

Down but not Out: A Cost of Capital Approach to Fair Value Risk Margins

B. John Manistre¹ FSA, FCIA, MAAA, CERA

Condensed Version June 27, 2014

This is a condensed version of a much longer technical paper with the same title to be presented as part of the 2014 ERM Symposium.

Abstract

The Market Cost of Capital approach is emerging as a standard for estimating risk margins for non-hedgeable risk on an insurer's fair value balance sheet. This paper develops a conceptually rigorous formulation of the cost of capital method for estimating margins for mortality, lapse, expense and other forms of underwriting risk. For any risk situation we develop a three step modeling approach which starts with i) a best estimate model and then adds ii) a static margin for contagion risk (the risk that current experience differs from the best estimate) and iii) a dynamic margin for parameter risk (the risk that the best estimate is wrong and must be revised).

We show that the solution to the parameter risk problem is fundamentally a regime switching model which can be solved by Monte Carlo simulation. The paper then goes on to develop four more pragmatic methods which can be thought of as short cut approximations to the first principles model. One of these short cuts is the Prospective method currently used in Europe. None of these methods require stochastic or stochastic projections to get useful results.

Introduction

There is a well-known quote, due to George E.P. Box, which goes, "All models are wrong but some are useful."² All of the methods outlined in this article take this concept to heart in the sense that the model structures themselves recognize that the models are wrong and will require adjustment as new information becomes available. The models are therefore intended to be applied in the context of a principles based, fair valuation system where continuous model improvement is an integral part of the process. One possible application would be to an internal economic capital model or an Own Risk and Self-Assessment (ORSA) process. The author believes the methods described here would also meet IFRS standards.

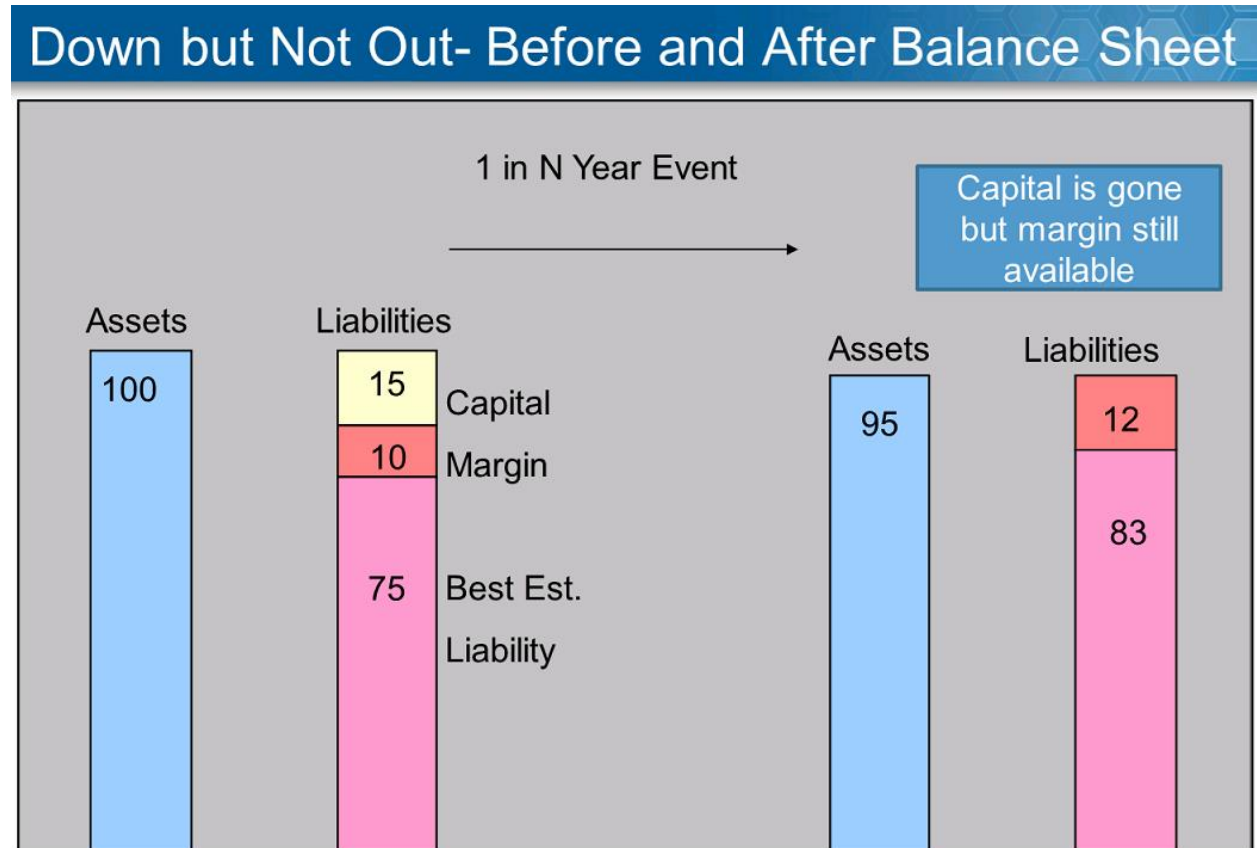
The cost of capital concept itself has been part of actuarial culture for many decades and this paper assumes the reader already has some familiarity with the idea. At a high level, the idea is that if a contract requires the enterprise to hold economic capital in the amount EC then we need to build an annual expense πEC into the value of the contract to price in the risk. The quantity π here is the cost of capital rate and it can vary from application to application. For non-hedgeable life insurance risk a typical cost of capital rate is $\pi = .06$.

¹ The author is a Research Actuary at GGY AXIS based in Baltimore Md.

² George E.P. Box (FRS) in 1987.

There are three themes or common denominators that run through all of the methods discussed here. These are:

1. **Down but not Out:** The idea is that if a 1 in N year event wipes out the economic capital of a risk enterprise there should still be enough risk margin on the balance sheet that the company can either attract a new investor to replace the lost capital or, equivalently, pay a similar healthy enterprise to take on its obligations. The chart below illustrates the idea graphically.



On the left side of the chart we see the risk enterprise's economic balance sheet at the beginning of the year. The right side of the chart shows the fair value balance sheet after a bad year. As a result of both poor experience in the current year, and adverse assumption revisions, all of the economic capital is gone. The risk enterprise is down. However, the economic balance sheet is still strong enough that it can either attract a new investor to replace the lost capital or pay another enterprise to take on its obligations i.e. the risk enterprise is not out because appropriate risk margins are still available.

This is clearly a desirable theoretical property for a model to have. In order to actually work in practice the revised balance sheet on the right must have enough credibility with the outside world that a knowledgeable investor would actually put up the funds necessary to continue. One way to get the needed credibility is for the actuarial

profession to develop standards of practice that are rigorous enough for the shocked balance sheet to be credible.

2. **Linearity:** All of the methods considered here can be formulated as systems of linear stochastic equations. This has two very general consequences.

- a. As is well known, a linear problem usually has a dual version. If you can solve the primal problem you can also solve the dual to get the same answer. In this case the primal version of the problem looks like an “actuarial” calculation where we project capital requirements into the future and then compute margins as the present value of the cost of capital.

As formulated here, the dual version of the problem looks more like a “financial engineering” calculation. The process above is reversed by starting with a concept of risk neutral or risk loaded mortality, lapse etc. and then determining the corresponding implied economic capital by seeing how the margins unwind over time.

Put another way, if the present value of margins M and the economic capital EC are related by an equation of the form

$$\frac{dM}{dt} = (r + \mu)M - \pi EC,$$

then the primal version of the method starts by projecting EC and then uses the above relation to calculate margins by discounting. The dual approach calculates M first and then uses a version of the relation above to estimate an implied economic capital EC .

