other Risks

7.8.1. Overview

This section discusses how a causal risk model may be developed for a number of other key insurance risks. The section is intended to provide a high-level summary for each risk rather than a detailed model.

7.8.2. Other protection risks

Other types of protection policy may include

- Critical illness.
- Income protection.
- Private medical care.
- Long-term care insurance.

These types of products share similar risks to that of mortality risk. However, an additional complexity is that multiple state modelling may be required to model transitions between different states (e.g. healthy, in deferral, in payment, retirement, death on an income protection policy.). As for mortality risk, there are a number of existing stochastic models to cover such risks. However, they may be less readily available than those for mortality and may therefore require development work to successfully implement.

Protection policies such as those above would be expected to not be materially dependant on other market or insurance risks with the exception of mortality. Where a firm has significant mortality and the above protection risks, it may be appropriate to construct a single stochastic model incorporating mortality with sickness/morbidity.

7.8.3. Expense risks

An insurer's expenses may be considered in a number of ways. Expenses result from areas such as policy administration, new business activity and investments. There may be a number of fixed costs (at least over the short term), together with other costs that vary with business volumes and assets under management. All expenses are likely to grow with inflation. This may include an element of macroeconomic inflation, but also other elements such as any excess of wage inflation in excess of macroeconomic inflation.

The key risks to expenses may be:

- Business volumes being insufficient such that fixed costs per policy increase.
- Expense inflation is higher than expected.
- Unexpected one-off cost (e.g. the introduction of Solvency II caused large project costs for firms).

Firms would typically use a detailed (deterministic) expense model in order to consolidate expenses from the different areas and use these to produce an expense allocation by product/policy, together with a long-term forecast of the firm's expenses per year taking into account business plans.

Such a model could be converted into a stochastic model by first expressing the different elements as a function of the key factors that affect them. For example, a model could be developed to give the total expenses in each future year as a function of:

- Business volumes inforce (by product).
- New business sales.
- Macroeconomic inflation.
- additional inflation (e.g. wage inflation above macroeconomic inflation).
- Assets under management.
- “one off” project costs.

Of these items, most would be available from other sources. For example, business volumes are a function of lapse risk, mortality risk and new business sales. Macroeconomic inflation would be available from market risk models.

Bespoke modelling may be required for additional inflation (there are existing models of wage inflation) and “one off” project costs. A simple stochastic model of these project costs could be developed through investigations of the extent of past company spending on these.

7.8.4. Counterparty risk module

This risk relates to the possibility of default of key counterparties such as those involved in reinsurance or hedging. It could also be considered to affect the future terms offered. For example, a firm using significant reinsurance is exposed to the risk that the reinsurance terms worsen at a future renegotiation.

Regarding the default of counterparties, the first step should be to consider the factors that could cause this to occur. These could include:

- Market factors – counterparties may have high market exposure and therefore be exposed to such risks.
- Mortality experience – very high experience (e.g. as may be seen in the COVID-19 pandemic) may weaken reinsurers balance sheets to the extent there may be a risk of default.
- General insurance experience – many reinsurers cover general insurance as well as life and therefore will have such exposure.
- Defaults – where reinsurers pass some of the risks to others then one default may trigger another. Similarly, the default of a major investment bank could cause defaults amongst others.

Using these or other factors, a model could be developed to give the probability of default of each major counterparty for a firm. Such a model could take into account the rating of each counterparty and the extent of recoveries expected under a default.

7.8.5. Operational risk module

The UK Internal model firms typically use scenario-based operational risk models. These normally contain a number of representative operational risk scenarios for which the frequency and severity are modelled. The events are combined together using copula simulation modelling in order to derive an overall operational risk distribution.

Such models may be developed for use in a causal risk approach. The key questions to consider are:

- To what extent are each of the scenarios dependent on other modelled factors?
- How long could the scenarios take effect over?

- If the scenario has occurred, could it occur again?

As an example of these, consider a scenario under which admin errors cause incorrect lapse payments and costs associated with this. The likelihood of such a scenario could conceivably be related to high lapse experience (making claims processing overstretched). The severity of such a scenario may be related to the value of the lapse payments being made at that time or the number of lapses (depending on the nature of the error).

The time period of an operational risk event is something not normally considered within the firm's existing operational risk models as these simply use the normal 1-year time step within the Solvency II framework. As part of a long-term stochastic model, this becomes a more important consideration as there are many forms of operational risk that include costs spread over many years. This could include, for example, compensation for mis-sold policies.

The final consideration here is whether each operational scenario could happen more than once and whether an occurrence affects the future likelihood. As an example, an operational risk event representing losses due to cybercrime may be able to occur more than once. However, there may be a reduced likelihood after the first instance on the expectation that a firm takes steps to reduce the risk in the future.

Using these considerations, an operational risk model could be constructed based on causal dependencies between risks rather than the statistical relationships assumed with copula simulation models.

7.9. Allowance for New Business Volumes

New business volumes may not be considered to be a risk in the same way as things like market risk or mortality risk. Any risk of excessive new business volumes giving rise to a cashflow strain is mitigated by the flexibility firms have to restrict volumes or increase prices to control them.

Low new business volumes certainly have the possibility of causing difficulties to a firm's business plans, and new business volumes may be related to:

- Economic factors.
- Product cost.
- Market competition.
- Company reputation.
- Advertising spend.
- Regulatory and legal changes.

Some of these factors can be included within a risk model. However, others are not well suited for stochastic modelling. For this reason, it may be preferable that new business volumes are modelled deterministically rather than including a stochastic element. Alternatively, a simple stochastic model could be designed taking into account some of the main elements above.

7.10. Risk Aggregation

Within risk models used under Solvency II Internal Models, risk aggregation is normally carried out through the use of copula modelling. These models use a copula to combine risk movements based on an assumed statistical relationship between risk movements over a single time period. Key inputs to the parameterisation of the copula include the assumed correlation between risk movements.

It is well known that correlation does not imply causation. The copula models make no inferences on the underlying cause or drivers of risk events. They instead simply model the statistical likelihood of events occurring together according to the assumed correlations.

By contrast, the long-term stochastic model described in this paper is a model based on understanding the true causation of risk events. Risk simulations may be generated using a number of independent random processes, together with the relationships developed within the model. These relationships mean that the correlation between different risk events still exists and may be estimated using simulation results. However, correlations are not an input to the modelling process.

7.11. Loss Model

7.11.1. Overview

The above sections describe how the risk model element of a long-term stochastic model may be developed. The sections below describe an approach to the construction of a loss model. The purpose of the loss model is to take the results of the risk model and use these to find the effect on key business metrics. These may include profits from assets and liabilities but also items such as elements of the Solvency II balance sheet.

Within the copula models commonly used within firm's Solvency II Internal Models, the loss model is normally used only to find the effects of risk movements on assets and liabilities (in order that these may be used to determine the SCR) rather than wider aspects such as elements of the Solvency II balance sheet.

These copula models are single-period models (normally based on a 1-year time frame). For this reason, they only need to generate a single output point (the gain/loss over the 1-year period), with the inputs being based on risk movements from the same single period. The loss model is therefore a function that takes as its input a vector of risk movements (over the 1-year period) and produces a single number representing the loss or gain over that period.

Within a long-term stochastic model, a key difference is that rather than inputs and outputs being based on a single period, the input variables are instead time series which are in turn used to produce output time series. For example, a simple model (based on just equity risk) could take as an input a time series of equity returns and use these to produce a time series of profits on a particular product.

The following sections describe how an appropriate loss model may be constructed.

7.11.2. Liability cashflows

A loss model for the liability cashflows needs to be able to take time series inputs in respect of risks such as mortality, lapse or market risks and be able to use these to produce an output time series of liability cashflows. This is essentially a discounted cashflow model. It may be quite simple (e.g. in the case of an immediate annuity), or more complex (such as a savings product with investment guarantees).

Modelling of demographic risks such as lapse or mortality may be carried out at different levels depending on the purpose of the model and the computation power available. For example, if using a stochastic lapse risk model gave an assumed lapse rate of 5% for a particular time step and simulation, this could be applied as either a 5% lapse rate for all policies, or instead, each policy could be assumed to lapse with probability 5%. The latter approach here allows for lapse volatility risk. Such a risk could be important for a product with a small number of very large policies.

7.11.3. Asset model (cashflows and valuation)

To go with the liability cashflow model, a similar model is needed for assets. This needs to be able to produce both the value of the assets as well as asset cashflows (such as bond coupons or equity dividends). Of course, such asset models are readily available. The models are able to allow

for cashflows from the assets, as well as valuation of the assets using data such as interest rates, spreads and volatilities.

Where the asset and liability cashflow model results for a product are combined, this gives the overall profit/loss time series for the product.

7.11.4. Best Estimate Liability (BEL) reserves

The stochastic projection of BEL reserves is similar to that of liability cashflow projection. However, the key difference is that in addition to the cashflows at each step, the models need to be able to produce the BEL in accordance with the Solvency II regulations. There are a number of points to be considered within the calculation methodology.

