



SOCIETY OF ACTUARIES

2014 Enterprise Risk Management Symposium
Sept. 29 - Oct. 1, 2014, Chicago, IL

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

By B. John Manistre¹, Ph.D., FSA, CERA, FCIA, MAAA

Copyright 2015 by the Society of Actuaries, Casualty Actuarial Society, and the Professional Risk Managers' International Association.

All rights reserved by the Society of Actuaries, Casualty Actuarial Society, and the Professional Risk Managers' International Association. Permission is granted to make brief excerpts for a published review. Permission is also granted to make limited numbers of copies of items in this monograph for personal, internal, classroom or other instructional use, on condition that the foregoing copyright notice is used so as to give reasonable notice of the Society of Actuaries', Casualty Actuarial Society's, and the Professional Risk Managers' International Association's copyright. This consent for free limited copying without prior consent of the Society of Actuaries, Casualty Actuarial Society, and the Professional Risk Managers' International Association does not extend to making copies for general distribution, for advertising or promotional purposes, for inclusion in new collective works or for resale.

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

By B. John Manistre,¹ Ph.D., FSA, CERA, FCIA, MAAA

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 that 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 that can be solved by Monte Carlo simulation. The paper then goes on to develop a number of more pragmatic methods that can be thought of as shortcut approximations to the first-principles model. One of these shortcuts is the prospective method currently used in Europe. None of these methods require stochastic-on-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 paper 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 principle-based, fair valuation system where continuous model improvement is an integral part of the risk management process. One possible application would be to an internal economic capital model or be the base for an Own Risk and Solvency Assessment (ORSA) process. The author believes the methods described here would also meet International Financial Reporting Standards (IFRS) for risk margins and could serve as a foundation for a market-consistent embedded value (MCEV) process.

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.³ The quantity π

¹ The author is a research actuary at GGY AXIS based in Baltimore, Maryland.

² George E.P. Box (FRS) in 1987.

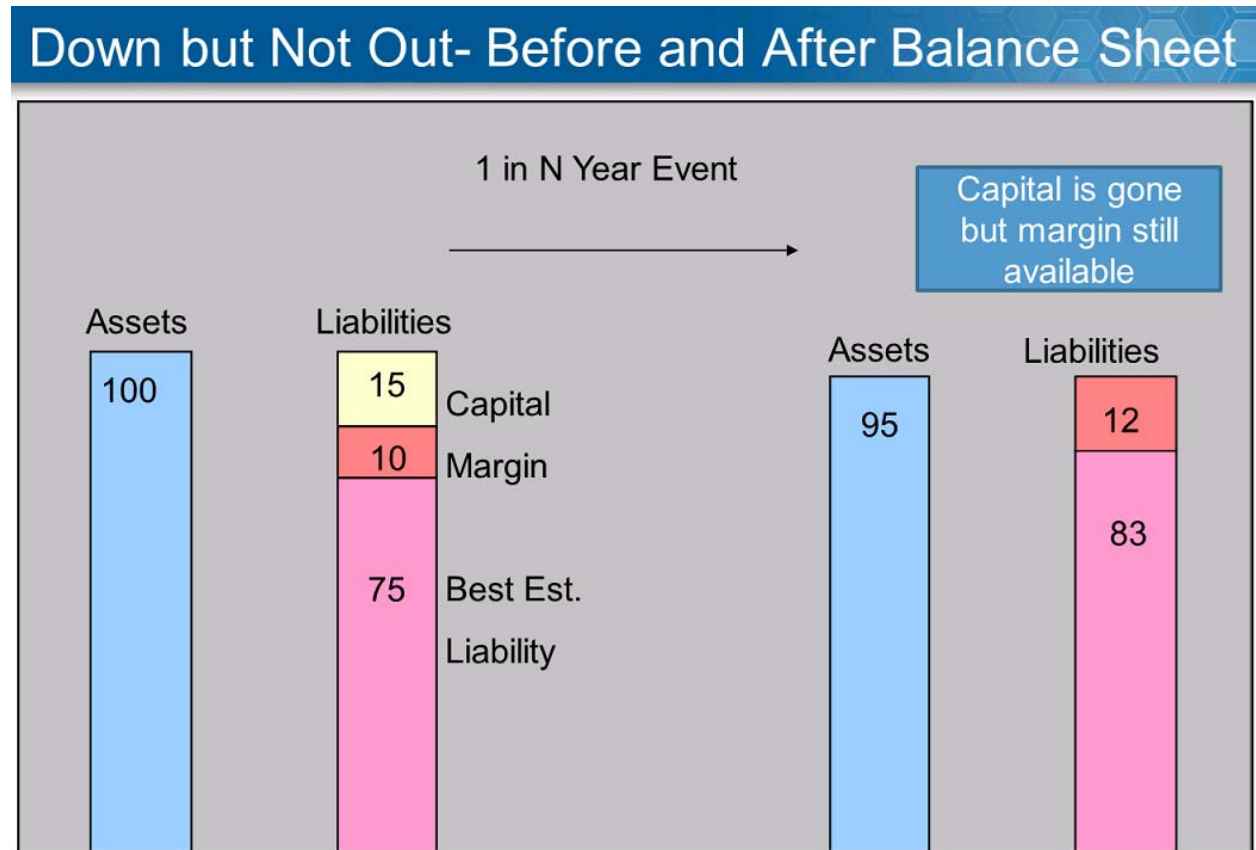
³ A prima-facie case for using the cost of capital method in risk management is developed in a 2006 paper published by the CRO Forum. It can be found at <http://www.thecroforum.org/a-market-cost-of-capital-approach-to-market-value-margins/>.

