The Potential Impact of Multi-Year Dependencies on the Design of the Solvency II Risk Margin
Malcolm Kemp, 30 November 2020

http://www.nematrian.com/docs/MultiYearDependenciesRiskMargin20201130.pdf

Abstract

This paper explores whether features of risks present in insurers' balance sheets could justify the inclusion of time varying features in the specification of the cost of capital parameter used in the calculation of the Solvency II risk margin. It argues that, on average, some attenuation through time for this parameter can likely be justified for Solvency II regulated insurers, particularly if the stresses for risks used by the Solvency II standard formula Solvency Capital Requirement (SCR) are believed to provide reasonable pictures of the multi-year dependencies expressed by these risks. For insurers using an internal model to determine their SCR, the attenuation used could form part of the specification and implementation of the internal model.

1. Introduction

1.1 In December 2019, the Actuarial Association of Europe published its commentary paper "A Review of the Design of the Solvency II Risk Margin", AAE (2019), to assist EU policymakers in their review of the Solvency II Directive. The paper reviewed the current design of the Solvency II risk margin, the aim of which is to provide an estimation of the cost a hypothetical third party would expect to charge (in addition to the Solvency II '*best estimate liability*') to take on a book of insurance liabilities.

1.2 One observation made by the paper was that multi-year dependencies could lead to situations that should ideally be tackled using a term-dependent cost of capital rate within the risk margin calculation that reduced (e.g. for mass lapses) or rose (e.g. for asbestos like liability) for increasing terms. At present the risk margin calculation does not include such a feature and instead involves a flat term structure for the cost of capital rate, as the calculation involves the following formula:

$$RM = CoC \cdot \sum_{t \geq 0} \frac{SCR(t)}{\left(1 + r(t+1)\right)^{t+1}}$$

where $SCR(t)$ is the projected (non-hedgeable) SCR at time t which a hypothetical reference entity would need to establish if it took over the liabilities, $r(t)$ is the Solvency II-specified (time-dependent) risk-free discount rate and CoC is the selected cost of capital rate, currently a 6% for all terms.

1.3 Early consultation papers contributing to the Solvency II 2020 Review published by the European Insurance and Occupational Pensions Authority (EIOPA) envisaged no change to the Solvency II risk margin calculation methodology. However, in EIOPA's 2020 holistic impact assessment (HIA) specification, a modification was proposed that in effect involved the cost of capital falling as the term, t, increased, see EIOPA (2020). The formula it proposed was:

$$RM = CoC \cdot \sum_{t \geq 0} \frac{SCR(t) \times \lambda(t)}{\left(1 + r(t+1)\right)^{t+1}} \quad where\ \lambda(t) = \max(\lambda^t, 0.5)\ and\ \lambda = 0.975$$

1.4 Such a formula in effect involves a time-varying cost of capital rate (or a corresponding adjustment to the projected SCR) that starts at 6% p.a. for a zero term, falls through time at 2.5% p.a. compound for c. 27 years until it reaches 3% p.a., staying at that level thereafter.

1.5 In this paper we comment further on this proposed modification to the risk margin calculation by amplifying comments previously made in the AAE commentary paper. The focus of this paper is not to explore whether a 'base' cost of capital rate of 6% is too high or low (or whether it should somehow alter in different economic conditions) as some industry bodies have, we understand, sought to do and as was separately commented on in the AAE commentary paper. Instead, our focus is to explore what might be plausible shapes that could potentially be justified for the $\lambda(t)$ in the formula in Section 1.3, based on arguments similar to those set out in the AAE commentary paper.

1.6 This paper adopts the following structure. Section 2 summarises the rationales set out in the AAE commentary paper for why a time-varying cost of capital rate might be justifiable and what, a priori, might be plausible time variation shapes. However it does not attempt to quantify any specific rates of decline or increase that might be involved. Section 3 attempts a more precise quantification of what rates of decline or increase might be justified for different sorts of risk that an insurer might bear. It focuses on a market consistent derivation of the risk margin and adopts the simplifying assumption that autocorrelation features for the risk in question can be correctly inferred merely by considering the wording of the relevant stress for that risk included in the current Solvency II standard formula SCR specification. Section 4 comments on whether this simplifying assumption is reasonable, i.e. whether the results articulated in Section 3 seem reasonable in the context of the inherent nature of the risks in question.

1.7 Conclusions drawn include:

(1) If the SCR equates to the magnitude of the unexpected one year loss an investor seeks compensation for via an assumed cost of capital charge (as is implicit in the Solvency II risk margin calculation) then a market consistent risk margin can be derived by assuming a risk-neutral probability equal to the cost of capital rate for the occurrence in a given year of a suitably sized unexpected loss. Solvency II aims for market consistency in how it sets its technical provisions and the Solvency II risk margin forms a part of these technical provisions. Ideally, therefore, the risk margin actually used within Solvency II should align where practical with such a market consistent risk margin.
(2) The risk margin can then be determined by constructing a (binomial) tree of potential unexpected loss outcomes, as long as it is possible to project what might be the unexpected loss arising in each future year conditional on the pattern of such losses in earlier years.
(3) Adopting some simplifying assumptions, the appropriate formula for the risk margin can then be solved analytically in many cases. It takes a form akin to that in section 1.3 for some suitable $\lambda(t)$. The relevant $\lambda(t)$ depends on how the $SCR(t)$ might be expected to vary depending on the pattern of unexpected losses arising prior to t for the relevant risk(s) under consideration.
(4) Whilst a flat or even increasing shape for $\lambda(t)$ may be justified for certain risks, on average a declining $\lambda(t)$ seems more justified, if averaged across the entire insurance industry and across all types of risk that the industry faces that are within the direct remit of current regulatory capital requirements. This average reflects the typically longer durations of life versus non-life policies and the greater contribution to the

industry aggregate risk margin coming from life insurance business, where justification for a declining rather than an increasing $\lambda(t)$ seems typically stronger.