7.11.4.1. Demographic assumptions

The demographic assumptions used will of course be a key part of the BEL calculation. The risk models described above are intended to be capable of stochastically modelling values such as the lapse and mortality rates. However, the BEL assumptions should not simply be set to align with the stochastically modelled demographic risks.

BEL assumptions reflect information known at the date of the assumption (including future projections such as assumed mortality improvements). If the BEL assumptions were to be aligned with the stochastically projected demographic risks, this would result in a model in which no demographic experience variances could occur. Such a model would of course not be realistic.

To make the BEL assumptions more realistic, they should be set taking into account a firm's assumption setting policy. For example, if a firm sets mortality rates as a percentage of a standard table based on the past 3 years of data, this method could be directly used to set the BEL demographic assumptions within each time step and simulation. Of course, this is likely a simplification from actual practice in that there would be an element of judgement included even where a mechanical rule is in place.

7.11.4.2. Discount rates

The basic discount rates used to value liabilities under Solvency II are the EIOPA swap-based rates. The simulation of swap rates and therefore the EIOPA curve would be included in a stochastic risk model. However, different discount rates are required for pension schemes and for business subject to the Volatility Adjustment (VA) or the Matching Adjustment (MA).

Firms' defined benefit pension schemes under Solvency II are required to be valued using the IAS19 discounting basis. Within this basis, the discount rates are assumed to be based on the yield of high-quality bonds (normally taken to mean AA rated). In order to allow for this within a BEL loss model, it is necessary to have a stochastic model of the IAS19 discount rates. These could, for example, be taken to be the swap rates plus the spread on AA rated bonds.⁹

In order to be able to value business subject to the VA, it is necessary to have a model that may be used to estimate the VA from market data. The VA is effectively calculated to be 65% of the spread on a representative portfolio of government and corporate bonds. The calculation may be used to construct a model of the VA depending on current market data. This may be a simple model applied at a high level or alternatively a more detailed model accurately aligned with the detailed methodology used by EIOPA for the published VA.

Unlike the VA, the MA is specific to a firm's own assets and in addition, is subject to qualifying conditions to test its suitability. Carrying out a full MA calculation within a simulation model is

⁹The IAS19 regulations also include further considerations such as with regards the depth and liquidity of the market for high-quality bonds.

likely to be unrealistic due to the complexity of the calculation. For this reason, it may be more pragmatic to use a model that allows for changes in the MA as a simple function of spread changes.

7.11.4.3. Options and guarantees

The valuation of options and guarantees under Solvency II is on a market-consistent basis. This is normally carried out through the use of risk-neutral simulation modelling calibrated such that the options and guarantees are consistent with market option prices. To use such risk-neutral simulations within a stochastic risk simulation model results in a “nested stochastic” problem within which the computational power required is likely to be very high. It is also not clear how risk-neutral simulations could be produced to represent future time periods or within different simulations. This is because risk-neutral models are calibrated to current observed market prices. Such prices of course are not observable for future periods derived through stochastic projections.

For this reason, it may be appropriate to use an approximate approach to the valuation of options and guarantees. Such an approach should give a reasonable approximation based on projected market data (such as interest rates, equity values, implied volatilities), but be available without the use of simulations. Techniques such as the use of Closed-Form Solutions (CFS) or replicating portfolios may be used for this purpose.

7.11.5. SCR model

The SCR model needs to be able to estimate the value of the SCR either on a Standard Formula or Internal Model basis as appropriate.

7.11.6. Standard Formula SCR

The Standard Formula SCR is calculated using an aggregation formula on the results of a series of individual risk stresses. The approach used to estimate the SCR may be simply to use the asset valuation and BEL models described above in order to calculate each of the stress valuations required for the Standard Formula. For example, the Standard Formula mortality stress may simply be calculated by applying the appropriate stress to the q_x before using the BEL model to derive the stressed BEL and hence the 1-in-200 capital requirement.

An additional consideration is that within the Standard Formula, the equity stress depends on past equity values (for the symmetric adjustment mechanism) and the interest rate stresses depend on the level of the curve. In order to allow for these, it is necessary to construct a simple model following the EIOPA methodology that may be used to calculate how these change under stress.

7.11.7. Internal Model SCR

Internal Model firms typically use a stochastic copula simulation model to derive the SCR (where firms use an aggregation formula approach then a similar method as has been described above for the Standard Formula may be used).

Where a copula simulation model is used, it is unlikely to be practical to expect that this could be applied over multiple time steps and simulations (such models may typically use around 1m simulations for an SCR calculation). For this reason, an approximation approach may be more appropriate. Such an approach could be, for example, use an aggregation formula approach based on 1-in-200 stresses. An aggregation formula approach is an approximation to the copula simulation model results. Such an approach could use an adjustment factor calibrated as the ratio of the time zero copula model result to the time zero aggregation formula result.

As an alternative, if sufficient computational processing power should be available, copula simulation modelling could be used at each time step and over each simulation. This could be done at a much lower number of simulations than the main SCR calculation (very high simulation volumes are often used to support capital allocation rather than being required for the SCR calculation).

To use such an approach, it is necessary to derive the proxy functions used within the SCR calculation for each time step and simulation. Recalibration of proxy functions is neither practical nor necessary. However, it is possible to adjust the existing time zero proxy functions. For example, by considering how individual stresses (at 1-in-200 level or another level) vary over time and by simulation, changes to proxy functions may be approximated. Where key areas of non-linearity exist, specific additional stresses could be applied in order to model this.

7.11.8. Risk Margin model

The calculation of the Risk Margin is based on a cost of capital approach taking into account the discounted value of future SCRs in respect of non-hedgeable risk. For this reason, it may be calculated at different time steps and over different simulations using a similar approach. The additional considerations are that:

- The Risk Margin model needs to allow for changes in the swap rates used to discount future SCRs.
- The Risk Margin calculation excludes the use of the MA and VA.

7.11.9. Transitional measures model

Solvency II transitional measures may be applied in the form of Transitional Measures on Technical Provisions (TMTP) or Transitional Measures on Risk-Free Interest Rates (TMRFIR).

TMTP are calculated using a comparison of Solvency I and Solvency II results to derive a transitional amount due to be run-off linearly over 16 years from 2016. The TMTP are recalculated every 2 years or more frequently should the risk profile of the firm be considered to have changed. TMTP are subject to the Financial Resources Requirements (FRR) test within which the TMTP are restricted, where the sum of technical provisions, non-technical liabilities and capital requirements are lower under Solvency II than under Solvency I.

Modelling of the TMTP over different time steps and different simulations are challenging as it requires the calculation of results on both the Solvency I and Solvency II bases. Therefore, consideration should be given to the materiality of TMTP to the firm, the extent it could change under a recalculation and the likelihood of an FRR restriction. Where modelling of TMTP changes is considered appropriate, this could take a simple form (such as simply allowing for how changes in Risk Margin affect TMTP), or it could take a more sophisticated form allowing for other key areas of methodology difference between Solvency I and Solvency II.

The TMRFIR allows for an adjustment calculated as a portion of the difference between the rate that applies under the current regime and the Solvency II discount rate and reduces linearly over the 16-year period. This may be modelled using an allowance for the differences in interest rates used in discounting.

7.11.10. Liquidity modelling

Although capital modelling normally takes on a much greater significance within firms, it is also important for firms to ensure they have sufficient liquidity to make payments as necessary. Liquidity modelling is not well suited to the single-period models, 1-year models commonly used within the industry. It is, however, something that can be accurately incorporated within a long-

term stochastic projection model. This is because such models allow for all cashflows associated with a firm's assets and liabilities (including the stochastic variation in these) rather than simply the current value of those assets and liabilities.

In order to allow for liquidity considerations within a stochastic long-term model, it is necessary first to include a robust measure of a firm's liquidity. This measure may then be included in the model taking into account the effects of asset and liability cashflows throughout the projections. The model could incorporate management actions designed to improve liquidity should it fall below target levels.

8. Example Model

8.1. Purpose

This section shows the use of a model to demonstrate the techniques discussed in this paper. The model is intended for illustration purposes. It does not reflect the extent of complexity within firms' actual business and uses relatively simple risk models with approximations for a number of areas.

8.2. Model Overview

8.2.1. Firm

The example is based on a proprietary firm that administers a simple annuity product only. The liabilities are backed by gilts with free assets invested in cash and equities. The firm calculates its capital requirements using the Solvency II Standard Formula. The firm is closed to new business¹⁰.

The firm does not make use of transitional measures, the MA or VA.

8.2.2. Annuities

The business consists of 100,000 annuities paying £10k p.a. paid annually in arrears. The annuitants are assumed to be females aged 60.

8.2.3. Expenses

The expenses of managing the business are initially assumed to be £300 per policy per annum, together with fixed costs of £5m p.a. Both elements of the expenses are assumed to increase with inflation.