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 $\pi = 0.06$.

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

- The Down but Not Out principle
- Linearity
- A three-step risk modeling process.

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. Near the end of the paper, we argue that 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. This is what most people understand the cost of capital method to mean.

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 enough set of risk-neutral scenarios.⁴

With this result in hand 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.

⁴ This is a standard result in stochastic calculus, which is outlined in the Appendix.

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.

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. We illustrate this later 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:** As stated in the abstract, we assume 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 contain parameters that are subject to revision.
 - b. Hold capital and margins for a single large contagion event. A contagion event is often one where the law of large numbers fails to work as a risk management tool. This is often because some unusual event has caused a large number of normally independent risks to behave in a correlated way.

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 1990s⁵ and the well-known problems with the U.S. residential mortgage market that led to the financial crisis of 2008.

⁵ This was caused by the overbuilding of office space during the 1980s 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.

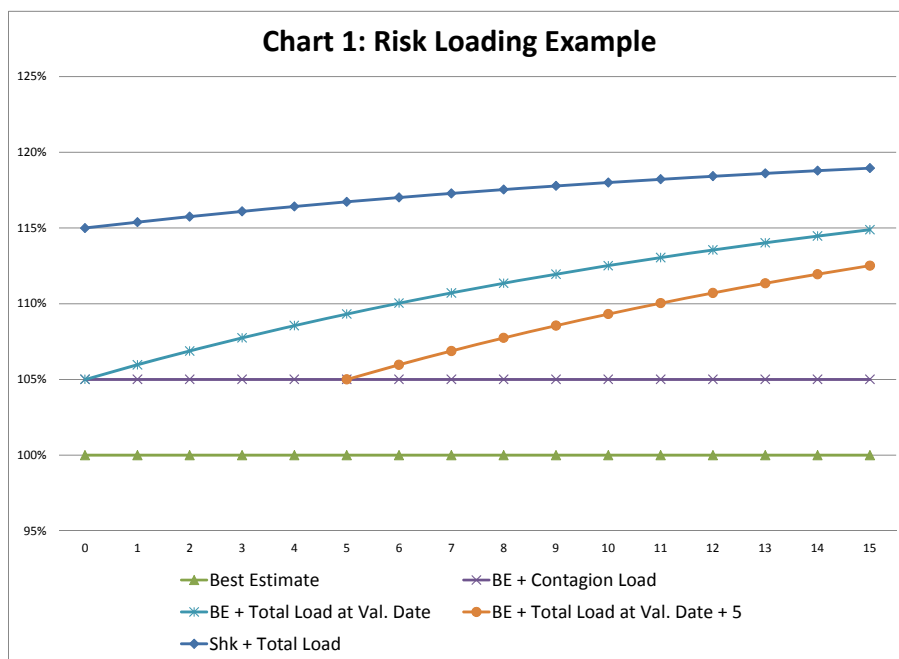
A risk enterprise should have sufficient capital and margins that it can withstand a plausible contagion event and still be able to continue as a going concern without regulatory intervention. We show that traditional, static, risk loadings in our parameters can usually deal with this issue.