(5) An upper limit on the attenuation that seems practically capable of being justified is $\lambda(t) = (1 - CoC)^t$, i.e. $\lambda(t) = 0.94^t$, with the 6% p.a. cost of capital rate as currently specified. However for most risks, including most life insurance risks, a slower decline seems more theoretically justifiable, particularly if the relevant stress is only a small fraction of the assets or liabilities exposed to the risk in question.

(6) The analysis does not support the inclusion of a floor of 0.5 on $\lambda(t)$ as is currently included in the HIA specification.

(7) If introduction of a common $\lambda(t)$ is considered undesirable, insurers using internal models to identify their SCR could be allowed to expand their internal models to incorporate features linked to $\lambda(t)$. Consideration could also be given to introducing an undertaking-specific parameter targeting $\lambda(t)$ for firms not using such internal models. Firms could also be asked to analyse the appropriateness of the risk margin calculation alongside the appropriateness of the SCR calculation in their own risk and solvency assessment (ORSA).

2. Potential rationales for a time-varying cost of capital rate

2.1 The AAE commentary paper noted that there were two main ways in which multi-year features of insurance risks (and how they might be financed) could directly justify the use of a time-varying cost of capital rate:

(1) The emergence of uncertainty in the insurance liabilities could exhibit some time-dependency and could be correlated through time. For example, there are some types of risks insurers face where there is some practical upper limit to what loss might arise. For these risks, if the overall term of the liabilities is T and nearly all the risk has emerged between $t = 0$ and $t = t_1$ then only a limited amount of the risk is capable of emerging between $t = t_1$ and $t = T$, even in cases where $t_1 \ll T$.

(2) The way in which the hypothetical reference entity receiving the insurance liabilities might price risks might in effect exhibit multi-year dependencies because of the existence of the so-called *shareholder put*. Shareholders of the reference entity can't be wiped out more than once by suffering losses from risks arising in the businesses in which they have invested (if they have suitably ring-fenced within a limited liability structure the reference entity from their other business activities). Thus emergence of large losses early in the life of the transferred book does, for these shareholders, reduce the amount of losses they might suffer later. It should be noted that this sort of multi-year dependency does not necessarily reduce the losses suffered by others (e.g. policyholders or any applicable insurance guarantee scheme) and does not depend on the risks themselves but rather on how they might be financed within a limited liability context.

2.2 An example given in the AAE commentary paper of a type of risk that exhibits the *negative autocorrelation* described in 2.1(1) is mass lapse risk. Suppose the mass lapse stress targeted by the SCR involves a 40% lapse (and all contracts are profitable and exposed to mass lapse risk). In the current risk margin calculation design, the projected SCR is calculated assuming progression is along the trajectory of the best estimate before the mass lapse is assumed to happen. So for contracts with a very low assumed lapse rate but still deemed subject to mass lapse risk, the current risk margin computation in effect assumes third parties would price in capital requirements as if 40% of the initial book could lapse in year 1, a further 40% of the initial book could lapse in year 2, a further 40% of the initial book could lapse in year 3 etc. However, it is not possible for more than 100% of the initial book to lapse. So the current computation overcompensates the third party assumed to be taking over the liabilities. All other things being equal, if a mass lapse of a given magnitude occurs early on in the life of the contract grouping, future mass lapses of the same absolute monetary magnitude are likely to be rarer, since more of the initial book would by then have lapsed. Where negative autocorrelation applies, the emergence of a particular type of uncertainty early on reduces the likelihood or quantum of uncertainty that might arise later on[1].

2.3 Conversely, the commentary paper also noted that some types of risk might exhibit *positive autocorrelation*. For example, a liability scenario like asbestos has the feature that the first court verdict establishing the scenario can increase uncertainty in future technical provisions quite dramatically. An unexpected loss early on in the life of such a liability can therefore *increase* the uncertainty the book can be subject to later on.

2.4 A priori, all other things being equal, we might expect the ideal $\lambda(t)$ to decline as t increases if the risks are typically negatively autocorrelated as per 2.2, but to increase as t increases if the risks are typically positively autocorrelated as per 2.3 (both relative to any pattern we might a priori expect for a risk that is neither positively nor negatively autocorrelated). Superimposed in either case might be a likely more modest non-risk specific decline as t increases due to the non-risk-specific 'shareholder put' effect noted in 2.1(2).

2.5 The 'shareholder put' effect can be illuminated by assuming that the market-clearing return in excess of the risk free rate demanded by equity investors who put up a specified amount of capital to carry a unit amount of a certain type of non-autocorrelated risk but do not cap their downside to just this amount by using a limited liability structure is x% per annum on that capital (for simplicity we assume x is constant through time). Some outcomes for such a project will involve possibly large cumulative losses, offset (the investor hopes) by some outcomes involving large cumulative profits.