- b. A second useful consequence of using linear models is that they allow us to avoid the “stochastic on stochastic” issue that bedevils many other approaches to the margin issue. Linear models can be calculated scenario by economic scenario. Any errors we make by ignoring the “stochastic on stochastic” nature of the problem average out to zero when we sum over a large set of risk neutral scenarios³.

With this result we can develop the cost of capital ideas in a simple deterministic economic model, and be confident that the results developed will continue to apply when we go to a fully stochastic economic model.

Looking at the dual gives us both new theoretical insight and an alternative way to compute any given model. In particular, the dual approach adds transparency in the sense that it tells us what the implied “risk neutral” assumptions for mortality, lapse etc. are.

³ This is a standard result in stochastic calculus which is outlined in the main technical paper.

For any particular application, the primal and dual approaches are equivalent but can differ in practice for a variety of reasons. One of the paper's general conclusions is that solving the primal problem works well for simple applications but the dual approach can be preferable as the complexity of the application increases. The main problem with the dual approach is the effort required to understand why the theory works. The actual implementation is not that difficult.

We take the view that both the primal and dual versions of a model should make theoretical sense and this leads to a critique of some approaches. For example, the primal version of the prospective model used in Europe usually looks simple and reasonable but the dual version may not. This is illustrated in the main paper by looking at the example of a lapse supported insurance product. It is possible for the dual problem to exhibit negative risk loaded lapse rates. We offer a modification to the method, as well as several other approaches, that can resolve this issue.

3. **The basic risk modeling process:** This article assumes a three step process for putting a value on non-hedgeable risk. In a bit more detail, the steps are
 - a. Develop a best estimate model that is appropriate to the circumstances of the application. Detailed discussion of this step is outside the scope of this paper although we do provide a number of examples from life insurance. The key assumption we make is that our best estimate models are not perfect and are subject to revision.
 - b. Hold capital and risk margins for a contagion event i.e. the risk that current experience may differ substantially from our best estimate.

Imagine, for the sake of clarity, that our best estimate model is a traditional actuarial mortality table. Even if our table is right on average, we could still have bad experience in any given year. The classic example of a contagion event would be a repeat of the 1918 flu epidemic – hence the name contagion risk.

More recent examples of contagion risk events would be the North American commercial mortgage meltdown in the early 1990's⁴ and the well-known problems with the US residential mortgage market that led to the financial crisis of 2008.

A risk enterprise should have sufficient capital and margins that it can withstand a plausible contagion event and still be able to continue as a going concern without regulatory intervention. We show that traditional, static, risk loadings in our parameters can usually deal with this issue.

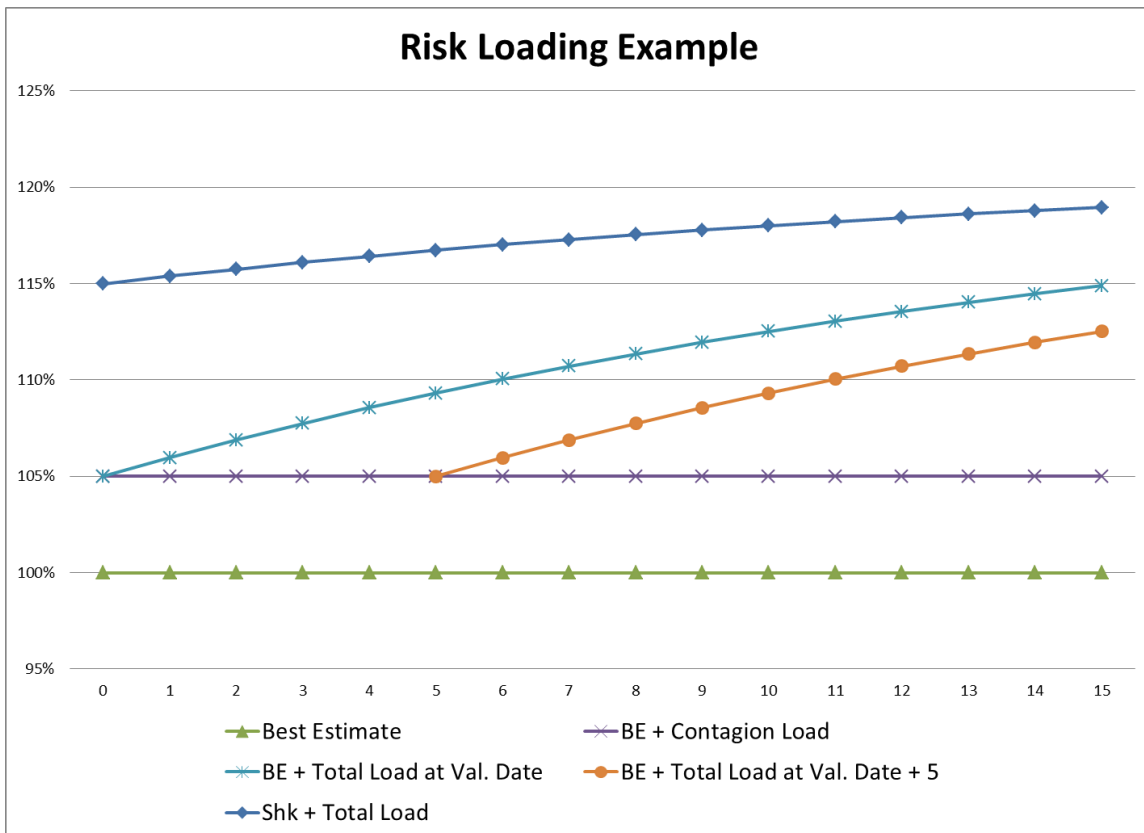
⁴ This was caused by the overbuilding of office space during the 1980's in many North American cities. When the oversupply became apparent, office rents plummeted. This dragged down property values and triggered defaults on many of the mortgages used to finance the office towers.

- c. Hold capital and margins for parameter risk: New information might arrive in the course of a year that causes the risk enterprise to revise one or more models. To the extent these model revisions cause the fair value of liabilities to increase, we need economic capital to absorb the loss. Again we need a margin model that allows the risk enterprise to withstand the loss and carry on without regulatory intervention. To deal with this issue, we introduce the concept of a dynamic loading which arises naturally out of the dual approach.

Static and dynamic loadings differ in the way margin gets released into income over time. If best estimate assumptions are realized, then any static margin emerges as an experience gain in the current reporting period. The risk loading is engineered so that the resulting gain is equal to the cost of holding capital for contagion risk. This is what most actuaries would expect.

By contrast, a dynamic margin is a time dependent loading, which is equal to zero at the valuation date, and then grades to an ultimate value discussed later. There is very little experience gain in the current reporting period. The risk margin gets released into income by pushing out the grading process as time evolves i.e., when we come to do a new valuation, we establish a new dynamic load which restarts from zero at the new valuation date. If we get the math right, this process releases the correct amount of margin to pay for the cost of holding economic capital for parameter risk, while still leaving sufficient margin on the balance sheet for the future.

Chart 1 below shows a simple example of the risk loading ideas introduced above.



In this example we have a model parameter whose best estimate value is $\theta_0 = 100\%$ and a static contagion loading of 5% has been added. At the valuation date ($t = 0$), we have added a dynamic load that takes the parameter up to the value of 115% over a 15 year period. This is the parameter path used to compute a fair value. A shocked fair value is calculated assuming a shocked path that starts at 115% (base + 10%) and then grades to about 119%. Economic capital, for parameter risk, is the difference between the shocked and base fair values.

When we come to do a new valuation 5 years later, the contagion loading has not changed but the dynamic loading for parameter risk has been recalculated to start at zero again. The risk margin released into income, if the assumptions do not change, is engineered to provide a target return on the risk capital.