- 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 parameters in a model. To the extent this model revision causes 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 margin that arises naturally out of the dual approach.

Static and dynamic margins 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 to the parameter that 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, one reporting period later, we establish a new dynamic margin that 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 contagion loading of 5 percent has been added. At the valuation date ($t = 0$), we have added a dynamic margin that takes the parameter up to the value of 115 percent 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 percent (base + 10 percent) and then grades to about 119 percent. Economic capital, for parameter risk, is the difference between the shocked and base fair values.

When we come to do a new valuation five 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 this paper is that the cost of capital method is, for most practical purposes, equivalent to an appropriate combination of static and dynamic margins.

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** to get useful results.

Following this introduction, the paper is organized as follows:

- 1) The first main section works through the risk modeling ideas introduced above for the simple example of a term life insurance risk with no lapses, i.e., a pure mortality risk. We solve the contagion issue and develop a first-principles approach to the parameter risk issue of a shock $\Delta\mu$ to the level of assumed mortality. The primal version of the model turns out to be, in theory, infinite-dimensional, and we show how the resulting dynamic margins can be calculated by Monte Carlo simulation. We call this the *Brute Force* approach.
- 2) The second main section starts by introducing two “actuarial” shortcuts which we call the *implicit* and *prospective* methods, respectively. Each shortcut makes a slightly different simplifying assumption that allows the primal version of the model to be solved fairly easily. We then formulate the dual version of each model, which allows us to compare the two approaches. The section concludes with a fairly simple algorithm for calculating the dynamic margins associated with each shortcut.
- 3) The third section develops two “financial engineering” shortcuts to the first-principles model, which we call the *simple mean* and *explicit* methods, respectively. These methods start by making a simplifying assumption about the dual model. Under certain useful simplifying assumptions, the explicit method turns out to be a one-dimensional model that is an exact dual to the infinite-dimensional first-principles model.

- 4) The fourth section summarizes the single risk models in their dual forms. We are able to rank the relative conservatism of the four shortcuts. The rank depends on the sign of the shock $\Delta\mu$. We give an extreme example where the four methods produce very different results but then show there is a wide range of practical problems where all four methods produce very similar results.
- 5) The fifth section discusses the issues of risk interaction and diversification that arise when we have more than one risk to consider.

Risk interaction deals with the issue of whether mortality margins are available to partially offset lapse rate risks and vice versa. We argue that the answer should be yes and call this “natural interaction.” If this position is taken, then the assumption margins determined at the single risk level are simply carried over to the multi-risk problem without modification. Other approaches are possible, though. The most common alternative is to calculate margins and capital one risk at a time and then sum the results. We call this “no interaction.”

Risk diversification deals with the traditional issue of how statistical diversification benefits among different risks could or should be handled. We give a very brief discussion here. A more detailed discussion can be found in Manistre⁶ [2008]. The main point we make is that there are reasonable, and practical, ways to handle the diversification issue within the “Down but Not Out” framework.

- 6) The sixth section summarizes the models developed in this paper and discusses some of their practical pros and cons. If the numerical answers are not materially different, then the choice of approach can be based on more pragmatic concerns. We argue that, for simple problems, the primal version of a model is fairly easy to understand and implement, but, as the complexity of the application grows, the dual approach can be easier to implement.
- 7) The final section discusses the idea that “Down but Not Out” won’t work in practice unless the models and assumptions used have credibility with the investing public. Putting an appropriate set of professional standards of practice in place is one way to develop the required credibility.
- 8) Lastly, we include an appendix that expands on the issue of why linear models can get around the “stochastic on stochastic” issue.

First Principles and the Brute Force Approach

In this section we will apply the cost of capital concepts described in the introduction to a simple life contingency issue. We simplify the economic assumptions in much the same way as is done when deriving the Black- Scholes option pricing formula. This means we use a deterministic continuous time interest rate $r(t)$.

⁶ Manistre, B.J. “A Practical Concept of Tail Correlation,” Proceedings of the 2008 ERM Symposium. This paper can be found on the Society of Actuaries website.