2.6 Equity investors who put up the same amount of capital and carry the same risk but via a limited liability structure (so putting only a limited amount of their overall wealth at risk) truncate the losses that they might suffer. Therefore, all other things being equal, limited liability investors ought to receive a higher return or conversely do not need as high a value of x to entice them to carry this risk. In a market consistent world as implicitly underlies Solvency II, this divergence should depend on the spread on debt that such investors would need to issue if they were to bulk up the capital they committed to the project by issuing debt (i.e. it should take into account the risk-neutral probability that they themselves default through such a structure). It should also depend on the capital structure of the vehicle carrying the risk and the relative priority of this debt vis-à-vis the customers impacted by the risk.

[1] Some commentators, e.g. some members of the Institute of Actuaries in Belgium, refer to this effect as a type of 'loss-absorbing capacity of the risk margin' (and by using such ideas derive results for the mass lapse stress similar to those given in Section 3.12(2)). This is because the risk margin would fall (more than otherwise expected) were such a loss to arise. The Solvency II Directive does include a loss-absorbing capacity of technical provisions (LACTP) within the computation of the SCR but the current wording of the Solvency II Delegated Regulation excludes the risk margin from the scope of technical provisions to which the LACTP relates. As noted in the AAE commentary paper, failure to include some such adjustment for the mass lapse stress can be shown in extreme circumstances to lead to demonstrably market inconsistent outcomes (in which the total capital that needs to be put up to support this risk exceeds the maximum loss the firm could possibly suffer from this risk). This paper can be viewed as seeking to put such ideas onto a firmer theoretical foundation and by doing so to inform better what if any $\lambda(t)$ might be most suitable to apply in practice.

2.7 Any non-flat term structure included in the cost of capital parameter due to this reason would potentially impact the calibration of the 'base' cost of capital level on which this term structure would be superimposed, as other firms in the market are in essentially all cases also limited liability corporations. Whilst Solvency II explicitly isn't a zero failure regime (e.g. its SCR is calibrated to a 1 in 200 year likelihood of the SCR being exhausted), policymakers still have a strong vested interest in designing regulation to keep very low the likelihood of a large loss being suffered by policymakers due to such failures. Moreover, it is unlikely to be politically expedient to draw attention to such a possibility, as to do so could undermine public confidence in the robustness of the firms being regulated. What term structure if any to include for this reason is also ultimately a socio-political matter, inherent in the nature of society allowing firms to be structured with limited liability, and therefore linked to how easy society wants it to be for firms to take advantage of these limited liability features. It is therefore assumed within this paper that the effect is sufficiently modest or sufficiently problematic from a political perspective not to warrant any specific adjustment to how the risk margin calculation might otherwise be specified.

2.8 Left unanswered is whether, ignoring the 'shareholder put' effect, the appropriate pattern for risks that are neither positively nor negatively autocorrelated should show an increasing, declining or flat pattern for $\lambda(t)$. Any such feature would not be insurer-specific, so it should be possible to identify from the behaviour of capital markets more generally whether investors demand such patterns. We are not currently aware of any evidence indicating that shareholders expect an increasing or a declining cost of capital rate for non-insurance risks, so for the remainder of this paper we have assumed that a constant cost of capital rate through time would apply for a risk that shows no autocorrelation through time[2].

2.9 Some commentators have argued that the cost of capital rate used by Solvency II should decline through time to reduce the 'excessive' size of the risk margin for long term insurance contracts and to avoid the risk margin introducing 'excessive' interest rate sensitivity for such contracts if they include implicit or explicit interest rate guarantees. Whilst the AAE commentary paper accepted that the size of the risk margin (and its volatility) was an important issue, it noted that merely because the figure was large did not necessarily mean that it was wrong. It noted that the apparently high sensitivity of the risk margin to interest rates seemed in part to be a manifestation of issues that led to the creation of Solvency II's Long Term Guarantees measures. So, ideally, this feature should be considered in conjunction with these measures. It is not the intention of this paper to consider this topic further. Instead, the paper aims to explore what insurance-risk-specific overlay(s) to the cost of capital rate might be most appropriate, taking as given any features it is agreed politically are appropriate to address issues within scope of the Long Term Guarantees measures.

3. Analysis of insurer-specific risks based on their standard formula SCR specifications

[2] We have not sought to explore further in this paper what would be the impact, if any, of potential changes to how Solvency II risk free rates might be defined, e.g. the possible introduction of own funds buffers, given the lack of clarity at the current time over whether and in what form such changes might take. This issue links to some extent to the question of whether the assumed cost of capital rate should vary according to economic conditions. This was a topic explored in the AAE commentary paper. As shown in Section 3, some dependency can be theoretically justified given that the cost of capital rate is an important driver of the theoretically correct $\lambda(t)$ (in the sense that if the theoretically correct cost of capital rate were to tend to zero then the theoretically correct $\lambda(t)$ would tend to 1).

3.1 In this Section, we assume that the relevant risk can be 'correctly' quantified by reference merely to the wording of the relevant SCR stress included in the current Solvency II standard formula SCR as set out in the current Solvency II Delegated Regulations, i.e. EU (2014) as subsequently amended. In Section 4, we consider whether this simplifying assumption results in plausible time dependency characteristics for the relevant risks given their inherent natures. For simplicity, we assume that quantum of loss scales linearly with respect to factors driving the risk and that the relevant Solvency II risk-free yield curve can be treated as if it is flat at zero at all time points. We also assume that if risks are not autocorrelated then a third party provider committing capital to protect policyholders against such risks would expect a flat through time cost of capital rate CoC.