A high level summary of the paper's theory is that the cost of capital method for calculating risk margins is, for most practical purposes, equivalent to using an appropriate combination of static and dynamic risk loadings.

The process described above is much easier to implement than it looks. The paper discusses a number of reasonable simplifying assumptions that allow the risk loaded parameters to be calculated fairly easily. **None of the methods discussed require any computationally expensive “stochastic on stochastic” or “projection within projection” algorithms.**

A Simple Term Life Example – the Best Estimate Model

Assume a contract that pays a death benefit F if the insured dies before an expiry date T . The insured pays a continuous premium g while alive. The insurer incurs maintenance expenses at the rate e per unit time while the contract is in force. We'll also assume a deterministic force of interest r .

The insurer's best estimate of the insured's force of mortality at time $s < T$ is a known function $\mu_0(s)$. Ignoring the policyholder's option to lapse (i.e. stop paying premiums), we can compute the insurer's best estimate of the fair value of the liability by solving an equation of the form

$$\frac{dV_0}{dt} + \mu_0(t)[F - V_0] = rV_0 + g - e.$$

An intuitive way to understand this equation is to say that the left side is the expected rate of change of the liability and the right side is the expected rate of change of the insurer's assets backing the liability.

Since the coverage expires at time T the appropriate boundary condition to assume is $V_0(T) = 0$.

The solution to this valuation problem is the well-known actuarial discounting formula

$$V_0(t) = \int_t^T e^{-\int_t^s (r + \mu_0(v)) dv} [\mu_0(s)F + e - g] ds.$$

The best estimate value V_0 is the actuarial present value of death benefits and expenses offset by the present value of gross premiums.

A Simple Term Life Example - Contagion Risk and Static Margins

Now assume that the insurer holds capital to protect its solvency in the event of a contagion loss such as a repeat of the 1918 flu epidemic. The insurer has determined that such an event would result in $\Delta Q(t)$ additional deaths per life exposed. If V is the value of the contract, which includes margin for this risk, then the amount of capital the insurer must hold is $\Delta Q[F - V]$ since this is the economic loss that would occur if additional ΔQ deaths were to occur at time t .

Letting π denote the insurer's cost of capital rate (e.g. 6.00%), the new valuation equation should include the cost of contagion risk capital as an additional expense i.e.

$$\frac{dV}{dt} + \mu_0(t)[F - V] = rV + g - e - \pi\Delta Q[F - V].$$

But this can easily be rewritten as

$$\frac{dV}{dt} + [\mu_0(t) + \pi\Delta Q][F - V] = rV + g - e.$$

This shows that including margin for the cost of holding contagion risk capital is equivalent to simply adding a load $\pi\Delta Q$ to the best estimate force of mortality μ_0 . We will refer to this process as one of adding a *static margin* or a *contagion loading*.

The result illustrated is simple, easy to implement, and makes sense as long as $\Delta Q[F - V] > 0$. For this example it is not hard to show what $F > V$ as long as the premiums and expenses are reasonable relative to the death benefit. Under this assumption, the contagion shock ΔQ should be a positive number and set at a level consistent with the insurer's overall capital target (e.g. one year *VaR* at the 99.5% level).

It is worth discussing why this model satisfies the "down but not out" principle. Assuming mortality contagion is our only risk issue, an investor is asked to put up economic capital in the amount $EC(t) = \Delta Q(t)[F - V(t)]$. The insurer then charges the customer a premium sufficient to cover the cost of expenses and death claims at the contagion loaded level.

To the extent best estimate assumptions are realized, the insurer will recognize economic profits equal to the margin release plus interest on the economic capital. The total economic return to the investor is then $(r + \pi)\Delta Q[F - V]$.

At the end of the period, the insurer returns the original capital, and profits, to the shareholder and then asks for a new capital infusion in the amount $\Delta Q(t + 1)[F - V(t + 1)]$ to finance the risk taking in the next period. We assume the investor is willing to do this because the product has been engineered to provide the same expected return on this new, higher or lower, capital amount in the following time period.

If experience is better than expected, the return in the current period will be higher than $(r + \pi)$ and, if worse, the return will be lower and possibly negative. We can imagine the following conversation between an investor and company management.

Management: *Hello Mr. Investor, welcome to the insurance business. I have some good news and some bad news.*

Investor: *I'm not sure I like the sound of that.*

Management: *The bad news is that we have had some adverse mortality experience this year and most of our available risk capital is gone. The good news is that there are still sufficient loadings in the future mortality rates that you can expect a reasonable return on your investment, if you replace the lost capital now.*

Investor: *How can I be confident this won't happen again?*

Management: *You can't. This is a risk business. The company's actuaries have followed all appropriate professional standards of practice in choosing methods, assumptions and performing the actual calculations. It is possible that we could have another bad year before the business has run off. If you are uncomfortable with that, you are investing in the wrong business.*

To the extent management's models and assumptions have credibility with the appropriate investor public, the company can withstand a loss up to the contagion capital level and still be strong enough to recapitalize and carry-on. There would be no need for regulatory intervention. This is what "down but not out" means in this paper.

Finally the investor asks, "What happens if you discover one or more of your assumptions is wrong and must be revised?" In order to answer this question we have to extend the model to cover parameter risk.

A Simple Term Life Example -Parameter Risk and Dynamic Margins

The previous section argued that the first two steps of our risk modeling process resulted in a contagion loaded force of mortality equal to $\mu(t) = \mu_0(t) + \pi\Delta Q(t)$. This would be the force of mortality used in a traditional actuarial valuation in order to provide a margin for contagion risk.

We now consider the risk that either the best estimate force of mortality assumption $\mu_0(t)$ or the contagion shock ΔQ could be wrong. New information might arrive which leads the insurer to set a new contagion loaded assumption $\mu + \Delta\mu$. Letting \hat{V} denote the relevant shocked fair value, we need to hold risk capital in the amount $\hat{V} - V$.

The size of the shock $\Delta\mu$ should reflect a plausible assumption change over the course of one year at something like the 99.5% VaR level. The size of the shock would then reflect the inherent riskiness or "liquidity" of the business. Shocks for blocks of traditional business that are well understood would presumably be smaller than shocks for newer or less liquid types of

business. Ideally, there should be some industry consensus around the principles used to choose the shock.

The fundamental valuation equation, which incorporates both contagion risk and parameter risk, now becomes

$$\frac{dV}{dt} + \mu_0(t)[F - V] = rV + g - e - \pi\Delta Q[F - V] - \pi[\hat{V} - V],$$

or

$$\frac{dV}{dt} + \mu(t)[F - V] = rV + g - e - \pi[\hat{V} - V].$$

This seems simple enough until we consider how we should calculate \hat{V} . This is a reserve, based on a mortality assumption $\mu + \Delta\mu$ which could again turn out to be wrong. The value \hat{V} also needs to include a margin for parameter risk, the risk that the mortality assumption might need to change again. Letting $\hat{V}^{(2)}$ denote a double shocked fair value, we would need to hold parameter risk capital in the amount $\hat{V}^{(2)} - \hat{V}$ in a shocked world.

The obvious extension of the equation above is then to write

$$\frac{d\hat{V}}{dt} + [\mu(t) + \Delta\mu(t)][F - \hat{V}] = r\hat{V} + g - e - \pi[\hat{V}^{(2)} - \hat{V}].$$

This equation makes the reasonable assumption that the shocked contagion loaded force of mortality used to calculate \hat{V} is $\mu + \Delta\mu$. The problem is that “down but not out” means we have had to introduce a second shocked reserve value $\hat{V}^{(2)}$ which, presumably, depends on a second level of parameter shock $\Delta\mu^{(2)}$ and a third contagion-loaded force of mortality $\mu + \Delta\mu + \Delta\mu^{(2)}$. We seem to be trapped in an impractical infinite regress. The reserve V depends on \hat{V} which depends on $\hat{V}^{(2)}$, and so on. This is known as the circularity problem.