8.2.4. Assets

The firm holds fixed interest assets (gilts and swaps) assumed to match the cashflows of the liabilities. Assets of £1bn Equity and £1bn cash are assumed to be held in excess of the value of the liabilities.

8.2.5. Hedging Strategy

The firm hedges the interest rate exposure of the liabilities at all times. Where changes occur (such as if expenses were to increase), the hedge would be rebalanced with additional assets taken

¹⁰In practice, it would be normally be expected that a firm paying dividends to shareholders would not be closed to new business.

from cash and equity where necessary. Rebalancing is assumed to be free of transaction costs with suitable assets for hedging assumed to always be available.

8.2.6. Equity strategy

The firm varies the extent of equity investment depending on the strength of its capital position. This is measured by the Capital Coverage Ratio (CCR) calculated as the ratio of surplus to the SCR. The following table shows the allocation used:

CCR%	Equity	Cash
>170%	50%	50%
<170%	20%	80%

8.2.7. Dividends

The value of the firm is measured primarily by its Own Funds. The firm, therefore, pays annual dividends based on a proportion of its Own Funds according to the following table:

CCR%	Dividend % of Own Funds
>200%	2%
160–200%	1%
<160%	0%

No further dividends are assumed should the value of the Own Funds ever fall below zero.

8.2.8. Firm objectives

The firm's key objectives are to:

- Increase the expected present value of the dividends.
- Reduce the variability (measured by the standard deviation) of the present value of those dividends.
- Avoid regulatory insolvency (CCR% falls below 100%).

The dividends are valued using a discount rate of 5% (reflecting the firm's cost of raising capital).

8.3. Risk Model

8.3.1. Overview

The firm is exposed to market risks, expense risks and mortality. In keeping with the approach described in this paper, these risks are implemented through stochastic models capable of simulating outputs over long-term periods consisting of multiple steps. The model is based on the use of a single interest rate curve rather than allowing for differences between gilt rates and the Solvency II discount curve.

The parameters used in each element of the model are set out in Appendix A.

8.3.2. Interest rates

Interest rate risk is modelled using a single-factor Vasicek model simulated using monthly time steps.

$$\Delta r = \alpha(b - r)\Delta t + \sigma\varepsilon\sqrt{\Delta t}$$

where r is the rate of interest

α is the “pullback”

b is the equilibrium

σ is the volatility

t is the time.

8.3.3. Equity

Equity risk is assumed to follow a lognormal model.

$$\text{Ln}(S) \sim \text{Normal}(\mu, \sigma^2)$$

where S is the equity price over a 1-year period;

μ is the mean of the log return;

σ is the standard deviation of the log return.

Note that this model is less sophisticated than the models typically found within ESGs. It is important to consider this as a model for illustration purposes. In practice, it would be preferable to use a firm’s existing real-world ESG.

8.3.4. Inflation

Inflation is not included with the scope of stochastically modelled risks. It is, however, assumed to vary over time according to the rates set out in Appendix A.

8.3.5. Mortality

The base table used for mortality rates is the CMI S3 series based on amounts (normal health). Stochastic mortality is assumed through the use of a zero mean random walk applied to the force of mortality at all ages.

$$\mu_x(t + 1) = \mu_x(t) + X_t$$

where $X_{t+1} - X_t \sim \text{Normal}(0, \sigma^2)$.

As for the other risk models, this is a simple stochastic model for illustrative purposes. In practice, it would be preferable to use one of the many more sophisticated stochastic mortality models available.

8.3.6. Expense

Expenses are assumed to increase with inflation as per the table set out in Appendix A. The level of expense is assumed to be affected by a multiplicative factor M . This factor is defined as a random walk with zero mean.

$$M_{t+1} = M_t + X_t$$

where $X_{t+1} - X_t \sim \text{Normal}(0, \sigma^2)$.

8.3.7. Aggregation

The above risks are all assumed to be independent. This is of course a simplification within the model as in practice, there would be a non-independent relationship between the risks (particularly equity and interest rate risks).

8.4. Modelling Approach

8.4.1. Overview

The model is intended to capture the key features of the exposure while making use of approximations in a number of areas.

8.4.2. Projection period

The projection period is 40 years. At the end of this time, very little liability exposure remains.

8.4.3. Time steps

The model is based on the use of annual time steps.

8.4.4. Simulations

Each projection is based on 1,000 stochastic simulations of the risk model. It is noted that a small element of simulation error remains with the results, which could be mitigated with the use of a greater number of simulations. The projections have been based on 1,000 simulations for practical purposes.

8.4.5. Base liability cashflows

The model has been constructed to be capable of allowing for the liability cashflows each year, which is required for annuity and expense payments. However, a key modelling assumption is that fixed interest assets are continually rebalanced to match the liabilities. Therefore, each liability cashflow is matched by a corresponding asset cashflow. The fixed interest asset cashflows are therefore not required to be explicitly modelled. Where cashflows differ from those expected (e.g. due to increased expenses), this results in an increased reserve (together with SCR and RM) rather than a direct cashflow impact. The increased reserve results in additional costs as cash or equity assets are sold to fund the hedge rebalancing.

8.4.6. Reserves (Best Estimate Liability or BEL)

The BEL at each time step is calculated through a discounted cashflow approach. The BEL is calculated as the discounted value of all future cashflows. The discounting is carried out using the term structure of interest rates specific to each time step and simulation.

The BEL calculations need to use an assumed mortality specific to each time step and simulation. Mortality rates are assumed to reflect the base table, multiplied by a factor calculated to reflect actual versus expected experience over the previous 3 years. This is intended to be similar

to the experience setting process used at firms, which may be based on consideration of actual versus expected experience over previous years.

The calculations also need to use an assumed rate of expenses. The rates of expenses for each time step and simulation are set equal to the (simulated) expenses in the previous year. Expenses are then assumed to increase with the rates of inflation set out in Appendix A.

8.4.7. Assets

The assets comprise gilts, equity and cash. Given the assumption of fixed interest being used to hedge the liabilities, the gilts will always have a value equal to the BEL. The remaining assets are the equity and cash. Equity is assumed to grow in value according to the lognormal model described in section 8.3.3. above (modelled as a total return rather than allowing for dividends). Cash is assumed to have zero return.

Cash and equity assets are reduced by dividends paid out. They are also affected by the cost of rebalancing hedges, where expenses and mortality experience differ from reserving assumptions.

Cash and equity assets are rebalanced in each time step taking into account the rule set out in section 8.2.6. above.

8.4.8. SCR

The SCR is assumed to be calculated according to the Solvency II Standard Formula. The appropriate risk modules to consider are those for longevity, expense and equity (interest rate exposure is zero due to the assumed hedging).

Longevity risk is calculated allowing for the 20% reduction in q_x assumed within the Standard Formula stress. A full discounted value calculation of the reserve under this stress is carried out. The SF longevity stress is then set to be the difference between this value and the BEL.

The SF expense stress includes an increase both to the level and inflation of expenses. A simplification used within this model is that only the (10%) level increase is applied. This is calculated using a full discounted cashflow approach.

The SF equity stress includes a symmetric adjustment mechanism within which the level of stress varies depending on past movements in equity markets. A simplification used within the model is that this is not included and the SF stress is modelled instead as a 40% fall. The equity SCR can then be directly calculated based on the value of equity at each time step and simulation.

Given the calculated SCRs in respect of longevity, expense and equity, the SF aggregation formula is used to calculate the overall SCR. No allowance is made for the operational risk element of the SII SF.

8.4.9. Risk Margin

The Solvency II Risk Margin is calculated as 6% of the discounted value of projected SCR in respect of non-hedgeable risk (taken to be the life risks in the model). This is modelled using a run-off factor approach. Within this approach, the SCR in respect of life risks only is assumed to run-off in line with the BEL. The future SCRs are then discounted using the yield curve specific to each simulation and time step allowing for the 6% cost of capital.

8.4.10. Key results

Having derived the above elements of the SII balance sheet, the Own Funds can be calculated simply as the assets, less BEL and RM. The CCR% can be calculated based on the surplus as a percent of SCR.

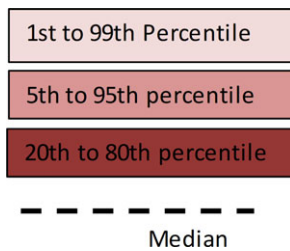
In addition to viewing the Own Funds and CCR%, key results include the expected present value of dividends, the probability of breach of 100% CCR (beginning of regulatory intervention) and the probability of Own Funds breaching zero (no transfer value of the business).

8.4.11. Liquidity

Liquidity is a key risk for insurers to consider and this form of the model is well suited to modelling the amount of liquidity at each time step. However, liquidity considerations are not included within the scope of this demonstration model.