3.2 Where needed, we assume that the book of business which is subject to the relevant risk involves liabilities for which the best estimate liability, $L(t)$, declines linearly through time over a period of T years. The assets (on which shareholders carry risk), $A(t)$, supporting these liabilities are assumed to move likewise, i.e. we assume that in the best estimate the present values of the liabilities and assets at time t take the form:

$$L(t) = \frac{t}{T} L_0 \quad and \quad A(t) = \frac{t}{T} A_0 = \frac{A_0}{L_0} L(t) = KL(t) \quad where\ K = \frac{A_0}{L_0}$$

3.3 A key insight for this Section is that the underlying valuation paradigm applicable to the Solvency II risk margin is a market consistent one. In a market consistent world, if a given risk is assumed to be correctly quantified by a loss of X between time t (in years) and $t + 1$ and to require a payment to a capital provider of $X \cdot CoC$ if the risk is borne by a third party for the following year then, all other things being equal, the risk-neutral probability of occurrence of the loss of X over that year is CoC (independent through time) and its risk-neutral cost is $X \cdot CoC$, discounted using a suitable risk-free yield curve. We can quantify a risk-neutral or market consistent risk margin ('market consistent RM') by summing these risk neutral costs. Solvency II targets a market consistent valuation for technical provisions, of which the risk margin forms a part. Ideally, therefore, the risk margin calculation actually specified within Solvency II should mirror this market consistent RM.

3.4 Therefore, the key issue for this Section is to identify what loss should be implicitly assumed for the projected Solvency II SCR stress for the risk in question. In some cases how to interpret this will be unclear, in which case we offer some possible alternatives.

3.5 To illustrate how we can use this insight consider first market risk. Certain types of market risk are specifically excluded from the calculation of the risk margin, so can be ignored for the purposes of this Section. For example, it is currently assumed that all interest rate risk can be hedged by the reference entity taking over the liabilities. Other market risks typically can or are excluded to the extent that they relate to hedge-able risk.

3.6 However, for any market risks that are deemed not hedge-able, the projected SCR used in the risk margin calculation usually involves the application of a stress test. For example, the stress test used for the property risk sub-module of the standard formula SCR involves a one-off 25% market value decline (at the valuation date) applied to property assets. If such a risk was deemed not hedge-able then the projected SCR would be determined by the impact that this would have on the insurer's own funds. Most non-hedge-able market risk is likely to relate to the assets on which shareholders carry the risk but it is convenient to expand the analysis at this point to cover the potential impact such a stress might also have on the liabilities. The sign of the stress is also potentially important, so we will assume that the relevant stress as applied to assets is S_A and as applied to liabilities is S_L, with a positive S_A (and S_L) corresponding to a decline in the underlying assets (and liabilities), and the fractions of the assets and liabilities to which the stress applies are F_A and F_L respectively (both assumed positive between 0 and 1). In principle, the signs of S_A and S_L can vary depending on how the relevant risk impacts the asset and liability side of the balance sheet.

3.7 On the asset and liability sides of the balance sheet, the relevant projected SCR stress amounts may be taken to be $A(k,t)F_AS_A$ and $L(k,t)F_LS_L$ for the relevant path dependent $A(k,t)$ and $L(k,t)$, where k indexes path dependent evolutions of the assets and liabilities in a risk-neutral world (assuming that the exposure fractions remain the same through time). The overall stress is the movement in own funds, i.e. the movement in assets minus liabilities, i.e. $A(k,t)F_AS_A - L(k,t)F_LS_L$.

3.8 Relatively straightforward is the picture at time 0. Implicit in most standard formula SCR market stresses is that the stress involves a one-off permanent market value decline. So in a risk-neutral world, there is a risk-neutral probability of $q = 0.06$ of a permanent decline of S occurring at $t = 0$, i.e. the assets and liabilities falling by $F_AA(0)S_A = KF_AL(0)S_A$ and $F_LL(0)S_L$ respectively. The overall stress is the movement in own funds, i.e. the movement in assets minus liabilities, so is $L(0)(KF_AS_A - F_LS_L)$.

3.9 The picture becomes more complicated thereafter. For example (if discrete yearly steps are assumed), there is a risk-neutral probability of 0.06 of a stress happening at $t = 1$ and:

(1) In paths where a stress event didn't happen at time 0, the assets and liabilities are $A(1) = KL(1)$ and $L(1)$ respectively and it is reasonable to assume that if the stress occurs at time $t = 1$ the assets less liabilities then fall by $L(1)(KF_AS_A - F_LS_A)$.
(2) However, in paths where a stress event happened at time 0, the assets will have already fallen to $A(1)(1 - F_AS_A)$ and the liabilities to $L(1)(1 - F_LS_L)$, so it is not clear what stress we should assume might happen in the risk-neutral world at time $t = 1$.