A large part of the paper is devoted to solving the circularity problem. The paper does enough theoretical analysis to come up with a true first principles approach to calculating the model and then develops four very practical short cut methods. There is a wide range of practical problems where all four short cuts produce very similar results.

It turns out that it is useful to think of the short cut methods as pragmatic approximations to a model where we have a geometric hierarchy of assumption levels $\mu, \mu + \Delta\mu, \mu + \Delta\mu + \Delta\mu^{(2)}, \dots$ where $\Delta\mu^{(n)} = \alpha^{n-1} \Delta\mu$ for some constant $0 \leq \alpha \leq 1$. The n' th level in the shock hierarchy would then be given by

$$\mu^{(n)} = \mu + \Delta\mu + \alpha\Delta\mu + \dots + \alpha^{n-1}\Delta\mu = \begin{cases} \mu + \frac{1-\alpha^n}{1-\alpha}\Delta\mu, & \alpha < 1 \\ \mu + n\Delta\mu, & \alpha = 1 \end{cases}$$

The first principles solution to the circularity problem is to calculate the fair value V as an expected present value, where the force of mortality is allowed to be a random quantity μ , which jumps from one mortality level to the next with a transition intensity equal to the cost of capital rate π ⁵.

In symbols, we can write the value as

$$V(t) = E_{C(t)} \int_t^T e^{-\int_t^s (r+\mu) dv} [\mu F + e - g] ds .$$

We use the symbol $C(t)$ to denote the regime switching probability measure governing the process $\mu \xrightarrow{\pi} \mu + \Delta\mu \xrightarrow{\pi} \mu + \Delta\mu + \Delta\mu^{(2)} \xrightarrow{\pi} \dots$. We will call this the C measure.

To calculate a shocked value we do the same kind of calculation but with a regime hierarchy that starts with the first shocked level $\mu + \Delta\mu \xrightarrow{\pi} \mu + \Delta\mu + \Delta\mu^{(2)} \xrightarrow{\pi} \dots$. Using the symbol \hat{C} to denote this shocked regime switching measure, we can write a formula for the shocked value as

$$\hat{V}(t) = E_{\hat{C}(t)} \int_t^T e^{-\int_t^s (r+\mu) dv} [\mu F + e - g] ds .$$

Economic Capital for parameter risk is then calculated as $EC = \hat{V} - V$.

In theory, there is no real obstacle to implementing the first principles model. One merely has to specify the shock hierarchy and then perform a large number of Monte Carlo simulations. A numerical example is given in the main paper.

In practice, there are two serious flaws with the approach outlined above. An obvious issue is the computational cost of running a large number of random mortality scenarios. The second issue is that most practitioners would think that specifying an infinite hierarchy of assumptions is over engineering. The short cut methods to come address both issues.

Short Cut # 1 the Simple Mean

For a valuation starting at time t , the probability of reaching the n 'th regime at time $t+s$ is given by the Poisson probability $\exp[-\pi s](\pi s)^n / n!$. The expected force of mortality for a geometric hierarchy can then be calculated in closed form as

$$E_{C(t)}[\mu(s)] = \sum_{n=0}^{\infty} e^{-\pi(s-t)} \frac{(\pi(s-t))^n}{n!} \begin{cases} [\mu(s) + \frac{1-\alpha^n}{1-\alpha} \Delta\mu(s)] & \alpha < 1 \\ \mu(s) + n\Delta\mu(s) & \alpha = 1 \end{cases}$$

$$= \begin{cases} \mu(s) + \frac{1-e^{-\pi(1-\alpha)(s-t)}}{1-\alpha} \Delta\mu(s), & \alpha < 1 \\ \mu(s) + \pi(s-t) \Delta\mu(s) & \alpha = 1 \end{cases}$$

⁵ See the main paper for more detail on why this is true.

We now make the simplifying assumption⁶ that

$${}_s\bar{p}_t = E_{C(t)} \exp \left[- \int_t^{t+s} \boldsymbol{\mu} dv \right] \approx \exp \left[- \int_t^{t+s} E_{C(t)}[\boldsymbol{\mu}] dv \right].$$

To the extent this simplifying assumption is reasonable (it usually is) we see that *the cost of capital method for parameter risk is, approximately, equivalent to using an assumption equal to the base $\mu(t)$ on the valuation date plus a load that grades from 0 towards an ultimate value $\Delta\mu/(1-\alpha)$ with a speed of mean reversion rate given by $\pi(1-\alpha)$.*

The parameter risk margin is released by continuously pushing the grading process out into the future as time evolves. This is clearly a dynamic loading structure.

To help compare this approximation with other short cuts we note that we can write $\bar{\mu}(t, s) = \mu(s) + \bar{\beta}(t, s)\Delta\mu(s)$ where

$$\bar{\beta}(t, s) = \begin{cases} \frac{1 - e^{-\pi(1-\alpha)(s-t)}}{1-\alpha}, & \alpha < 1 \\ \pi(s-t) & \alpha = 1 \end{cases}$$

satisfies the dynamical rule

$$d\bar{\beta} = \pi[1 - (1-\alpha)\bar{\beta}], \quad \bar{\beta}(t, t) = 0. \quad (1)$$

We can get a sense of what capital means for this model by calculating the expected mortality under the shocked regime switching measure. The same algebra as before yields the following result

$$E_{\hat{C}(t)}[\boldsymbol{\mu}(s)] = \mu(s) + \Delta\mu(s) + \alpha\bar{\beta}(t, s)\Delta\mu(s).$$

This result suggests that *the shocked valuation scenario looks like a world where the base assumption μ has been replaced by $\mu + \Delta\mu$ and the risk loadings have been multiplied by the factor α .* This result makes intuitive sense for the geometric shock hierarchy⁷.

Short Cut # 2 the Explicit Margin Method

The Explicit Margin Method is a fine tuning of the Simple Mean Method so that the qualitative statements made in italics above are exactly true. The idea is to assume a risk loaded force of mortality of the form $\mu + \beta^e \Delta\mu$ where the margin variable $\beta^e = 0$ at the valuation date t and then evolves forward in time according to a dynamical rule of the form

$$d\beta^e = B(s, \beta^e) ds, \quad s > t.$$

⁶ This is a conservative approximation to the exact geometric hierarchy model if $\Delta\mu > 0$.

⁷ We do not make a prima fascia case for using a geometric shock hierarchy. It is a simple place to start and, as we will see, all of the short cut methods can be thought of as pragmatic approximations to this model.

The fair value V is calculated using $\mu + \beta^e \Delta\mu$ and a shocked fair value \hat{V} is calculated using the shocked assumption $\mu + \Delta\mu + \alpha\beta^e \Delta\mu$. We now ask how we should choose the function $B(s, \beta^e)$ to make the margin release rate equal to the cost of capital $\pi[\hat{V} - V]$. The answer

$$d\beta^e(s) = (\pi - \beta^e \Delta\mu(s))(1 - (1 - \alpha)\beta^e(s))ds, \beta^e(t) = 0, \quad (2)$$

is derived in the main paper.

This turns out to be an exact theoretical solution to the first principles model, with geometric shock hierarchy, when $\alpha = 0$ or 1 . It is a good approximation when $0 < \alpha < 1$ and has a convenient discrete time implementation detailed in the appendix to this document.