9. Example Model Results

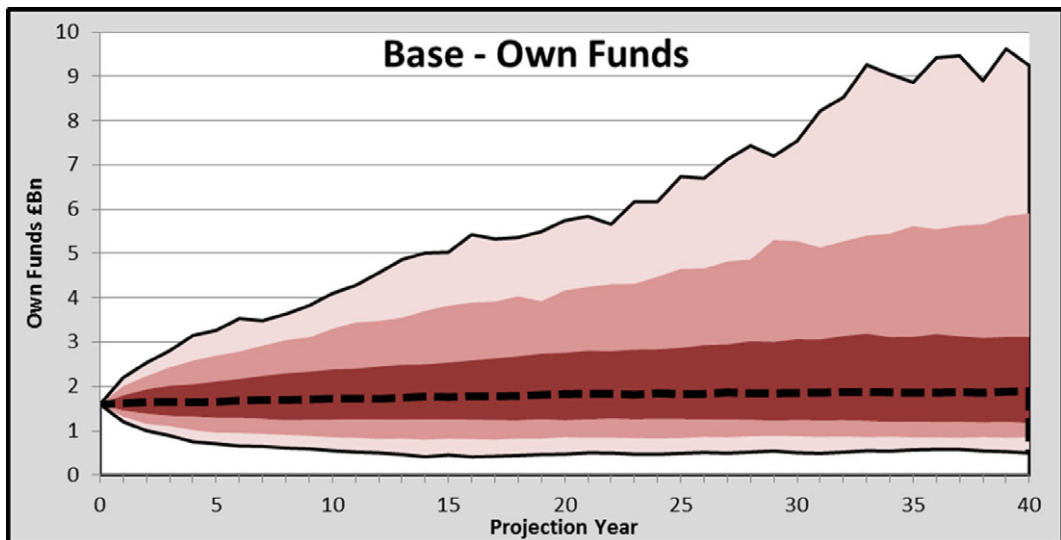
9.1. Key



9.2. Base Results

9.2.1. Own Funds

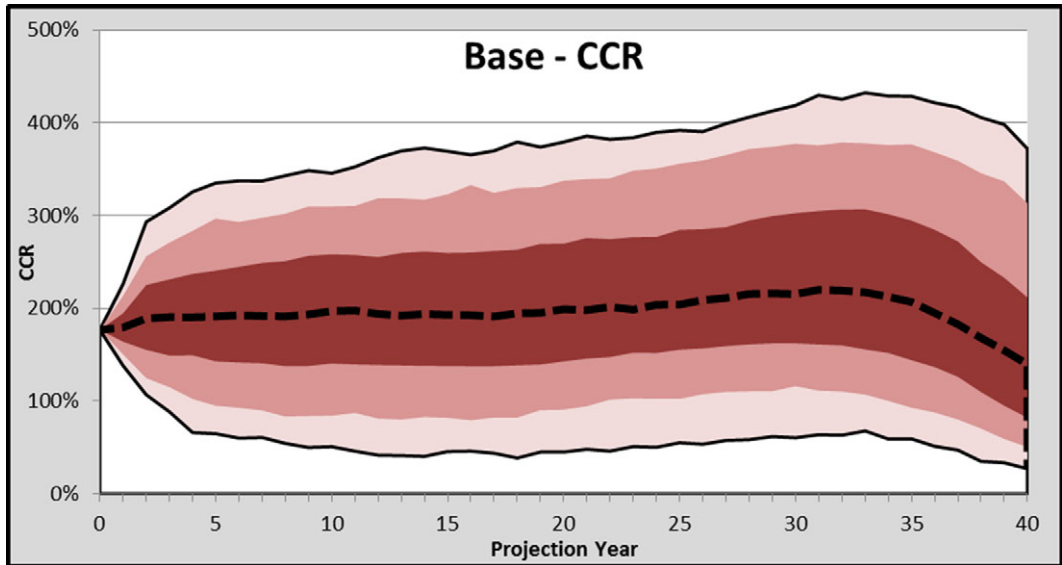
The following chart shows the projected Own Funds.



The chart shows that at 99% confidence (the lowest line on the chart), the Own Funds are not expected to become negative. The results show the median level of Own Funds is stable across the projection with the potential for significant increases in the later years.

9.2.2. CCR%

The following chart shows the projected CCR%:



The results show an expected spread of possible outcomes over the initial few years, but beyond 5 years, the projection results show a similar spread of results (amount of risk). This represents the possibility of different outcomes affecting results over the early years, but with correcting actions (such as restrictions on equity investment, dividends) stopping results from spreading further over time. The results also reflect the extent of mean reversion in the Vasicek model used for interest rates. This prevents the results from spreading further out over the term of the projection.

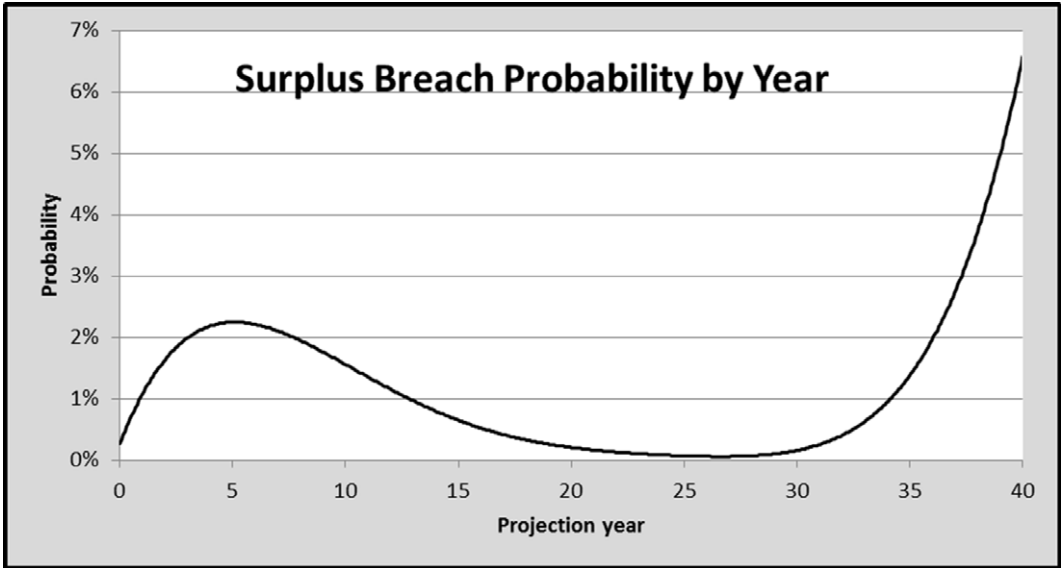
The results show that while the median projection gives a healthy level of CCR, there are a significant proportion of scenarios that result in CCR coverage below 100%.

9.2.3. Breach of SCR by year

The graph in the next page shows the likelihood of breach of SCR in each projection year.

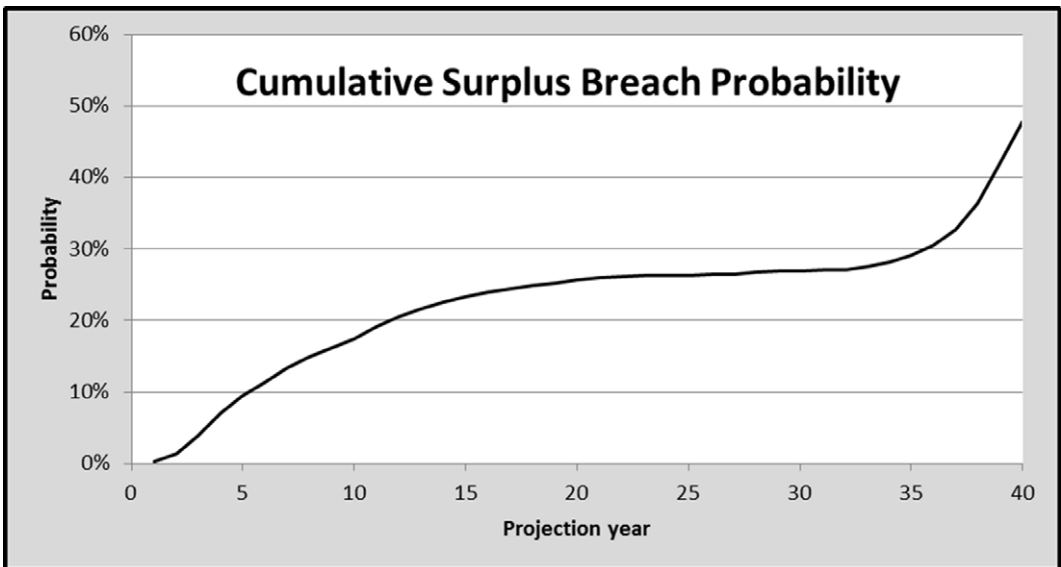
The results show a high risk of a breach in SCR at the later stages of the projection. However, by this stage, much of the insurance risk has run-off with just equity risk remaining. It would perhaps be unrealistic for this to occur without a change in company strategy.

In the earlier stages of the projection, it can be seen that the level of risk of surplus breach peaks around year 5 before reducing in the following years. This reflects the likelihood that, if the early years can be survived, sufficient levels of free assets will have been accumulated to withstand further shocks.