A Simple Term Life Example—The Best-Estimate Model

For this example we 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. 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 the differential equation

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

An intuitive way to understand this result 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. To see this, note that at time t , when the insured is alive, the liability is $L(t) = V_0(t)$. At time $t + \Delta t$ the liability has two possible values depending on whether the insured life dies during time interval Δt . If the probability of death is $\mu_0(t)\Delta t$ then the expected change in liability is

$$\begin{aligned} E[\Delta L] &= [F - V_0]\mu_0(t)\Delta t + [V_0(t + \Delta t) - V_0(t)][1 - \mu_0(t)\Delta t], \\ &= \left\{ [F - V_0]\mu_0(t) + \frac{dV_0}{dt} \right\} \Delta t + o\Delta t^2. \end{aligned}$$

Assuming we have assets equal to liabilities at the beginning of the time step, the change in assets is

$$\Delta A = (rV_0 + g - e)\Delta t + o\Delta t^2.$$

Setting $E[\Delta L] = \Delta A$ and then dividing by Δt gives the differential equation stated above when we take the limit as $\Delta t \rightarrow 0$.

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

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

This equation says that V_0 is the actuarial present value of death benefits and expenses offset by the present value of gross premiums.

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 $V = V_0 + M$ 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 capital⁷ rate, the new valuation equation should include the cost of contagion risk capital as an additional expense, i.e.,

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

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 very 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 percent level).

We can turn this into a payout annuity model if we set $F = g = 0$ in this example and we use a rate of payment p :

$$\frac{dV}{dt} + \mu_0(t)[0 - V] = rV - p - e - \pi\Delta Q[0 - V].$$

We would need to set $\Delta Q < 0$ if we want to add a conservative margin to the annuity value. This raises an interesting issue.

If a company had both life and annuity risks in its portfolio, then one could argue that there is a natural hedge between the life and annuity blocks and, if there were a repeat of the 1918 flu epidemic, conclude that using $\Delta Q > 0$ for both types of business makes sense. Not all actuaries would agree with this conclusion. Ideally, the issue would be resolved by some form of industry consensus or professional standard.

A detailed discussion of exactly how the contagion shock ΔQ should be set is beyond the scope of this paper. However, there are some useful research resources, available in the public domain,⁸ which summarize mortality contagion event experience in the 20th century and also discuss the additional issues one should consider when developing such a shock. For example, even if we take the 1918 flu epidemic to be a representative data point, $\Delta Q \approx 4/1,000$, we still have to consider how modern health care systems would react to such an event if it were repeated today⁹.

⁷ In this paper the cost of capital rate refers to the expected pre-tax return to the shareholder in excess of the risk-free rate. A typical value is $\pi = 6\%$.

⁸ One reference is the 2007 Swiss Re publication "Pandemic Influenza: A 21st Century Model for Mortality Shocks": http://media.swissre.com/documents/pandemic_influenza_a_21st_century_model_en.pdf.

⁹ The current Solvency II specification in Europe is $\Delta Q = 1.5/1,000$.

It is worth discussing how 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 insurance 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. Nevertheless, 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.

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 that 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 percent 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 level of industry consensus around the principles used to choose the shocks.

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 shortcut methods. There is a wide range of practical problems where all four shortcuts produce very similar results.

Before introducing the shortcuts, we do some analysis to understand what a no-compromise or first-principles approach would look like.

The Brute Force Method

Suppose we were actually willing to contemplate an infinite hierarchy of mortality assumptions $\mu, \mu + \Delta\mu, \mu + \Delta\mu + \Delta\mu^{(2)}, \dots$. For a simple example, we might assume a geometric hierarchy where $\Delta\mu^{(n)} = \alpha^{n-1}\Delta\mu$ for some constant $0 \leq \alpha \leq 1$. The n' th level in the hierarchy would then be given by

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

However we choose the shock hierarchy, consider the n' th equation for $\hat{V}^{(n)}$, which we can write as

$$\frac{d\hat{V}^{(n)}}{dt} + \mu^{(n)}(F - \hat{V}^{(n)}) = r\hat{V}^{(n)} + g - e - \pi(\hat{V}^{(n+1)} - \hat{V}^{(n)}).$$

If we rewrite this as

$$\frac{d\hat{V}^{(n)}}{dt} + \mu^{(n)}(F - \hat{V}^{(n)}) + \pi(\hat{V}^{(n+1)} - \hat{V}^{(n)}) = r\hat{V}^{(n)} + g - e,$$

then standard actuarial ideas allow us to say that $\hat{V}^{(n)}$ should be calculated by assuming the force of mortality is $\mu^{(n)}$ and π is the force of transition for a jump from the n' th level of the shock hierarchy to the $(n + 1)$ 'st level.