On comparing the evolution equations (1),(2) for $\bar{\beta}$ and β^e we see that the explicit margin variable evolves like $\bar{\beta}$ with an effective cost of capital rate $\pi - \beta^e \Delta\mu$. This means that if $\Delta\mu > 0$ then $\beta^e < \bar{\beta}$ and the relationship reverses if $\Delta\mu < 0$. For a typical mortality application we have $\Delta\mu \approx \pm 1/10,000$ so the two models are often close if π is much bigger than $\Delta\mu$ e.g. $\pi = 6/100$.

One reasonable property of these models is that the longer the contract persists, the more loading is built into the assumed mortality. In general, this makes sense because parameter risk is more of an issue with a longer contract than it is with a shorter one. However, it can lead to the models being misused.

As an example of misuse, consider the situation of an annuity risk where it is appropriate to assume the mortality shock is negative i.e. $\Delta\mu < 0$. If we set α close to 1 then the ultimate value $\mu + \Delta\mu/(1 - \alpha)$ could be negative. This issue is typically not material when dealing with mortality risk but it can be important when developing lapse rate loadings for lapse supported products.

The main paper refers to Short Cuts 1 & 2 as the “Financial Engineering” models because the risk loaded mortality has the form $\mu + \beta\Delta\mu$ where we can think of β as a variable that is zero in the real world (P Measure) but has a dynamic of the form $d\beta = B(s, \beta)ds$ in the risk neutral or valuation world (C measure).

Once we think in those terms, we can think of the fair value as a function $V = V(s, \beta)$ in the valuation world. This raises the question, what is the meaning of the “greek” $\partial V/\partial\beta$? The main paper shows that if $B(s, 0) = \pi$ then, on the valuation date, $\partial V/\partial\beta$ is the implied economic capital for the model.

Short Cuts 1 & 2 are therefore only two particular examples of models, of that form, that could be useful. We are free to specify the dynamics $d\beta = B(s, \beta)ds$ any we like as long as the resulting implied economic capital $\partial V/\partial\beta$ makes sense.

Both of the simple mean and explicit methods have convenient discrete time implementations if we assume the shock $\Delta\mu$ is piecewise constant by year of age. The details are in the appendix.

Two Actuarial Short Cuts (3 & 4)

The previous models started with a version of the first principles model and then made a simplifying assumption which allowed us to analyze the continuous time version of the model.

For Short Cuts 3 & 4 we make a more “actuarial” simplifying assumption which, somewhat surprisingly, ends up taking us to almost the same place as before.

Short Cut #3 The Implicit Margin Method

The implicit method is based on the simplifying assumption that there is constant $\alpha \geq 0$ such that the capital required in a shocked world can be approximated by $\hat{V}^{(2)} - \hat{V} \approx \alpha(\hat{V} - V)$. In continuous time, this breaks the circular system of valuation equations down to a two dimensional system

$$\begin{aligned}\frac{dV}{dt} + \mu(t)[F - V] &= rV + g - e - \pi[\hat{V} - V], \\ \frac{d\hat{V}}{dt} + (\mu(t) + \Delta\mu(t))[F - \hat{V}] &= r\hat{V} + g - e - \pi\alpha[\hat{V} - V].\end{aligned}$$

Here is a simple discrete time version for the term insurance example. This particular discretization bases the margin on the cost of capital held at the beginning of the time period. A more precise approximation to the continuous time model would base the margin on some kind of average capital held over the period.

Let ${}_tV$ and ${}_t\hat{V}$ denote the discrete time values at time t and let i be the one period effective interest rate. Letting q, \hat{q} be the base and shocked mortality rates we can write

$$\begin{aligned}({}_tV + g - e)(1 + i) &= qF + (1 - q)_{t+1}V + \pi({}_t\hat{V} - {}_tV), & (a) \\ ({}_t\hat{V} + g - e)(1 + i) &= \hat{q}F + (1 - \hat{q})_{t+1}\hat{V} + \alpha\pi({}_t\hat{V} - {}_tV). & (b)\end{aligned}$$

Subtracting one equation from the other we find

$$({}_t\hat{V} - {}_tV)(1 + i) = [\hat{q}F + (1 - \hat{q})_{t+1}\hat{V}] - [qF + (1 - q)_{t+1}V] - \pi(1 - \alpha)({}_t\hat{V} - {}_tV).$$

This can be solved for the capital requirement

$$({}_t\hat{V} - {}_tV) = \frac{[\hat{q}F + (1 - \hat{q})_{t+1}\hat{V}] - [qF + (1 - q)_{t+1}V]}{1 + i + \pi(1 - \alpha)} \quad (c)$$

Equations (a),(b) and (c) above form a simple recursive system that allows us to determine values at time t given that we know the relevant values at time $(t + 1)$. The table below shows a spreadsheet implementation of the above logic when $g = e = 0$, $i = 4.00\%$, $\alpha = 100\%$ and $\pi = 6.00\%$. The shocked mortality is 10% higher than the base mortality.

| Table 1: Implicit Method Example | | | | | | | | |
|----------------------------------|---------|-------------|---------------------|--------|--------|---------|---------|-----------|
| Level Shock= 10.0% | | | Face= 10,000 | | | i=4.00% | | |
| alpha= 100.0% | | | Values per survivor | | | Margin | Capital | |
| Year | Base qx | Shocked qx^ | V0 | V | V^ | V-V0 | V^-V | RoC Check |
| 0 | | | 121.53 | 125.63 | 137.70 | 4.10 | 12.07 | |
| 1 | 0.00102 | 0.00112 | 116.36 | 119.91 | 131.46 | 3.55 | 11.56 | 6.00% |
| 2 | 0.00111 | 0.00122 | 110.07 | 113.07 | 124.01 | 3.00 | 10.94 | 6.00% |
| 3 | 0.00121 | 0.00133 | 102.52 | 104.99 | 115.18 | 2.46 | 10.20 | 6.00% |
| 4 | 0.00132 | 0.00145 | 93.55 | 95.50 | 104.81 | 1.95 | 9.31 | 6.00% |
| 5 | 0.00144 | 0.00159 | 83.00 | 84.47 | 92.74 | 1.48 | 8.27 | 6.00% |
| 6 | 0.00157 | 0.00173 | 70.70 | 71.74 | 78.79 | 1.04 | 7.05 | 6.00% |
| 7 | 0.00172 | 0.00189 | 56.47 | 57.13 | 62.76 | 0.66 | 5.63 | 6.00% |
| 8 | 0.00187 | 0.00205 | 40.13 | 40.48 | 44.48 | 0.35 | 4.01 | 6.00% |
| 9 | 0.00204 | 0.00224 | 21.37 | 21.50 | 23.63 | 0.12 | 2.14 | 6.00% |
| 10 | 0.00222 | 0.00245 | - | - | - | - | - | 6.00% |

The product being illustrated is a 10 year term insurance. The column labeled V0 was calculated using only the base mortality qx. The last column (Return on Capital) was calculated as the margin released, if best estimate assumptions are realized, divided by the required capital at the beginning of the contract year.

$$ROC_{t+1} = \frac{Margin_t(1+i) - (1-q_{t+1})Margin_{t+1}}{Capital_t}$$

If the math is working properly, we should get back the cost of capital rate that we used as an input. The expected return to the shareholder is then 6% coming from margin release plus 4% interest on capital for a combined total of 10%.

The main practical shortcoming of this version of the implicit method is that it cannot be directly applied when dealing with more complex products such as universal life or joint life cases. In these situations the business being valued must first be broken down into components to which the method can be directly applied. This can be difficult.

The implicit method gets its name from the fact that the fair value margins are implicit in the discounting process used to solve the linear system.

Short Cut # 4 The Prospective Method (Solvency II)

The prospective method starts by calculating best estimate values V_0, V_1 using assumptions $\mu, \mu + \Delta\mu$ respectively. The margined values are then assumed to be $V = V_0 + M$ and $\hat{V} = V_1 + \hat{M}$.