9.2.4. Cumulative probability of SCR breach

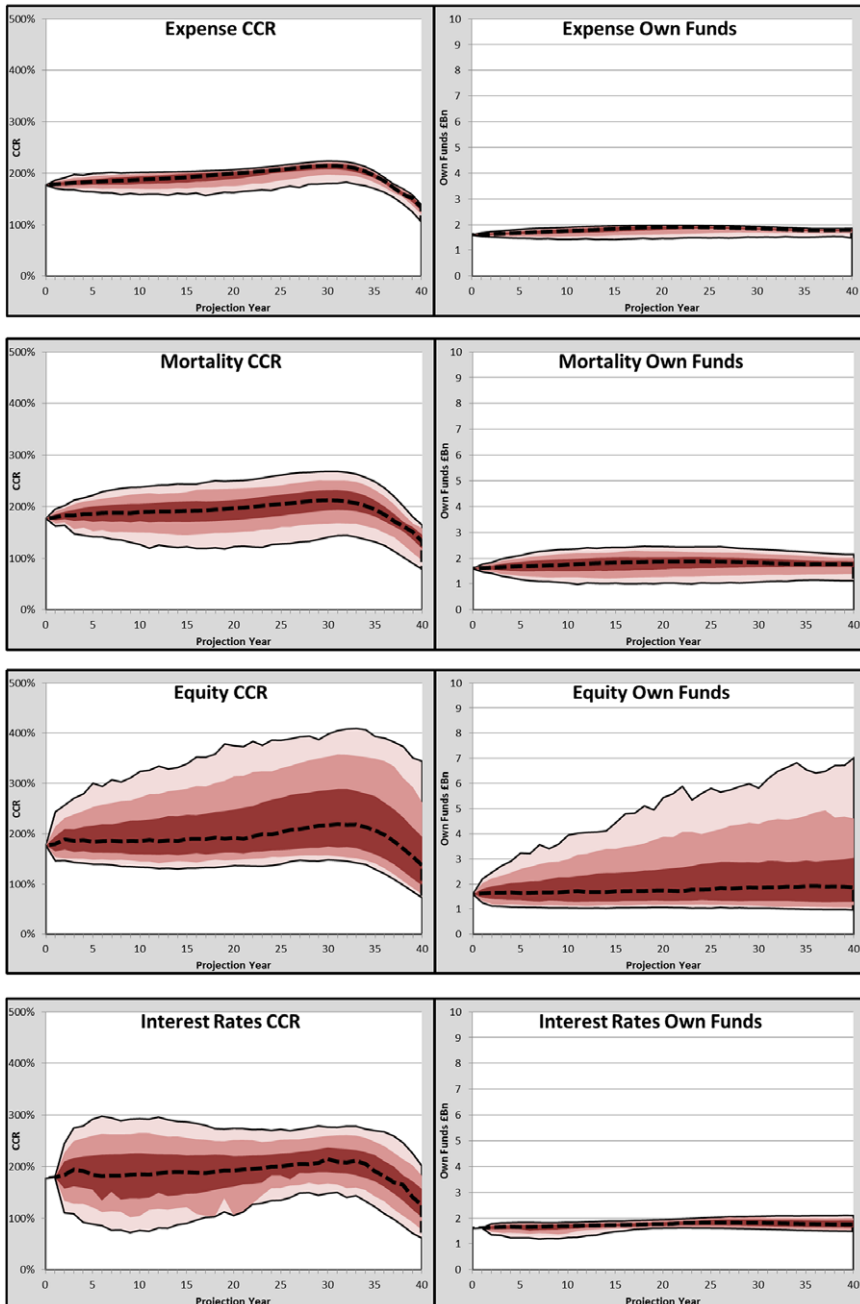
An alternative way to consider the risk of SCR breach is to consider the cumulative probabilities as shown in the graph below:



The results displayed in the graph may be used to answer questions regarding the long-term probability of company failure (defined here by breach of surplus). Such questions are not possible to answer with the single-period copula models commonly used within the UK.

9.2.5. Results by risk

The following graphs show the variation in CCR and Own Funds due to individual risks. The results can be used to see the contribution of each risk to the overall risk to the firm and to see how each risk varies over time.



9.2.6. The results provide insights such as the following:

- Before the fund gets very small (around year 30), individual risk events are unlikely to result in a breach of SCR (100% CRR). Should such a breach occur, it is likely to result from multiple risks. This is particularly something to note for risks such as mortality and interest rates for which the effect of a combined event can be much worse than the sum of the individual events (due to non-linearity).
- Interest rate risk has only a small effect on Own Funds, but it can have a much bigger effect on CCR% as it materially affects the value of the SCR. The risk is high in the early years but reduces throughout the projection.
- Equity risk can give significant variation in Own Funds and CCR. However, the amount of downside risk is restricted (due to the EBR and dividend actions).
- Expense risk is not highly material. The risk is not symmetric with a greater potential downside than upside.

9.3. Strategy Testing

9.3.1. Overview

A key benefit of long-term stochastic risk modelling is that, in addition to looking in detail at existing plan projection, it is also possible to test the effects of different strategies in detail. As an example, three different alternative strategies are set out below.

9.3.2. Option 1 – Outsourcing

The firm is considering outsourcing its administration costs. A deal has been negotiated such that another firm takes over all of the expenses, in return for a payment of the BEL of those expenses plus £60m.

9.3.3. Option 2 – Increased level of dividends

Within this strategy, the level of dividends is increased by 50%.

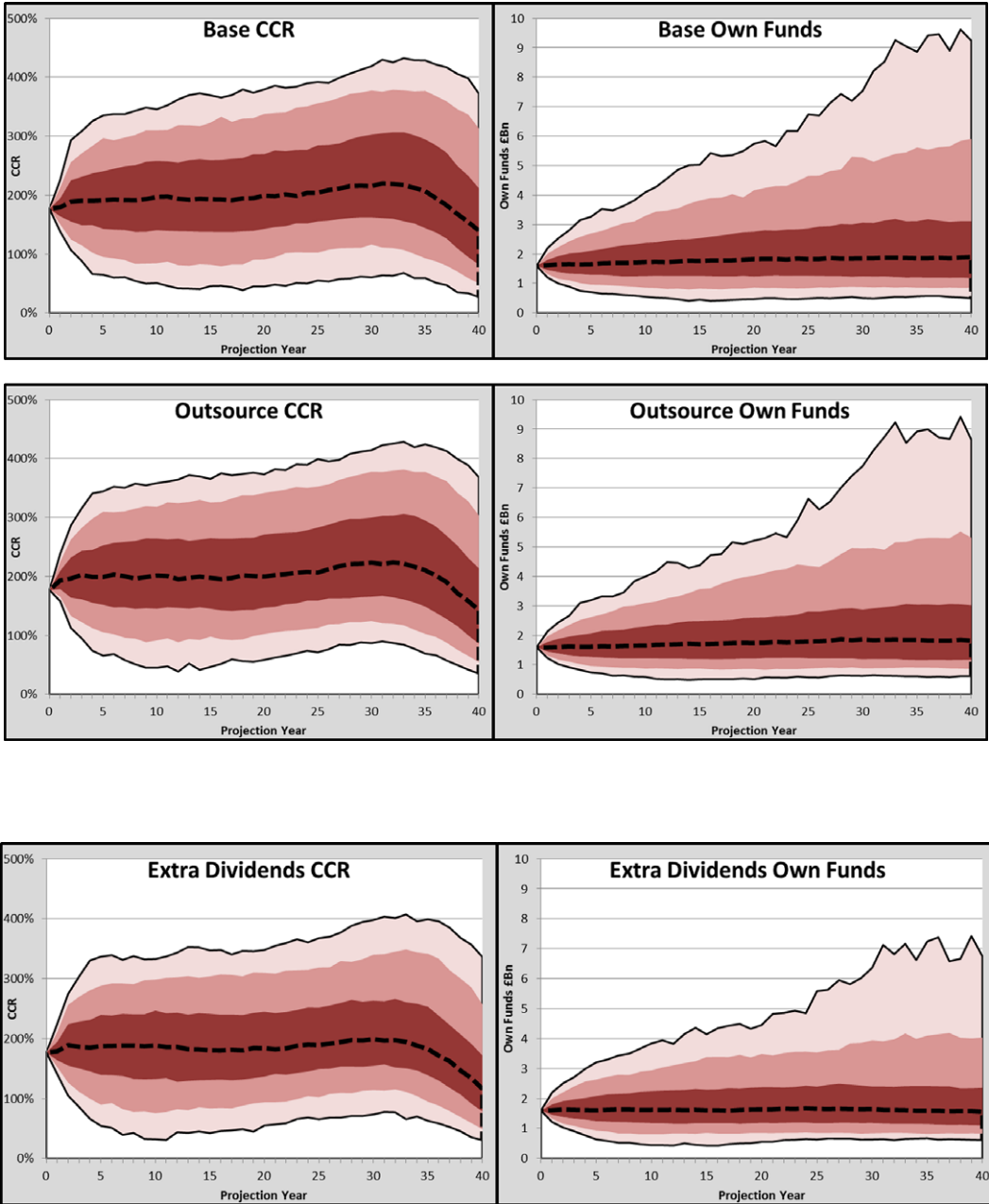
CCR%	Dividend % of Own Funds	
	Base	Option 2 Strategy
>200%	2%	3%
160–200%	1%	1.5%
<160%	0%	0%