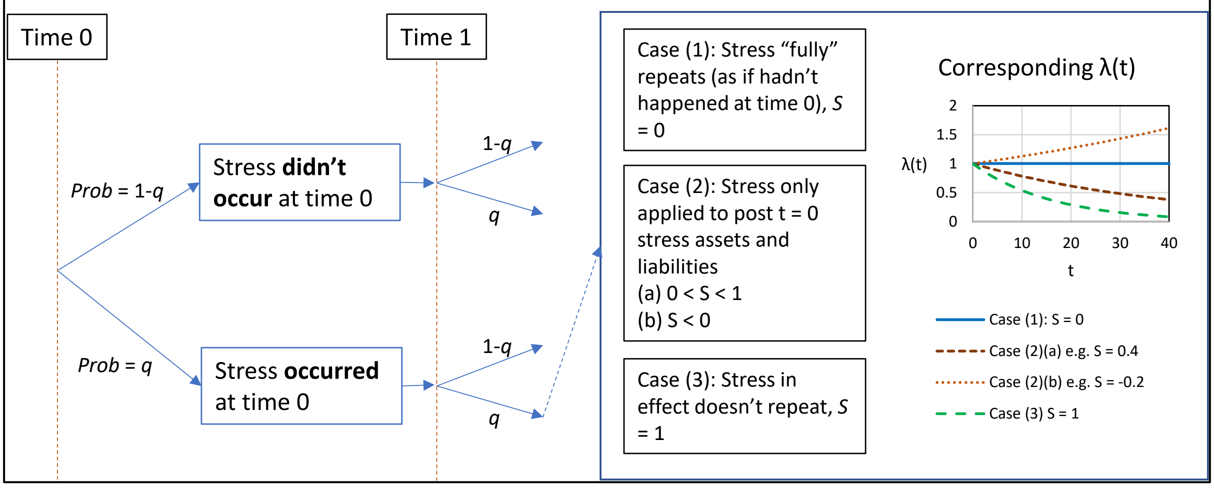
3.10 Possible approaches for the path in 3.9(2) include:

(1) We might assume the same absolute market value decline (attenuated only by the run-off of assets and liabilities through time) that happened at time $t = 0$ occurs again in this path, i.e. the assets and liabilities decline by a further $F_AA(1)S_A = F_AA(0)\frac{T-1}{T}S_A$ and $F_LL(1)S_L$=$F_LL(0)\frac{T-1}{T}S_L$ respectively. The stress to use would then be $F_AA(1)S_A - F_LL(1)S_L$. However, implicit in such an approach is that the decline that previously occurred at $t = 0$ somehow unwinds itself over the following year, so that it can be repeated again in full at $t = 1$. This seems implausible for most market risks, and is also arguably not how the standard formula stress is specified, since it assumes an instantaneous permanent decline in the market value of the relevant asset;
(2) We might assume that a further market value decline of S should be assumed to apply to the then present assets and liabilities at time $t = 1$, i.e. the risk margin should be priced as if the then applicable stress is $F_A(1 - S_A)A(1)S_A - F_L(1 - S_L)L(1)S_L$. This seems to be the most obvious interpretation of how market stresses operate in the standard formula, bearing in mind intuition on how markets move (which is that just because they have declined once does not mean that they cannot suffer further declines thereafter); or
(3) We might assume that no further market stress should be assumed at time $t = 1$, since relative to their originally expected level the assets have already declined to

$F_A(1 - S_A)A(1)$ and $F_L(1 - S_L)L(1)$, i.e. to levels that they would have fallen too had the only the time the stress happened been at $t = 1$. Where the SCR stress automatically applies to all future years a case could be made for this interpretation, but the justification seems weak for market risks, given the intuition noted above.

3.11 These possibilities are illustrated schematically in Figure 1 and can be thought of as prototypical of nearly all the risks included in the standard formula SCR. Essentially, the task is to identify which of the three cases mentioned above seems most applicable to the risk in question. If Case (2) applies, it also becomes relevant to identify the magnitudes and signs of S_A and S_L and, if these signs and magnitudes differ, the relative magnitudes of F_A and F_L.

Figure 1: Potential path dependency characteristics of stresses



3.12 This is because in most cases we can calculate analytically the $\lambda(t)$ that correspond to each of the above cases. The calculations are as follows, if q is the risk neutral probability of the stress occurring at any given year end (i.e. the cost of capital rate, following the line of reasoning given in section 3.3) and $SCR(t)$ is the projected value of the SCR as used in the current RM calculation:

(1) Full absolute value of stress applied at each year end as if no stress had happened at previous year ends

The market consistent RM calculation is then as follows, i.e. as per the current specification:

$$\sum_{t \geq 0} \frac{qSCR(t)}{(1 + r(t+1))^{t+1}} = CoC \cdot \sum_{t \geq 0} \frac{SCR(t) \times \lambda(t)}{(1 + r(t+1))^{t+1}} \quad where\ \lambda(t) = 1^t = 1$$

(2) Stress is applied to merely 'depleted' asset and liability values at a given year end, reflecting falls that would have happened at previous year ends in the path in question

Assuming $S_A = S_L = S$ (or the relevant F_A or F_L are zero so the stress is not applicable to one side of the balance sheet), the projected SCR stress in the current RM calculation is