The key simplifying assumption is that there is constant $\alpha \geq 0$ such that

$$\hat{M} = \alpha M.$$

This is clearly similar in spirit to the assumption $\hat{V}^{(2)} - \hat{V} \approx \alpha(\hat{V} - V)$ underlying the implicit method. It is another way to get around the circularity issue.

Having made this assumption, the economic capital is given by $\hat{V} - V = V_1 - V_0 - (1 - \alpha)M$.

The present value of margins is then calculated by discounting the cost of capital using interest and base mortality i.e.

$$\frac{dM}{dt} = (r + \mu)M - \pi[V_1 - V_0 - (1 - \alpha)M].$$

This implies that

$$M(t) = \int_t^\infty e^{-\int_t^s (r + \mu + \pi(1 - \alpha)) dv} \pi[V_1(s) - V_0(s)] ds.$$

The prospective method was adopted by European regulators in 2010 for the Solvency II Quantitative Impact Study #5. Their specification set $\alpha = 1$ for all products and they also allowed an illiquidity premium ϑ to be added to the risk free rate when calculating V_0, V_1 , but not M .

Table 2 below shows numerical results for the same 10 year term product that was used to illustrate the implicit method. We have set $\vartheta = 0$ for this example and used a simple discretization scheme that bases the margin on the beginning of period capital amount.

The actual discrete time equations used for the example are

$$\begin{aligned} {}_tV_0(1 + i + \vartheta) &= qF + (1 - q) {}_{t+1}V_0, \\ {}_tV_1(1 + i + \vartheta) &= \hat{q}F + (1 - \hat{q}) {}_{t+1}V_1, \\ {}_tM(1 + i) &= (1 - q) {}_{t+1}M + \pi[{}_tV_1 - {}_tV_0 - (1 - \alpha) {}_tM]. \end{aligned}$$

| Year | Values per survivor | | | | |
|------|---------------------|--------|--------|---------|-----------|
| | V0 | V1 | Margin | Capital | RoC Check |
| 0 | 121.53 | 133.60 | 4.10 | 12.07 | |
| 1 | 116.36 | 127.92 | 3.55 | 11.56 | 6.00% |
| 2 | 110.07 | 121.01 | 3.00 | 10.94 | 6.00% |
| 3 | 102.52 | 112.72 | 2.46 | 10.20 | 6.00% |
| 4 | 93.55 | 102.86 | 1.95 | 9.31 | 6.00% |
| 5 | 83.00 | 91.27 | 1.48 | 8.27 | 6.00% |
| 6 | 70.70 | 77.75 | 1.04 | 7.05 | 6.00% |
| 7 | 56.47 | 62.10 | 0.66 | 5.63 | 6.00% |
| 8 | 40.13 | 44.13 | 0.35 | 4.01 | 6.00% |
| 9 | 21.37 | 23.51 | 0.12 | 2.14 | 6.00% |
| 10 | - | - | - | - | 6.00% |

The actual values for margins and capital are not identical to the implicit method but they are close enough that they are equal at the displayed level of precision. If we went out further in time we would eventually see a difference.

The main paper shows that the continuous time version of both actuarial short cuts can be formulated as financial engineering methods, if we choose to do so.

Comparing the Short Cuts – Duality

One way to compare the four practical methods is to apply them all to the same simple problem. The problem we pick is that of fair valuing a simple pure endowment in a 0% interest rate environment. For the two actuarial methods we have applied the mechanics illustrated above to value a sequence of pure endowments that vary by maturity. For the two financial engineering methods we first calculated the risk loaded mortality rates (in the appendix) and then used those rates to compute the pure endowments.

The table below shows the results of this test for a range of maturities varying from 0 to 10 years.

| Table 3a | | | | | | |
|-----------------|-----------|-----------------------------|--------|--------------|----------|-------------------|
| Input Mortality | | Pure Endowment Calculations | | | | |
| | | $\alpha = 100\%$ | | $i = 0.00\%$ | | |
| Base | Shocked | | Simple | Explicit | Implicit | Prospective |
| 1000 qx | 1000 qx^A | $\pi =$ | Mean | | | $\theta = 0.00\%$ |
| | 10% | | 5.83% | 5.83% | 6.00% | 6.00% |
| | | 0 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 1.0150 | 1.1165 | 1 | 0.9990 | 0.9990 | 0.9990 | 0.9990 |
| 1.1063 | 1.2170 | 2 | 0.9979 | 0.9979 | 0.9979 | 0.9979 |
| 1.2078 | 1.3286 | 3 | 0.9966 | 0.9966 | 0.9966 | 0.9966 |
| 1.3195 | 1.4514 | 4 | 0.9953 | 0.9953 | 0.9953 | 0.9953 |
| 1.4413 | 1.5854 | 5 | 0.9938 | 0.9938 | 0.9938 | 0.9938 |
| 1.5732 | 1.7306 | 6 | 0.9922 | 0.9922 | 0.9922 | 0.9922 |
| 1.7153 | 1.8869 | 7 | 0.9905 | 0.9905 | 0.9904 | 0.9904 |
| 1.8676 | 2.0543 | 8 | 0.9885 | 0.9885 | 0.9885 | 0.9885 |
| 2.0401 | 2.2441 | 9 | 0.9864 | 0.9864 | 0.9864 | 0.9864 |
| 2.2228 | 2.4451 | 10 | 0.9841 | 0.9841 | 0.9840 | 0.9840 |

At the displayed level of precision we see almost no difference in value until we get out to 10 years. Since the first two “financial engineering” methods use a continuous time approach the cost of capital parameter used is $5.83\% = \ln(1.06)$ for them.

A more sensitive way to compare the methods is to calculate the implied mortality rates that connect the calculated sequence of endowment values. We do this in table 3b below. As noted earlier, the risk loaded mortality rates were calculated first for the financial engineering methods.

| Table 3b: Implied Risk Loaded Mortality 1000 qx~ | | | | |
|---|-----------------|-----------------|-----------------|--------------------|
| | $\alpha= 100\%$ | | $i= 0.00\%$ | |
| | Simple | Explicit | Implicit | Prospective |
| | Mean | | | $\theta=0.00\%$ |
| $\pi=$ | 5.83% | 5.83% | 6.00% | 6.00% |
| 1 | 1.0179 | 1.0179 | 1.0211 | 1.0211 |
| 2 | 1.1160 | 1.1160 | 1.1196 | 1.1196 |
| 3 | 1.2254 | 1.2254 | 1.2296 | 1.2296 |
| 4 | 1.3464 | 1.3464 | 1.3512 | 1.3512 |
| 5 | 1.4791 | 1.4791 | 1.4845 | 1.4845 |
| 6 | 1.6237 | 1.6236 | 1.6299 | 1.6299 |
| 7 | 1.7803 | 1.7803 | 1.7874 | 1.7874 |
| 8 | 1.9492 | 1.9491 | 1.9572 | 1.9572 |
| 9 | 2.1412 | 2.1411 | 2.1502 | 2.1503 |
| 10 | 2.3459 | 2.3458 | 2.3561 | 2.3561 |

We now see small differences between all four methods with a larger apparent difference between the two financial engineering models and the two “actuarial” discrete time models. To help explain this difference we compute the effective beta factor for each model in Table 3c below.

| Table 3c: Implied beta factor $qx\sim = q x + \beta (qx^{\wedge}-qx)$ | | | | |
|---|-----------------|-----------------|-----------------|--------------------|
| | $\alpha= 100\%$ | | $i= 0.00\%$ | |
| | Simple | Explicit | Implicit | Prospective |
| | Mean | | | $\theta=0.00\%$ |
| $\pi=$ | 5.83% | 5.83% | 6.00% | 6.00% |
| 1 | 2.91% | 2.91% | 6.00% | 6.00% |
| 2 | 8.74% | 8.74% | 12.00% | 12.00% |
| 3 | 14.57% | 14.57% | 18.00% | 18.00% |
| 4 | 20.40% | 20.39% | 24.00% | 24.00% |
| 5 | 26.22% | 26.21% | 29.99% | 29.99% |
| 6 | 32.05% | 32.04% | 35.99% | 35.99% |
| 7 | 37.88% | 37.86% | 41.99% | 41.99% |
| 8 | 43.70% | 43.68% | 47.98% | 47.99% |
| 9 | 49.53% | 49.50% | 53.98% | 53.98% |
| 10 | 55.36% | 55.31% | 59.97% | 59.98% |

Because the first two models use continuous time, the cost of capital is based on an average capital requirement for the year. Our particular approach to discretizing the second two methods, bases the cost of capital on the beginning of year value. As both Tables 1& 2 show the required capital is declining over time so we expect to see a bias in Table 3c.