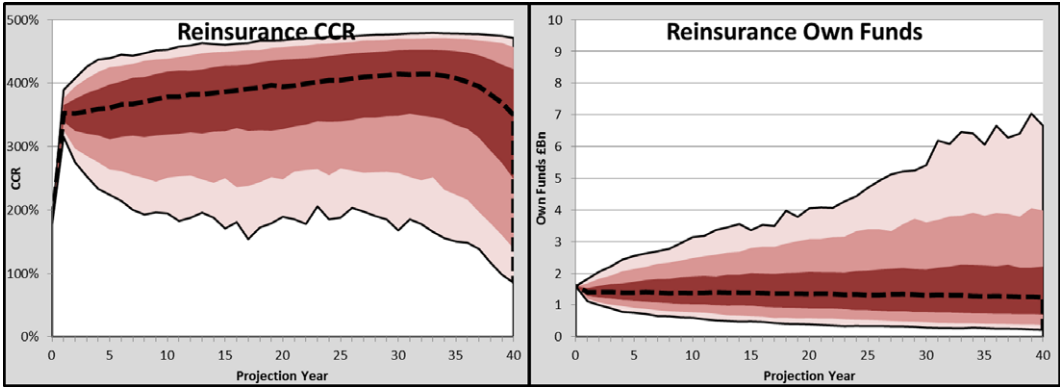
9.3.4. Option 3 – Reinsurance

In this strategy, a reinsurer agrees to take 90% of the annuity liabilities in return for the BEL of the amount reinsured plus £500m.

9.3.5. Own Funds and CCR

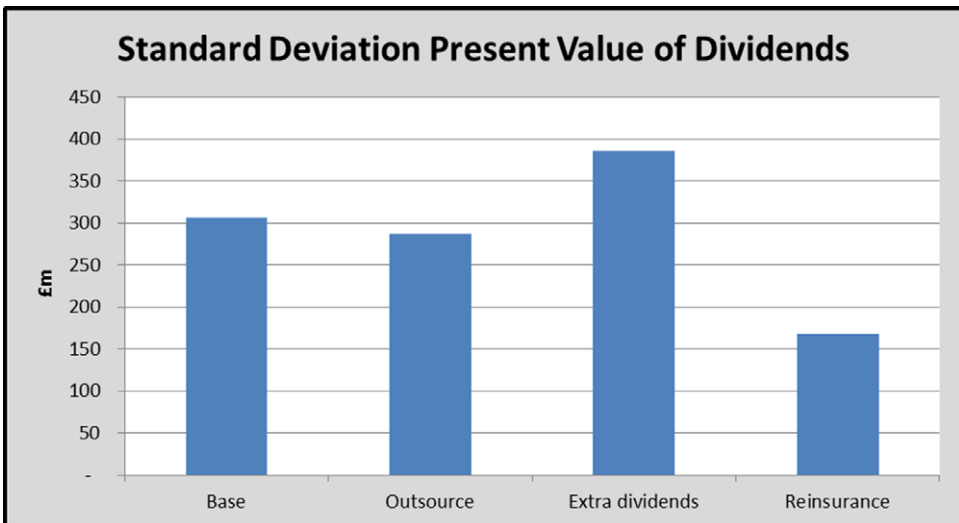
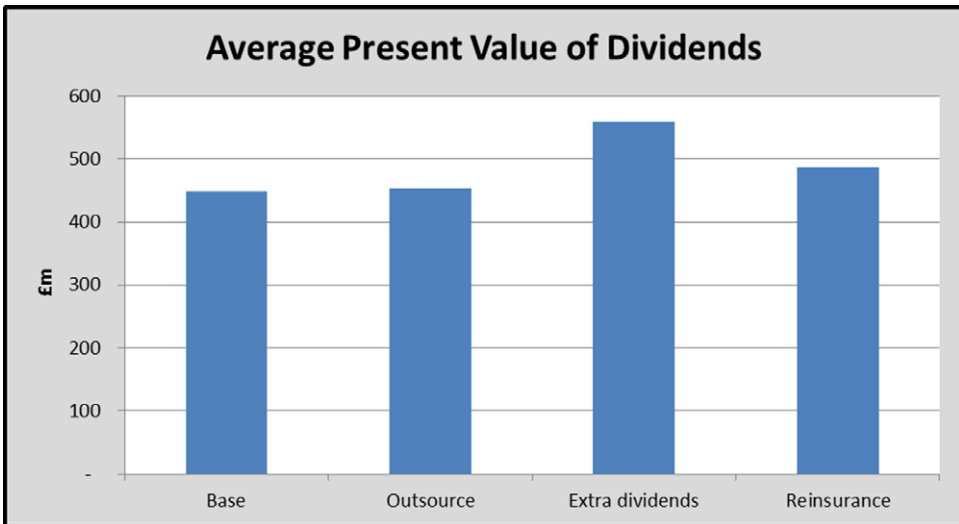
The following graphs show how the Own Funds and CCR projections are affected by the strategies.

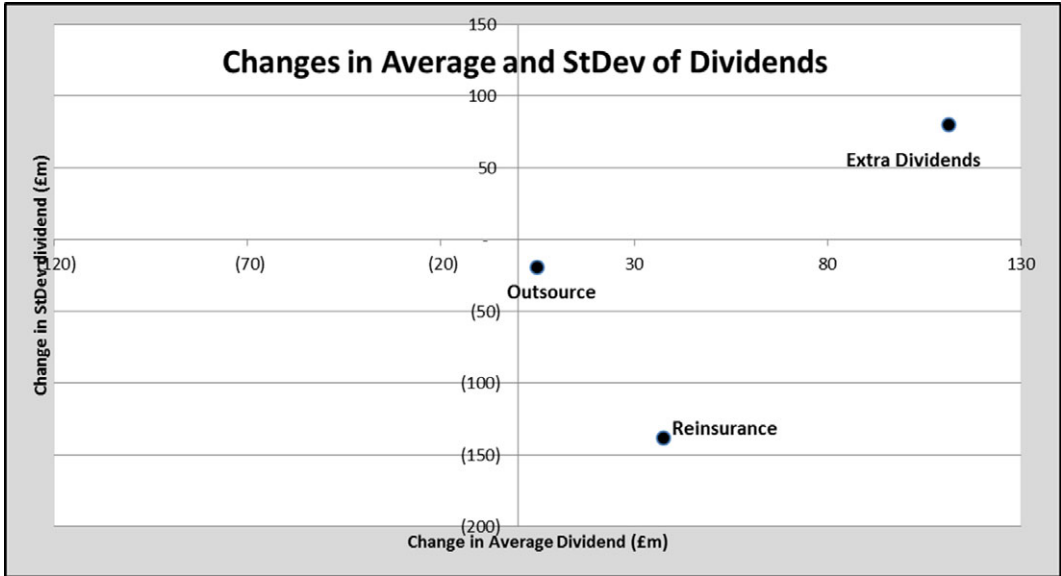




9.3.6. Present value of Dividends

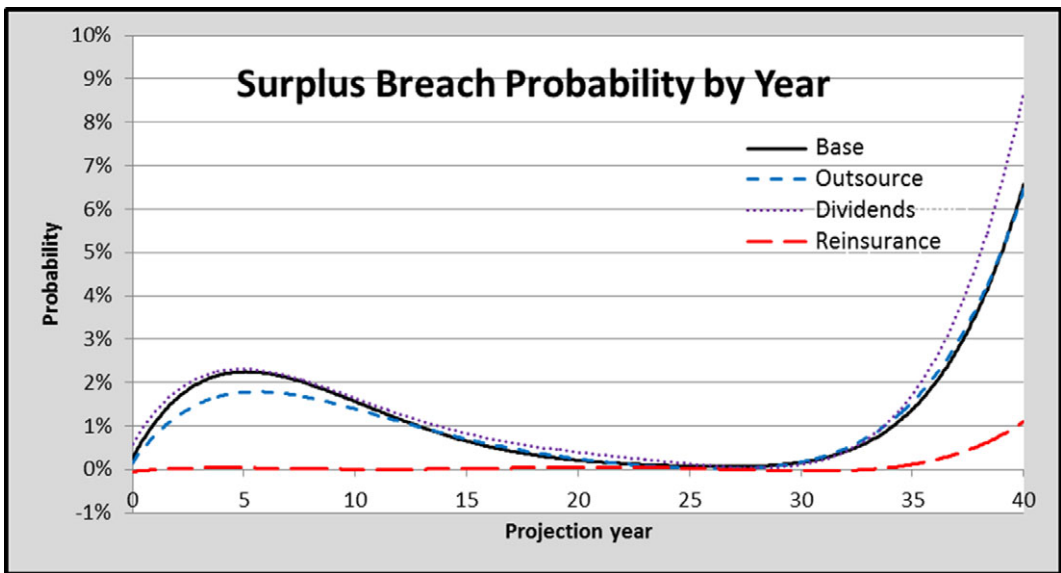
The following charts show how the different strategies affect the present value of dividends.