$SCR(t)$. However, if the stress has applied n times before by time t in the particular path being considered then the risk-neutral stress amount is $(1-S)^n SCR(t)$ (before risk-neutral discounting). The risk-neutral probability of the stress happening if it has happened n times already at the previous t year ends is $qC(t,n)q^n(1-q)^{t-n}$ where $C(t,n)$ is the binomial coefficient, i.e. $\frac{t!}{(t-n)!n!}$. So the market consistent RM calculation is then:

$$\sum_{t\geq 0} q \sum_{n=0}^{t} C(t,n)q^n(1-q)^{t-n}(1-S)^n \frac{SCR(t)}{\big(1+r(t+1)\big)^{t+1}}$$

This can be simplified to:

$$CoC \cdot \sum_{t\geq 0} \frac{SCR(t) \times \lambda(t)}{\big(1+r(t+1)\big)^{t+1}} \quad where\ \lambda(t) = (1 - Coc \cdot S)^t$$

If S is positive (i.e. application of the stress reduces both assets and liabilities by the same fraction), Case (2) can be thought of as intermediate between Cases (1) and (3). Intuitively this can be explained by noting that if S is close to 0, application of the stress has very little impact on the assets and liabilities at future time points, so should be approximated by one in which the calculation assumes no stresses applied at previous year ends. Conversely, if S is close to 1, there will be little further assets or liabilities available to stress thereafter, so the situation approaches that for Case (3).

It should be noted that if S is negative then $1 - Coc \cdot S$ can be greater than one, i.e. the appropriate $\lambda(t)$ would rise rather than fall through time.

If S_A and S_L have different signs and/or magnitudes then a more complex expression applies, reflecting the relative contributions to $SCR(t)$ from the different sides of the balance sheet.

(3) *No stress applied at a given year end if the stress has already happened in that path at a previous year end.*

If the stress is applied at the year-end then its value will be $SCR(t)$. However, whilst the risk-neutral probability of it occurring at time zero is q, the risk-neutral probability of it being applied at time 1 is only $(1-q)q$ and at time t is $(1-q)^t q$. Hence the market consistent RM calculation is then:

$$\sum_{t\geq 0} (1-q)^t q \frac{SCR(t)}{\big(1+r(t+1)\big)^{t+1}}$$

This can be simplified to:

$$CoC \cdot \sum_{t\geq 0} \frac{SCR(t) \times \lambda(t)}{\big(1+r(t+1)\big)^{t+1}} \quad where\ \lambda(t) = (1 - Coc)^t$$

3.13 In Table 1 we set out for various types of risk an insurer might carry (that are included in the standard formula SCR) which of the above three Cases seems most in line with the specification for that risk given in the standard formula SCR computation. For short duration business, $\lambda(t)$ will be approximately 1 at all relevant durations. We therefore concentrate in this table on risks to which life insurers are subject, as their liabilities will typically have longer durations than those for non-life insurers.

Risk (SCR sub-module)	**Relevant case (and for case (2) likely range for S)**	**Justification**
Market risks	Case (2) various S (potentially in range c. +0.1 or +0.2 to c. +0.59)	Mostly assumed to be hedge-able. But if not, likely impact principally or wholly to asset side of balance sheet. Nearly all asset stresses are expressed just as downside movements (i.e. positive S). An exception is interest rate risk but this risk is currently specifically excluded from the RM. In principle currency risk can also be two-sided. Most stresses in range c. 0.25 (e.g. property) to c. 0.59 (maximum for Type 2 equity stress within the confines of possible values the symmetric equity adjustment can take) although the spread risk and market concentration risk and certain other risks may have smaller S or its equivalent.
Counterparty default risk	Case (2) various S (likely to be positive but typically smaller than for market risk)	In some jurisdictions and for some business models mostly assumed to be hedge-able, but if business reinsured or assets e.g. include a significant amount of residential mortgages then can be more significant. Stress more complicated, but probably typically lower per quantum of exposure than for market risks. Risk again specified principally in terms of downward movement in exposure
Mortality risk	Perhaps close to Case (3) and at least Case (2) with $S = 0.15$	Stress explicitly involves an increase in mortality rates at all future valuation dates (reducing liabilities). But specification less clear on whether stress should be assumed capable of repeating. Probably stress can repeat, i.e. compound, in subsequent years, so probably not fully in line with Case (3).
Longevity risk	Probably somewhere between Case (2) (but with negative S, i.e. $S = -0.20$) and Case (3)	Stress involves a decrease in mortality rates (thus increasing future liabilities for relevant policies), but again is specified to

		apply to all future time points if it occurs.
Disability-morbidity risk	Close to Case (1) and no higher than Case (2) (with positive S of c. 0.2 to 0.35) except for that part relating to recovery rates which might be closer to Case (3)	Stress typically reduces liabilities to which future stresses might apply but stresses are mostly limited to the following 12 months' experience. An exception is the recovery rate as its stress is in respect of all years thereafter as well.
Expense risk	Arguably close to Case (3)	Expense stress framed as an instantaneous increase of 10% in expenses plus a further 1% pa thereafter, so spans whole future time horizon. If the stress has already applied at e.g. time 0 then at time 1 the expenses would already be taken as 11% lower (and increase at a further 1% pa thereafter) which is beyond the stress that would then apply at time 1 if that was the first time the shock struck. Conversely, the specification is less clear on whether stress should be assumed capable of repeating.
Revision risk	Close to Case (1)	Even if some elements of Case (2) applied, stress involves a 3% change to benefits, i.e. close to no change
Mass lapse risk	Case (2) typically $S = 0.4$ (but higher for some institutional business)	Applied only to profitable business. If stress applied then these profits deplete. Business cannot 'un-lapse'.
Lapse up	At least Case (2) with a positive S but typically lower than 0.4 but probably not Case (3)	Stress is applied to all future lapse rates. Application of stress reduces amount that might lapse in future. In principle repeated application of such a stress might compound through time, but this does not seem to be envisaged by the standard formula SCR wording. Effective S likely to be smaller than for mass lapse.
Lapse down	Somewhere between Case (2) (but with a negative S although typically smaller in absolute size than 0.4) and Case (3)	Akin to Lapse up, except that a decline in the lapse rate will typically increase future liabilities. Again size of stress likely to be smaller than for mass lapse.
Operational risk	Case (1)	Standard formula seems to focus on operational risks expected to arise