Other discretization approaches are possible. If the calculations were being done on a monthly time step, as opposed to annual, the results would be much closer.

The discrepancies do grow if go further out in time but the impact on calculated present values is often small as a result of discounting with interest and persistency.

Another small difference between the continuous and discrete time methods arises when we use a non-zero interest rate. Table 4c below show what happens if we use an annual time step with $i = 5.00\%$ interest rate.

| Table 4c: Implied beta factor $qx \sim = q x + \beta (qx^{\wedge} - qx)$ | | | | |
|--|------------------|-----------------|-----------------|--------------------|
| | $\alpha = 100\%$ | | $i = 5.00\%$ | |
| | Simple | Explicit | Implicit | Prospective |
| | Mean | | | $\theta = 0.00\%$ |
| $\pi =$ | 5.83% | 5.83% | 6.00% | 6.00% |
| 1 | 2.91% | 2.91% | 5.71% | 5.71% |
| 2 | 8.74% | 8.74% | 11.43% | 11.43% |
| 3 | 14.57% | 14.57% | 17.14% | 17.14% |
| 4 | 20.40% | 20.39% | 22.85% | 22.85% |
| 5 | 26.22% | 26.21% | 28.57% | 28.57% |
| 6 | 32.05% | 32.04% | 34.28% | 34.28% |
| 7 | 37.88% | 37.86% | 39.99% | 39.99% |
| 8 | 43.70% | 43.68% | 45.70% | 45.70% |
| 9 | 49.53% | 49.50% | 51.41% | 51.41% |
| 10 | 55.36% | 55.31% | 57.12% | 57.12% |

The continuous time model results have not changed but this particular discretization of the actuarial methods is now behaving as if the cost of capital rate had dropped down to $5.71\% = 6.00\%/1.05$. Again, this is an issue that goes away if we use a shorter time step.

There is final issue that can arise if we use the Prospective (Solvency II) method with an illiquidity premium. Table 5c below shows what happens if we assume a 50 bp illiquidity premium.

| Table 5c: Implied beta factor $qx \sim = q x + b (qx^{\wedge} - qx)$ | | | | |
|--|---------------|-----------------|-----------------|--------------------|
| | $\alpha =$ | 100% | $i =$ | 5.00% |
| | Simple | Explicit | Implicit | Prospective |
| | Mean | | | $\theta = 0.50\%$ |
| $\pi =$ | 5.83% | 5.83% | 6.00% | 6.00% |
| 1 | 2.91% | 2.91% | 5.71% | 5.71% |
| 2 | 8.74% | 8.74% | 11.43% | 11.46% |
| 3 | 14.57% | 14.57% | 17.14% | 17.22% |
| 4 | 20.40% | 20.39% | 22.85% | 23.02% |
| 5 | 26.22% | 26.21% | 28.57% | 28.84% |
| 6 | 32.05% | 32.04% | 34.28% | 34.69% |
| 7 | 37.88% | 37.86% | 39.99% | 40.57% |
| 8 | 43.70% | 43.68% | 45.70% | 46.47% |
| 9 | 49.53% | 49.50% | 51.41% | 52.40% |
| 10 | 55.36% | 55.31% | 57.12% | 58.37% |

The results for the first three methods have not changed. When a liquidity premium is used for the Prospective Method we are allowed to use the interest rate $5.50\% = 5.00\% + .50\%$ when calculating best estimate values but a discount rate of only 5.00% is used to compute the present value of risk margins. What Table 5c tells us is that we can reproduce the Solvency II result by using mortality rates with slightly higher loadings and then discounting the resulting risk loaded cash flows using the liquidity enhanced valuation rate of 5.50% .

The modified prospective method would then, in total, produce a more liberal result than the other methods. The mathematical details of this equivalence are fully documented in the main paper.

One of the key conclusions of this analysis is that we can approach the cost of capital method from either the financial engineering or actuarial perspectives. The main paper calls this duality. The approach you take can be decided by practical rather than theoretical considerations. The main paper argues that the actuarial approach is fine for simple problems but, as the complexity of the application increases, the financial engineering approach becomes easier to implement.

Examples of more complex applications include universal life, mortality improvement and long term care. The main paper also gives references to where the author has extended this cost of capital method to value long dated options and credit risky bonds.

Appendix: Discrete Time Implementation of the Financial Engineering Short Cuts

Both of the financial engineering short cuts have practical discrete time implementations provided we are willing to assume the force of decrement is piecewise constant by contract year. This is a simplifying assumption often used by practicing actuaries.

Suppose the issue age of our life is x and we measure time from the issue date. On the valuation date the life is aged $x + t$ and our best estimate mortality rate plus contagion load is $q_{[x]+s}$ and the parameter shocked value is $\hat{q}_{[x]+s}$ for $s \geq t$. We are using standard select and ultimate notation. The goal is to come up with a practical way to calculate risk loaded mortality rates $q_{\{[x]+t\}+s}$ for $s \geq 0$ that reflect the impact of adding an appropriate dynamic margin for parameter risk after the valuation date. This is a doubly select and ultimate structure since the risk margin process begins on the valuation date.

Define the decrement shock $\Delta\mu_s$ by

$$1 - \hat{q}_{[x]+t+s} = (1 - q_{[x]+t+s})e^{-\Delta\mu_s}, \quad s = 0,1,2, \dots$$

For the simple mean approximation, we can go from one loaded persistency factor ${}_s\bar{p}_{[x]+t}$ to the next by

$$\begin{aligned} {}_{s+1}\bar{p}_{[x]+t} &= {}_s\bar{p}_{[x]+t} e^{-\int_s^{s+1} (\mu(v) + \bar{\beta}(v)\Delta\mu_s) dv} \\ &= {}_s\bar{p}_{[x]+t} (1 - q_{[x]+t+s}) e^{-\int_s^{s+1} \bar{\beta}(v)\Delta\mu_s dv}, \\ &= {}_s\bar{p}_{[x]+t} (1 - q_{[x]+t+s}) e^{-\Delta\mu_s \int_s^{s+1} \bar{\beta}(v) dv} \end{aligned}$$

Since we know

$$\bar{\beta}(v) = \begin{cases} \frac{1 - e^{-\pi(1-\alpha)v}}{1-\alpha}, & \alpha < 1 \\ \pi v & \alpha = 1 \end{cases}$$

$$\text{we can calculate } k_s = \int_s^{s+1} \bar{\beta}(v) dv = \begin{cases} \frac{1 - e^{-\pi(1-\alpha)s} (1 - e^{-\pi(1-\alpha)}) / (\pi(1-\alpha))}{1-\alpha}, & \alpha < 1 \\ \pi \left(s + \frac{1}{2} \right), & \alpha = 1 \end{cases}.$$

The risk loaded mortality rate is then given by

$$1 - q_{\{[x]+t\}+s} = (1 - q_{[x]+t+s}) \left(\frac{1 - \hat{q}_{[x]+t+s}}{1 - q_{[x]+t+s}} \right)^{k_s}, \quad s = 0,1,2, \dots$$

This is obviously easy to implement. A corresponding margined mortality rate for a shocked scenario is approximately⁸

$$1 - \hat{q}_{\{[x]+t\}+s} = (1 - \hat{q}_{[x]+t+s}) \left(\frac{1 - \hat{q}_{[x]+t+s}}{1 - q_{[x]+t+s}} \right)^{\alpha k_s}.$$

The table below shows a spreadsheet implementation of the method described above.