Another way to see this is to think in terms of a health impairment model based on two very simple assumptions.

1. Each level in the shock hierarchy has a different force of mortality corresponding to a different degree of impairment.
2. The probability of a life becoming more impaired in any given time interval Δt is $\pi\Delta t$.

This model is a one-way street. Lives never recover and drop back down the shock hierarchy.

We have basically shown that the formal solution to the infinite system of valuation equations can be calculated 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 .

In symbols, we can write

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

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 the 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 this model. One merely has to specify the shock hierarchy and then perform a large number of Monte Carlo simulations, hence the name “Brute Force.”

There is a mathematical way to simplify the result of the Monte Carlo approach and present the answer as an actuarial present value calculated along an equivalent single, loaded, mortality scenario. Let

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

be the expected persistency factor in the C measure. If we can get our hands on this quantity, perhaps by Monte Carlo simulation, then we can calculate an effective loaded force of mortality $\tilde{\mu}$ from the standard relation

$$\begin{aligned} {}_s\bar{p}_t \tilde{\mu}_{t+s} &= - \frac{d}{ds} {}_s\bar{p}_t, \\ &= - \frac{d}{ds} E_{C(t)} \exp \left[- \int_t^{t+s} \mu dv \right], \\ &= E_{C(t)} \left\{ \exp \left[- \int_t^{t+s} \mu dv \right] \mu(t+s) \right\}. \end{aligned}$$

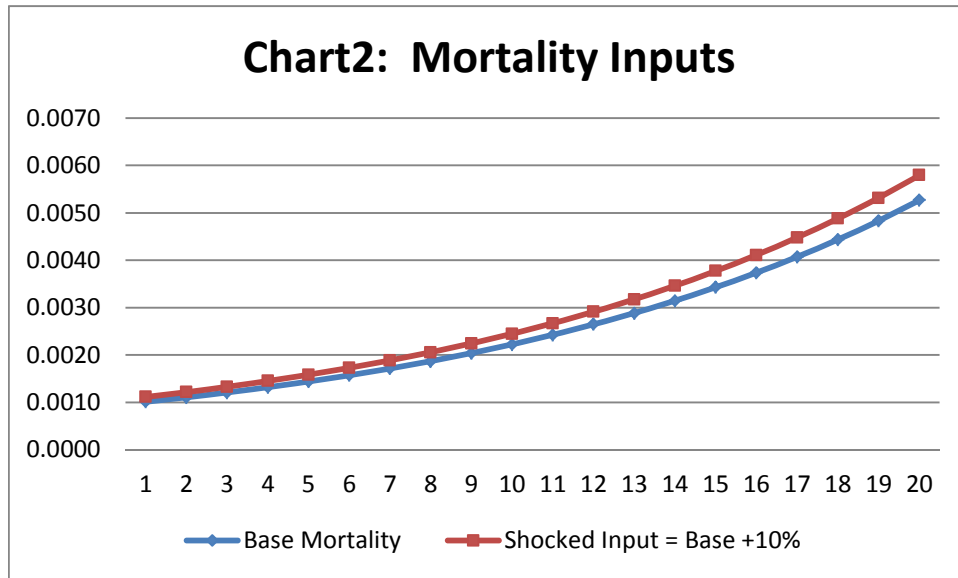
The solution to the valuation problem can then be written¹⁰ as a traditional actuarial present value using the risk-loaded force of mortality $\tilde{\mu}_{t+s}$, i.e.,

¹⁰ This is possible as long as the actuarial application being developed is linear in the persistency factor. This is true for most life insurance and annuity products. It would not be true for stop loss reinsurance. Additional work would be required to adapt the ideas presented here to a non-linear application.

$$V(t) = \int_t^\infty e^{-\int_t^s r dv} {}_s\bar{p}_t [\tilde{\mu}_{t+s} F + e - g] ds.$$

This shows that all we really need to know is the expected persistency factor ${}_s\bar{p}_t$ and that is equivalent to knowing the risk loaded mortality scenario $\tilde{\mu}_{t+s}$. We will call this quantity the Equivalent Single Scenario (ESS) mortality assumption associated with the model.