		in the coming year, without any indication that an event this year will influence what might happen in future years.

3.14 It therefore appears that for nearly all risks relevant to life insurers, a risk-neutral RM computation focusing on how the stresses are specified in the standard formula SCR would have a $\lambda(t)$ that is either flat or declining through time, arguably in some cases by close to CoC per annum, although in many cases by a materially lower rate. For operational risk, little or no decline in $\lambda(t)$ seems applicable given how that part of the standard formula SCR is specified[3].

3.15 The AAE commentary paper noted that some types of risk might intrinsically be expected to exhibit positive autocorrelation through time, an example being certain forms of liability risk. The standard formula SCR for the liability risk sub-module of the non-life risk module is set out in Article 133 of the Solvency II Delegated Regulation. This refers to the application of a risk factor (dependent on the liability risk group type) to the premiums earned by the insurance or reinsurance undertaking during the following 12 months. To the extent that the projected premiums for such business last for more than 12 months from the valuation date, such a risk would likely fall within the scope of Case (1) as above, assuming that premiums do not change as a result of past experience. For this business line, referring merely to the standard formula specification would therefore imply a constant $\lambda(t) = 1$ rather than one that increases through time. Some other non-life risk sub-modules exhibit similar features.

3.16 For comparison, we note that the $\lambda(t)$ included in the HIA is akin to Case (2) (with $S = 42\%$) for the first c. 27 years and then Case (1) thereafter.

3.17 The correlation matrix used to quantify diversification effects within the standard formula SCR is time-independent. There is therefore little obvious reason just from the wording of this part of the standard formula SCR to adjust the (average) shape of $\lambda(t)$ to reflect diversification effects. However, it should be borne in mind that the relative magnitudes of the risks can alter over the lifetime of the liabilities. The theoretically correct mix to use at any given point in such an analysis will also depend on the multi-year dependency features of each individual risk involved. This in theory should impact on the most appropriate way to average the applicable $\lambda(t)$ for different risks. For firms using an internal model, a possibility would be to develop suitable $\lambda(t)$ for each risk in isolation and then to combine them in an appropriate manner after separately allowing for multi-year time dependency features for each risk. However, this level of complexity is likely to be excessive for standard formula firms.

4. Further comments based on intrinsic natures of insurer-specific risks

4.1 In Section 3, we have considered what form of dependency might be inferred for $\lambda(t)$ if we assume that the autocorrelation features for the risk in question can be derived purely from the standard formula SCR wording. However, we do not believe that the stresses included in the standard formula were formulated bearing in mind this possible use. Indeed, at the time they were formulated it was less clearly recognised that the risk margin might form a significant fraction of insurers' regulatory capital requirements, averaged across the entire

[3] A potential topic for further research might be the actual mix of risks typically present within EEA insurers, which might inform a suitable choice of an average $\lambda(t)$ applicable to the industry as a whole.

industry. In this Section we therefore consider how the intrinsic nature of the risks might alter the picture revealed in Section 3.

4.2 For operational risk, we might argue on intrinsic grounds that possibly a firm that is repeatedly hit by operational risk losses will eventually place greater emphasis on mitigating future operational risks. Conversely, initially a past loss might in part be indicative of a control weakness that might not initially be rectified, so perhaps superimposed on a possible longer term negative autocorrelation might be a shorter term positive autocorrelation, making it difficult to tell whether in aggregate a divergence from a flat $\lambda(t)$ is justified.

4.3 Many commentators have argued that longevity risk exhibits features somewhat akin to that postulated above for mass lapse risk, i.e. an increase in life expectancy one year can be followed by further increases thereafter, but it is implausible to assume that e.g. the same sort of cancer can be "cured twice". This suggests that it would be optimistic to assume that this stress justified a decline as strong as implied by Case (3), i.e. as strong as $\lambda(t) = 0.94^t$, but some decline is still likely to be justifiable. Likewise mortality risk.

4.4 As noted in Section 3, the standard formula SCR expense risk specification has the feature that the later the risk is assumed to hit the smaller is the effective impact it is assumed to have at that point and thereafter. It is therefore probably the closest to a Case (3) risk. It is of course possible for expense overruns to happen repeatedly, as most people who have any project experience are aware. However, if the expense base is set by reference to the expenses that a third party might incur when managing the transferring book (as seems most in line with the market consistent principles that underlie the risk margin calculation), the issue becomes how plausible it might be for repeated industry-wide expense stresses to occur. Arguably, given the severity of expense stress in the standard formula SCR, such industry-wide repeats seem unlikely.