⁸ The theoretical error in this approximation is usually immaterial. For more details on the nature of the error see the main paper.

Simple Mean Approximation to Margined Mortality

| s | 1000 qx | | $\alpha = 100\%$ | $\pi = 5.83\%$ | |
|---|--------------|-----------------------|------------------|----------------|-------------------------|
| | Inputs | | ks | Loaded Results | |
| | Base | Shocked | | Base | Shocked |
| | $q[x]_{t+s}$ | $q^{\wedge}[x]_{t+s}$ | | $q\{[x]_{t+s}$ | $q^{\wedge}\{[x]_{t+s}$ |
| 0 | 1.0150 | 1.1165 | 2.91% | 1.0179 | 1.1194 |
| 1 | 1.1063 | 1.2170 | 8.74% | 1.1160 | 1.2266 |
| 2 | 1.2078 | 1.3286 | 14.57% | 1.2254 | 1.3462 |
| 3 | 1.3195 | 1.4514 | 20.39% | 1.3464 | 1.4783 |
| 4 | 1.4413 | 1.5854 | 26.22% | 1.4791 | 1.6232 |
| 5 | 1.5732 | 1.7306 | 32.05% | 1.6237 | 1.7810 |
| 6 | 1.7153 | 1.8869 | 37.87% | 1.7803 | 1.9518 |
| 7 | 1.8676 | 2.0543 | 43.70% | 1.9492 | 2.1359 |
| 8 | 2.0401 | 2.2441 | 49.53% | 2.1412 | 2.3452 |
| 9 | 2.2228 | 2.4451 | 55.36% | 2.3459 | 2.5681 |

The discrete time implementation of the *explicit margin method* is a bit more involved but, as we saw earlier, it is a theoretically superior method. The starting point for this method is the relation

$$1 - q_{\{[x]_{t+s}\}} = (1 - q_{[x]_{t+s}}) e^{-\int_s^{s+1} \beta(v) \Delta\mu_s dv}.$$

For explicit margin method we can derive the identity⁹ $1 - (1 - \alpha)\beta(s) = e^{-(1-\alpha)\int_t^s [\pi - \beta\Delta\mu] dv}$. Using this expression we can write, assuming $0 \leq \alpha < 1$

$$1 - q_{\{[x]_{t+s}\}} = (1 - q_{[x]_{t+s}}) \left[\frac{1 - (1 - \alpha)\beta(s+1)}{1 - (1 - \alpha)\beta(s)} \right]^{\frac{1}{1-\alpha}} e^{-\pi}.$$

The next step is to consider the quantity $J(s) = \frac{\beta(s)}{1 - (1 - \alpha)\beta(s)}$. The evolution equation for β implies that J satisfies the linear differential equation

$$\frac{dJ}{ds} = \pi + J[\pi(1 - \alpha) - \Delta\mu(s)], \quad J(t) = 0.$$

This can be solved in closed form under our simplifying assumption that $\Delta\mu(s)$ is piecewise constant. The result is

$$J(s+1) = J(s) e^{\pi(1-\alpha) - \Delta\mu_s} + \pi \frac{e^{\pi(1-\alpha) - \Delta\mu_s} - 1}{\pi(1-\alpha) - \Delta\mu_s}.$$

This is easily programmed in any language.

If we have $J(s)$ we can recover β by using $\beta = J/[1 + (1 - \alpha)J]$ or, better yet, use the relation

⁹ This follows by rewriting the evolution equation $d\beta^e = (\pi - \beta^e \Delta\mu(s))(1 - (1 - \alpha)\beta^e) ds$ in the equivalent form $d\beta^e / (1 - (1 - \alpha)\beta^e) = (\pi - \beta^e \Delta\mu(s)) ds$ and then integrating.

$$1 - (1 - \alpha)\beta = 1/[1 + (1 - \alpha)J]$$

to write

$$1 - q_{\{[x]+t\}+s} = (1 - q_{[x]+t+s}) \left[\frac{1 + (1 - \alpha)J(s + 1)}{1 + (1 - \alpha)J(s)} \right]^{\frac{1}{1-\alpha}} e^{-\pi}.$$

If $\alpha = 1$ the limiting form of this equation is

$$1 - q_{\{[x]+t\}+s} = (1 - q_{[x]+t+s}) e^{-\pi+J(s+1)-J(s)}.$$

For the shocked scenario, the relevant results are

$$1 - \hat{q}_{\{[x]+t\}+s} = (1 - \hat{q}_{[x]+t+s}) \left[\frac{1 + (1 - \alpha)J(s + 1)}{1 + (1 - \alpha)J(s)} \right]^{\frac{\alpha}{1-\alpha}} e^{-\pi\alpha}, \quad 0 \leq \alpha < 1$$

$$1 - \hat{q}_{\{[x]+t\}+s} = (1 - \hat{q}_{[x]+t+s}) e^{-\pi+J(s+1)-J(s)}, \quad \alpha = 1 .$$

The main point to be emphasized here is that the actual implementation is not onerous. The spreadsheet table below illustrates the method.

Explicit Margin Method Example

| s | 1000 qx | | $\alpha = 100\%$ | $\pi = 5.83\%$ | Loaded Results | |
|---|---------------|------------------------|-----------------------------|----------------|-------------------|----------------------------|
| | Inputs | | | | Loaded Results | |
| | Base | Shocked | | | Base | Shocked |
| | $q_{[x]+t+s}$ | $q^{\wedge}_{[x]+t+s}$ | $\pi(1-\alpha)-\Delta\mu_s$ | J_s | $q_{\{[x]+t\}+s}$ | $q^{\wedge}_{\{[x]+t\}+s}$ |
| 0 | 1.0150 | 1.1165 | -0.01% | 0.00% | 1.0179 | 1.1194 |
| 1 | 1.1063 | 1.2170 | -0.01% | 5.83% | 1.1160 | 1.2266 |
| 2 | 1.2078 | 1.3286 | -0.01% | 11.65% | 1.2254 | 1.3462 |
| 3 | 1.3195 | 1.4514 | -0.01% | 17.48% | 1.3464 | 1.4783 |
| 4 | 1.4413 | 1.5854 | -0.01% | 23.30% | 1.4791 | 1.6232 |
| 5 | 1.5732 | 1.7306 | -0.02% | 29.12% | 1.6236 | 1.7809 |
| 6 | 1.7153 | 1.8869 | -0.02% | 34.95% | 1.7803 | 1.9518 |
| 7 | 1.8676 | 2.0543 | -0.02% | 40.77% | 1.9491 | 2.1359 |
| 8 | 2.0401 | 2.2441 | -0.02% | 46.59% | 2.1411 | 2.3451 |
| 9 | 2.2228 | 2.4451 | -0.02% | 52.40% | 2.3458 | 2.5680 |

As the numerical examples show, there is not much difference, in practice, between the two short cuts. Given that they are both fairly easy to implement, but the explicit method has better theory, the author prefers the explicit method to the simple mean.