The next two charts present the results of estimating dynamic margins by Monte Carlo simulation. Chart 2 shows the input contagion-loaded mortality rates q_{t+s}, \hat{q}_{t+s} . The parameter shock is a 10 percent increase, i.e., $\hat{q}_x = 1.10 q_x$.



The next chart assumes a geometric shock hierarchy with $\alpha = 1$ and $\pi = 0.06$. All results are presented as a ratio to the base mortality and are based on 1,000 Monte Carlo simulations so there is some sampling error in the results.

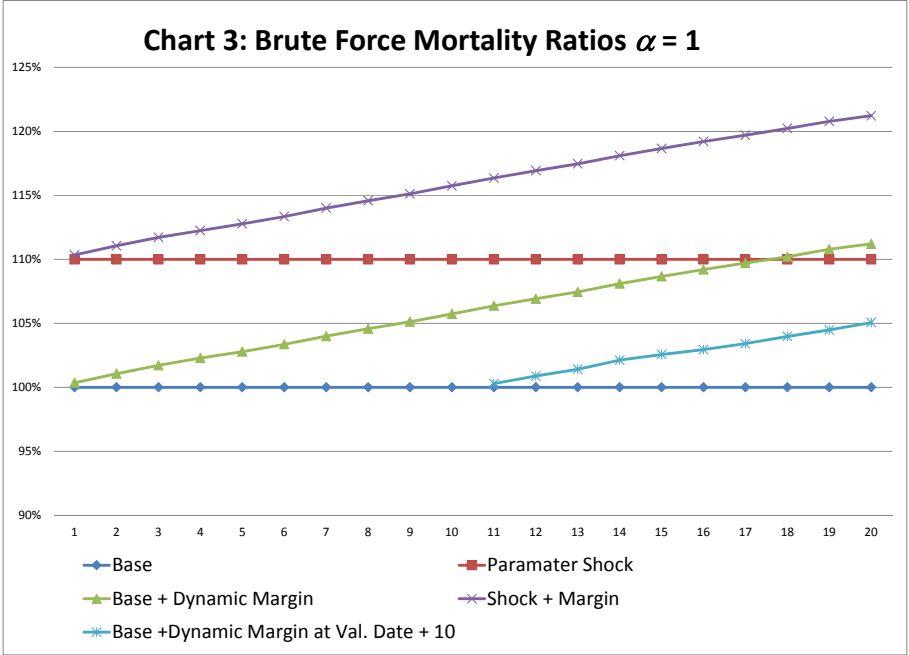
The actual simulations were done using continuous time mathematics¹¹ under the assumption that the forces of mortality $\mu_x = -\ln(1 - q_x)$ and $\mu_x + \Delta\mu_x = -\ln(1 - \hat{q}_x)$ are piecewise constant by year of age. The simulation results were used to calculate the average persistency ${}_sP_t, {}_s\hat{P}_t$ under both the base and shocked regime-switching measures. Risk loaded mortality rates were then estimated using