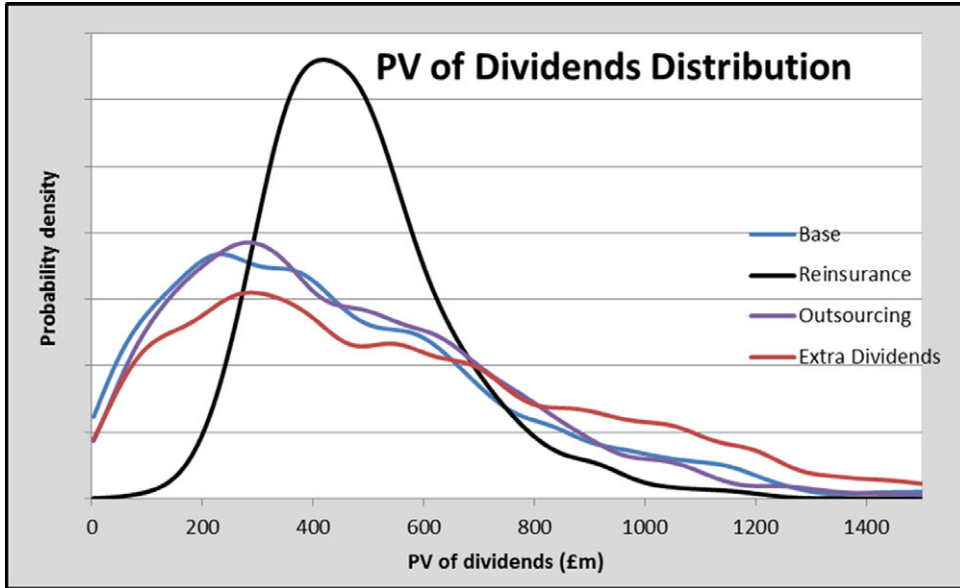
9.3.7. Surplus Breach

The following graph shows how the probability of a surplus breach by year is affected by the strategy.



9.3.8. Dividend distribution

The estimated probability density functions for the present value of dividends for the strategies discussed are set out below:



9.3.9. The graph shows that:

- The reinsurance strategy gives a significant reduction in downside risk, at the cost of also reducing the potential for high dividends.
- The outsourcing strategy has only a small effect on the results
- The extra dividends strategy gives a much greater probability of high dividends, but at the cost of significantly higher downside risk.

9.3.10. Analysis

The above graphs provide some examples of the kind of analysis that is possible using long-term stochastic models. Noting that the aims of the firm were to increase the average present value of dividends, reduce the standard deviation and avoid insolvency, some of the key insights from these results are as follows:

- The outsourcing scenario is beneficial but of low materiality. It gives a small increase in the average dividends and a reduction in their standard deviation. It reduces the risk of a CCR breach in the early years.¹¹
- The extra dividends scenario gives an additional expected value of dividends but at the cost of extra standard deviation. It has a small detrimental effect on the risk of a CCR breach.
- The reinsurance scenario gives the benefits of both higher expected dividends and a reduction in their standard deviation. It also almost entirely removes the risk of a surplus breach.

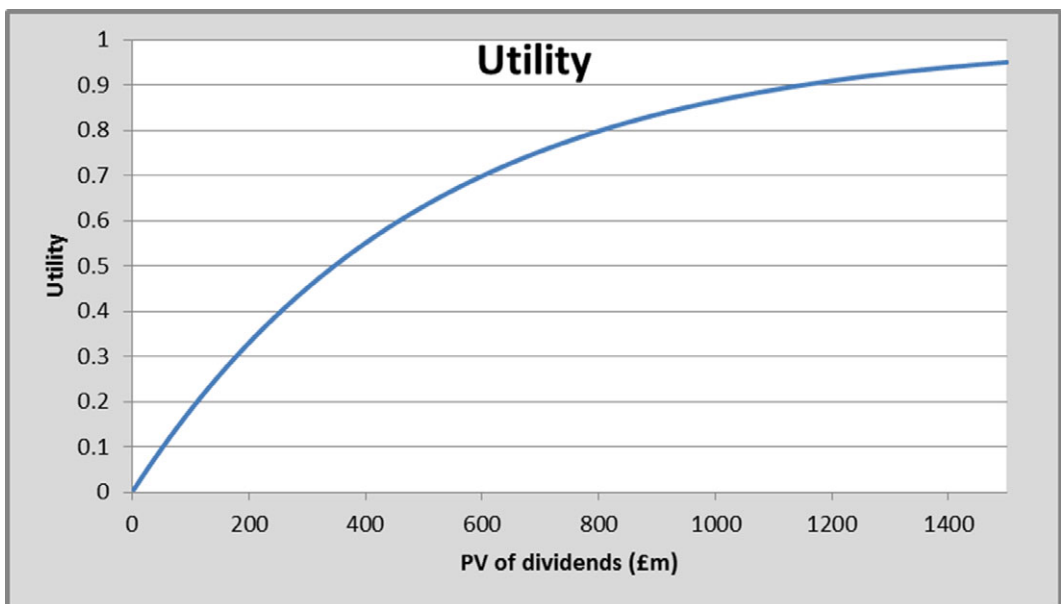
¹¹It is noted that the model does not allow for any additional operational risks that may arise in this scenario. Such model limitations need to be considered as part of any decision making process.

- 9.3.11. Of the different strategies, it appears that the reinsurance strategy is the strongest. However, there are of course other considerations. These include discussion of the limitations of the model and the risks that aren't included (particularly counterparty risk in this case). The firm may also consider applying more than one of these strategies.
- 9.3.12. Strategy optimisation

In this example, the firm's objectives were set out in high-level terms. If it can be expressed in precise terms, it may be possible to identify an optimal strategy. In the demonstration example, the company clearly needs to find a strategy that gives the expectation of high derivatives. However, this must be balanced against the risks to these dividends. One way this can be formalised is through the use of a utility function (Schoemaker, 1980).

- 9.3.13. A utility function may be used to map from a particular outcome (in this case, the PV of dividends), to a utility value. The utility value represents the firm's preference for this outcome. The utility function should have a concave shape to reflect the expectation that the firm would be risk averse. As an example, the following function could be used:

$$U(x) = 1 - e^{-0.002x} \text{ (an exponential utility function).}$$



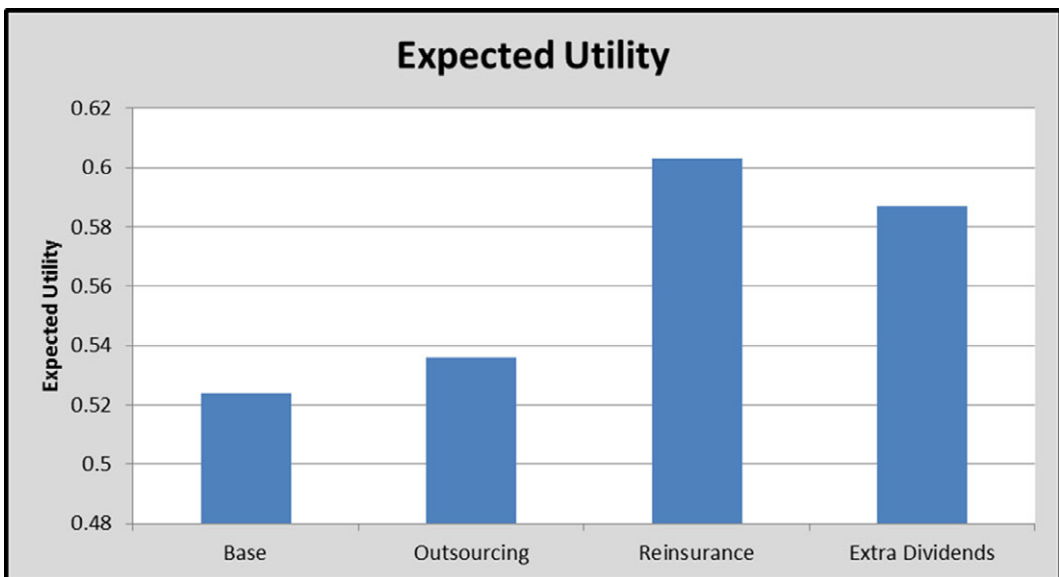
- 9.3.14. Given the distribution of the present value of dividends under a given strategy, it is possible to calculate the expected utility of that strategy. The strategy may, therefore, be optimised by varying the input parameters in order to maximise the expected utility.
- 9.3.15. By applying the utility function within the model, the expected utility of each strategy can be calculated as

$$E[U(x)] = \int_{-\infty}^{\infty} f(x)u(x)dx$$

where $f(x)$ is the probability density function of the PV of dividends
 $U(x)$ is the utility function $U(x) = 1 - e^{-0.002x}$.