4.5 For non-life liability risk as per Delegated Regulation Article 133, the standard formula SCR stress merely refers to premium income. It therefore arguably underestimates the impact that occurrence of stresses could have on potential variability in future claims cash flows. The duration of the relevant cash flows can also be quite long. Most other non-life insurance business lines where the standard formula SCR specification refers only to premium income but where intrinsically the risk might also link to claim amounts have shorter durations. The non-life liability risk sub-module therefore seems to be an outlier in the sense that the referring merely to its standard formula wording could lead to an inappropriately optimistic choice for $\lambda(t)$.

4.6 Market risk is typically seen as a type of risk that is particularly amenable to intrinsic quantitative analysis. However, nearly all market risks are treated as hedge-able and excluded from the risk margin. Market risks that aren't included are probably ones that are less capable of accurate quantitative analysis. We have not therefore tried to explore their intrinsic characteristics further in this section.

4.7 Overall, a plausible averaging of risks across the entire insurance industry seems likely to favour an average $\lambda(t)$ that attenuates as term increases, if a single average term structure is adopted for all risks and for insurer types, bearing in mind the relative duration-weighted liabilities of life versus non-life insurers. For no risk does it appear to be easy to justify on theoretical grounds an attenuation rate above 6% p.a., introducing a practical bound on what average attenuation rate might be appropriate of between 0 and 6% p.a.. For comparison, as noted above, the EIOPA HIA proposal involves an attenuation rate of 2.75% p.a. for the first c. 27 years and 0% thereafter.

4.8 Some EEA insurers use an internal model to set their SCR. As noted previously, the theoretically most appropriate choice of $\lambda(t)$ to use in the risk margin calculation depends on the magnitudes of the risks projected to contribute to the firm's overall risk profile at a given future point in time, the multi-year dependency characteristics of these risks and their diversification characteristics. If introduction of a common (non-flat) $\lambda(t)$ is considered undesirable, an alternative would be to allow such firms to include such a feature within their internal models. Advantages and disadvantages include:

(a) A better analysis of the intrinsic nature of the risks involved should be capable of being captured within such a framework
(b) Doing so would allow the firm's risk margin to reflect better the firm's own risk profile, the multi-year dependencies present within this risk profile and how the risk profile might change through time.
(c) Multi-year diversification dependencies otherwise ignored by use of a time-independent correlation matrix could be incorporated, where justified.
(d) Solvency II SCR Internal models need prior supervisory approval, mitigating the risk that firms might be incentivised to choose inappropriately optimistic shapes for $\lambda(t)$.
(e) It might be possible for regulators to adopt such an approach without altering existing primary and secondary legislation, instead just altering regulatory guidance on how internal models should be formulated, approved and implemented. This might be done by interpreting a firm's "internal model" as a package that (in effect) not only includes an algorithm specifying how the SCR should be calculated at time $t = 0$ but also how it should be projected to apply in the risk margin calculation at times $t > 0$.
(f) Conversely, such a proposal would add complexity to such firms' internal models and to their implementations.

4.9 The Solvency II framework also includes the concept of undertaking-specific parameters (USPs). USPs can be viewed as introducing limited internal model like flexibility into specific parts of the calculation of a firm's SCR, for firms that do not use an internal model to set their SCR. Use of USPs are also subject to prior supervisory approval. Places in the SCR calculation where USPs currently exist are specified in Article 218 of EU (2014). If introduction of a common (non-flat) $\lambda(t)$ is considered undesirable but there is still a desire to allow non-internal model firms to use a non-flat $\lambda(t)$ where justified, a USP could be added to those currently listed in Article 218 targeting this aspect of the firm's SCR / risk margin calculation, by modifying the Delegated Regulation in an appropriate manner.

4.10 A more thorough analysis of these aspects of the risk margin calculation might then be needed within the ORSA that insurers are required to carry out under Solvency II. There is no bar on firms valuing assets or liabilities differently in their ORSA to how they are valued in their Pillar 1 regulatory capital computations. However, Guideline 9 of EIOPA (2014) indicates that firms then need to explain "*how the use of such different recognition and valuation bases ensures better consideration of the specific risk profile, approved risk tolerance limits and business strategy of the undertaking, while complying with the requirement for a sound and prudent management of the business*". The same guideline also then requires such firms to estimate quantitatively the impact that such differences would have on the firm's overall solvency needs assessment[4].

4.11 Firms are required within their ORSA to assess whether their risk profile deviates from the assumptions underlying the SCR calculation and whether these deviations are significant. This is known as "standard formula appropriateness" for firms that do not use an internal model to calculate their SCR. Such an assessment could be expanded to cover how the resulting SCRs are then brought into the risk margin calculation.

References

AAE (2019). A Review of the Design of the Solvency II Risk Margin. *AAE Commentary paper*

EIOPA (2014). Guidelines on own risk and solvency assessment. *EIOPA-BoS-14/259*

EIOPA (2020). Technical specification of the information request on the 2020 review of Solvency II - Holistic impact assessment. *EIOPA-BoS-20/107*

European Union (2014) Commission Delegated Regulation (EU) 2015/35 of 10 October 2014 supplementing Directive 2009/138/EC of the European Parliament and of the Council on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II). *Eur-Lex (consolidated version as at October 2020)*

[4] Another area of potential further research not explored in this paper is the extent to which firms already in effect use different valuation bases in this regard within their ORSAs.