$$1 - q_{\{t\}+s} = \frac{{}_{s+1}P_t}{{}_sP_t},$$

$$1 - \hat{q}_{\{t\}+s} = \frac{{}_{s+1}\hat{P}_t}{{}_s\hat{P}_t},$$

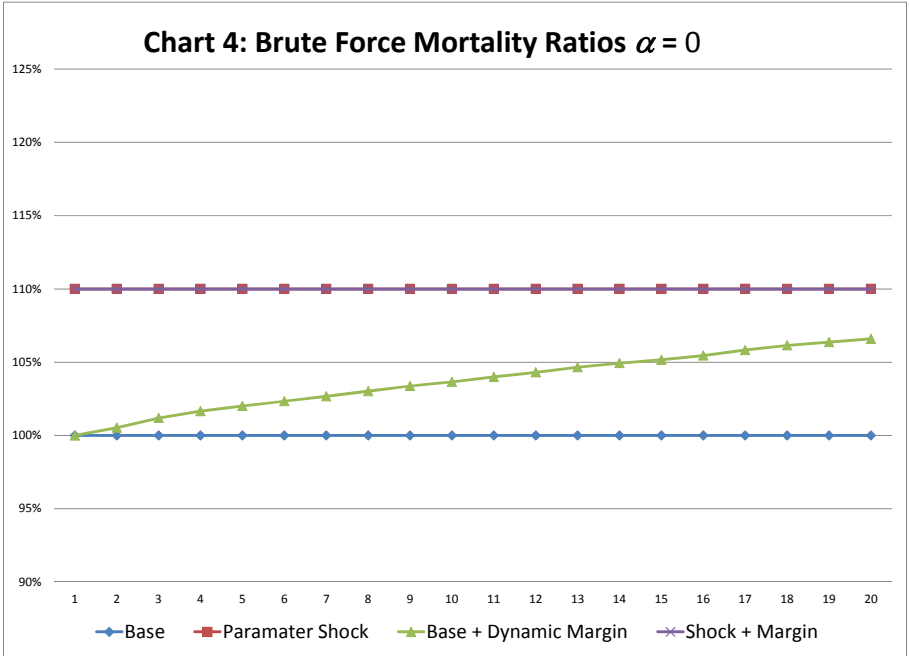
and presented as a ratio to the base input q_{t+s} .

¹¹ The Monte Carlo model will not be used again in this paper.



The two level curves correspond to the input assumptions at 100 percent and 110 percent, respectively. The first two sloping curves show the estimated ESS under the base and shocked measures at $t = 0$. The final curve shows the base ESS as it would be recalculated at time $t = 10$. This is clearly an example of a dynamic margin structure.

The final chart in this section shows how the results change, if we base the shock hierarchy on the parameter value $\alpha = 0$. Making this choice is equivalent to assuming that, after the first shock, things can never get any worse. There is therefore no margin required, once we are in the shocked regime.



This shock hierarchy is clearly less conservative than assuming $\alpha = 1$ so there is less risk margin than in the base case, as would be expected.

ESS mortality is not the same thing as the expected mortality $\bar{\mu}(s) = E_{C(t)}[\boldsymbol{\mu}(s)]$ but it is often close if the decrement shock $\Delta\mu$ is small, as it usually is for mortality problems.

We can then get some insight into why the ESS looks like it does by calculating the expected force of mortality for the geometric shock hierarchy $\Delta\mu^{(n)} = \alpha^{n-1}\Delta\mu$. As stated earlier, for this hierarchy, the n 'th level force of mortality is

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

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 shock hierarchy can then be calculated in closed form as

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

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

To the extent we are willing to accept the simplifying assumption¹² that

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

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 margin that grades from 0 toward 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.

To help compare this approximation with other methods described later, 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

¹² We will see later that this is a conservative approximation to the exact geometric hierarchy model if $\Delta\mu > 0$.

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

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

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

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 hierarchy.¹³

The Monte Carlo numerical examples presented in Charts 3 and 4 are clearly consistent with the approximations described above.

We will see later that most of the theoretical error that arises from assuming $\bar{\mu}(s) \approx \tilde{\mu}(s)$ can be corrected by modifying the dynamics of the margin variable to be

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

We will call this the *explicit margin method*. $\mu + \beta^e\Delta\mu$ turns out to be an exact theoretical solution for the ESS of the geometric shock hierarchy when $\alpha = 0$ or $\alpha = 1$. It is a very good approximation when $0 < \alpha < 1$ and has a convenient discrete time implementation that is detailed later.

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., $\pi = 6/100$.

The closed form results derived above depend on the special assumption of a geometric shock hierarchy but the general conclusions do not. If one specified a more general hierarchy, the table of Poisson probabilities below shows that, if the cost of capital is on the order of 6 percent, then only the first four or five levels in the hierarchy will ever be significant.

¹³ We are not making a prima-facie case for using a geometric shock hierarchy. It is a simple place to start and, as we will see later, all four shortcut methods can be thought of as pragmatic approximations to this model.