9.3.16. The expected utility under each of the strategies may, therefore, be estimated from the simulation results to give the following:

Strategy	Expected Utility
Base	0.524
Outsourcing	0.536
Reinsurance	0.603
Extra Dividends	0.587



9.3.17. Therefore, of these different strategies, the reinsurance strategy gives the greatest expected utility and may be considered as the strongest strategy (other factors should of course also be considered). Note that in this example, the utility is a function of the present value of dividends only. It would be possible to have a function that also took into account other factors such as the possibility of breaching the SCR.

For example,

$$U(x) \text{ is the utility function } U(x) = I(x) \cdot (1 - e^{-0.002x})$$

where $I(x)$ is an indicator function such that

$$I(x) = 1 \text{ if there has been no breach of SCR after 30 projection years} \\ = 0 \text{ otherwise.}$$

The function assigns a significant utility penalty should there be a breach of SCR.

- 9.3.18. Where the firm looks to optimise the strategy, this should be done by considering all the different variables available. Actions such as outsourcing the administration costs are effectively arrangements with a binary value (they either go ahead or they do not). For others, there is a continuous range of different values that could be used. For example, strategic actions on the EBR and dividend payments could take a continuous range of values.
- 9.3.19. The optimal strategy for a firm therefore can be derived as the strategy that maximises the expected utility over all possible values of the input strategic variables. The above example demonstrates how this may be achieved with binary variables that represent whether a particular strategy item is being used or not. More generally, strategy optimisation may also be used for continuous variables such as the level of EBR to be used.

10. Conclusions

10.1. Summary

- 10.1.1. Within this paper, it has been demonstrated how a long-term stochastic model can be constructed by combining a number of existing components with some additional elements under a single framework. Such models are already in use within the industry but commonly limited to the stochastic modelling of market risks only. The paper shows the benefits that a long-term model can bring. It shows the advantages such a model has over the traditional deterministic models used for business planning purposes or the single-period copula simulation models used within the Solvency II framework.
- 10.1.2. A key factor that affects the model design is the availability of suitable computational power to support it. Historically, this has certainly been the case with, for example, discounted cashflow valuation models coming into widespread use as computers were introduced into insurers. Today, the most recent available processing power again gives a step change in capability compared to what was available a decade ago. This gives further opportunity to develop new modelling techniques whether those are long-term stochastic models as discussed in this paper, or another form of model (perhaps full nested stochastic models without the need for proxy modelling).
- 10.1.3. The opportunity to develop the next generation of models gives great opportunity within firms to understand the nature of their risks and exposures in an increasingly sophisticated way and to use this increasingly in ORSA and business planning models.
- 10.1.4. A remaining question is how the regulatory framework of the future could look as increasingly sophisticated modelling capabilities arise. The significant regulatory change in the UK in 2004 with the introduction of realistic balance sheets was possible because the computation power had become available to facilitate stochastic models. As the next generation of actuarial models is developed, could this provide a drive to a new form of regulatory capital requirements?

- 10.1.5. The existing Solvency II Pillar 1 framework is very focussed on a current valuation of business using market-consistent techniques, with capital held against changes in that value. Some of the criticisms of this approach are that:
- Market-consistent valuation doesn't work well where markets don't exist (e.g. property implied volatilities).
 - Markets don't act rationally. They are prone to bubbles and crashes.
 - It gives rise to procyclicality. For example, if equity markets crash and volatilities spike, this makes firms' regulatory solvency much worse therefore prompting them to sell equities¹².
 - While the concept of a Risk Margin is fully justified, it is highly questionable whether any form of Risk Margin reflects the transfer value of non-hedgeable risks with any accuracy.
- 10.1.6. It may be preferable for a future solvency regime to instead be based on the use of long-term stochastic models. Such a regime could have at its core the requirement that a firm can meet its liabilities to policyholders with a specified (high) probability. This would give firms a greater incentive to focus on long-term value rather than a short-term regulatory position. It would also appropriately allow for the long-term nature of insurance risks. For example, it would be able to take into account the risk of, for example, 10 years of economic stagnation or the risk of long-term changes in mortality. Such risks are much more relevant within life insurance than the short-term changes incorporated in the 1-year framework key to Solvency II.
- 10.1.7. Finally, it would avoid a key problem area within market-consistent techniques – valuation within stressed conditions. In a stressed market situation, it is likely that:
- Implied volatilities spike.
 - Equity and bond markets could crash.

These effects may only last for a very short period. However, their effects can be significant enough that an insurer could be in breach of its SII capital requirement in that period (such effects for the industry can be procyclical). Solvency II contains measures designed to mitigate these effects. For example, the VA (intended to counter “artificial volatility”) and the equity symmetric adjustment mechanism. Such measures in effect take the Solvency II framework away from its underlying principle of using market-consistent valuation.

- 10.1.8. Through the use of long-term stochastic modelling instead of market-consistent valuation, such problems would not exist. Insurers would not need to be concerned about the short-term valuation of assets and liabilities under stress conditions. Instead, they could more appropriately consider the long-term requirement to meet policyholder liabilities with a high level of confidence. To put it another way, in a market stress situation, firms could focus on their responsibility to pay policyholders rather than focus on a hypothetical stressed transfer valuation of the business.
- 10.1.9. As a final point, while the contents of this paper focus on a number of advanced modelling techniques, it is important not to place sole reliance on models. Key business decision making should be carried out through the support of models to provide insights to complex problems. For this purpose, it is imperative that the limitations of the model are understood and where possible, sensitivity testing or other such analysis is used to give further information.

¹²The Solvency II Standard Formula includes the symmetric adjustment mechanism on equity risk to mitigate this risk.

Acknowledgements. The author would like to thank Tim Williams and Jack Silman for their contributions to this paper.

References

- Artzner, P., Delbaen, F., Eber, J.M. & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3), 203.
- Cairns, A.J.G., Blake, D. & Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, **73**(4), 687–718.
- Cox, J.C., Ingersoll, J.E. & Ross, S.A. (1985). A theory of the term structure of interest rates. *Econometrica*, **53**(2), 385–407.
- Jorion, P. (2006). *Value at Risk: The New Benchmark for Managing Financial Risk* (3rd ed.). McGraw-Hill.
- Lee, R.D. & Carter, L.R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, **87**(419), 659–671.
- Schoemaker, P.J. (1980). *Experiments on Decisions under Risk: the Expected Utility Hypothesis*.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, **5**, 177–188.

Appendix A. Risk Model Parameters

Interest Rates and Inflation Initial Forward Rates

The following table shows the initial interest rate and inflation forward curves used. Inflation is modelled deterministically and so no changes are assumed to occur in the rates below.

Duration (years)	Interest Rates	Inflation
1	5.0%	2.0%
2	5.0%	1.9%
3	4.9%	1.9%
4	4.9%	1.8%
5	4.8%	1.7%
6	4.7%	1.6%
7	4.6%	1.5%
8	4.5%	1.4%
9	4.4%	1.2%
10	4.2%	1.1%
11	4.1%	0.9%
12	3.9%	0.8%
13	3.8%	0.6%
14	3.6%	0.5%
15	3.5%	0.4%
16	5.0%	2.0%
17	5.0%	1.9%
18	4.9%	1.9%

Interest Rate Vasicek Model

Changes to interest rates are modelled through the use of a Vasicek one-factor model with the following parameters:

Parameter	Description	Value
r_0	Initial Rate of interest	5%
α	Pullback	7%
β	Equilibrium	5%
σ	Volatility	1%

Equity

Equity returns S are assumed to follow a lognormal model.

$$\ln(S) \sim \text{Normal}(\mu, \sigma^2)$$

Parameter	Description	Value
μ	Mean log return	2%
σ	Standard deviation of log return	20%

Expenses

The expenses (both per policy and overheads) are assumed to vary according to a lognormal distribution. This can be expressed as

$$E_{t+1} = E_t S_t$$

where E_t is a factor applied to the base expense assumptions

$$E_0 = 1$$

$$\ln(S) \sim \text{Normal}(\mu, \sigma^2)$$

Parameter	Description	Value
μ	Mean change	0%
σ	Standard deviation of changes	5%

Mortality

Mortality risk is modelled through the use of a series of factors μ_t , which is applied as a multiplier to the base table at all ages.

$$\mu_t = \mu_t + X_t$$

where $\mu_0 = 1$

$$X_{t+1} - X_t \sim \text{Normal}(\mu, \sigma^2)$$

Parameter	Description	Value
μ	Mean change	0%
σ	Standard deviation of changes	0.8%