<i>Poisson Probabilities</i> $p(n,s) = \exp[-\pi s](\pi s)^n/n!$						
$\pi = 6\%$	<i>s in Years</i>					
<i>n</i>	1	2	5	10	20	35
0	94.2%	88.7%	74.1%	54.9%	30.1%	12.2%
1	5.7%	10.6%	22.2%	32.9%	36.1%	25.7%
2	0.2%	0.6%	3.3%	9.9%	21.7%	27.0%
3	0.0%	0.0%	0.3%	2.0%	8.7%	18.9%
4	0.0%	0.0%	0.0%	0.3%	2.6%	9.9%
5	0.0%	0.0%	0.0%	0.0%	0.6%	4.2%
6+	0.0%	0.0%	0.0%	0.0%	0.2%	2.0%

The expected parameter value, under the regime-switching measure, will always start out equal to the base and then change over time as the upper levels in the hierarchy acquire more weight, i.e., a dynamic margin structure.

One reasonable property of this model is that the longer a 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 model 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.

In practice, there are two serious flaws with the Brute Force 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 will think that specifying a large hierarchy of assumptions is over-engineering. The shortcut methods to come will address both issues.

Two Actuarial Shortcuts and Their Duals

The methods discussed in this section start by simplifying the first-principles model down to a finite dimensional system. The primal forms of both methods have very convenient discrete time implementations, when they can be applied.

The author has had practical experience implementing both approaches.

The Implicit Margin Method—Primal Version

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

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

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

This approach turns out to have some extremely useful practical properties.

1. When it can be applied, it is usually a good approximation to the Brute Force model using a geometric shock hierarchy driven by the same parameter α . This will become more apparent when we look at the dual version of the method.
2. The discrete time version of the model is very efficient from both a computational perspective and a software development point of view, when it can be applied.

Here is a simple discrete time version of the mortality 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

$$({}_tV + g - e)(1 + i) = qF + (1 - q)_{t+1}V + \pi({}_t\hat{V} - {}_tV), \quad (a)$$

$$({}_t\hat{V} + g - e)(1 + i) = \hat{q}F + (1 - \hat{q})_{t+1}\hat{V} + \alpha\pi({}_t\hat{V} - {}_tV). \quad (b)$$

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 $(t + 1)$. Table 1 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 percent 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 percent coming from margin release plus 4 percent interest on capital for a combined total of 10 percent.

The main practical shortcoming of the primal 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.

The Prospective Method —Primal Version

The prospective method starts by calculating values V_0, V_1 using assumptions $\mu, \mu + \Delta\mu$ respectively. The margined values are then given by $V = V_0 + M$ 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 No. 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 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. The analysis of the dual models will explain why they are so close for this example.

The Implicit Method —Dual Version

Dual analysis is a standard topic in linear mathematics, but, for completeness, we go through a detailed development of the idea for the implicit method. We then state the corresponding results for the prospective method. The dual approach provides us a second way to calculate the models and also gives us additional insight that can be used to understand the theoretical errors in the practical shortcuts.

We return to the linear system of differential equations defining the implicit method

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

Let p, \hat{p} be two variables that are considered dual to V, \hat{V} and form the quantity

$$W(s) = p(s)V(s) + \hat{p}(s)\hat{V}(s).$$

Take the time derivative of W to find

$$\frac{d}{ds}W(s) = \dot{p}(s)V(s) + p\frac{d}{ds}V + \dot{\hat{p}}\hat{V}(s) + \hat{p}(s)\frac{d}{ds}\hat{V}(s).$$

Now use the defining equations for $\frac{d}{ds}V, \frac{d}{ds}\hat{V}$ and collect terms proportional to the primal variables.

$$\begin{aligned}\frac{d}{ds}W(s) &= [\dot{p} + (r + \mu + \pi)p + \alpha\pi\hat{p}]V + [\dot{\hat{p}} + (r + \mu + \Delta\mu - \alpha\pi)\hat{p} - \pi p]\hat{V} \\ &\quad - [\mu F + e - g]p - [(\mu + \Delta\mu)F + e - g]\hat{p}.\end{aligned}$$

We now choose to evolve the dual variables in such a way that the first two square brackets in the equation above are zero, i.e.,

$$\dot{p} + (r + \mu + \pi)p + \alpha\pi\hat{p} = 0,$$

$$\dot{\hat{p}} + (r + \mu + \Delta\mu - \alpha\pi)\hat{p} - \pi p = 0.$$

This is the dual system of linear equations.

The equation for W now simplifies to

$$\frac{d}{ds}W(s) = -[\mu F + e - g]p - [(\mu + \Delta\mu)F + e - g]\hat{p}.$$

Define new variables $p^T = p + \hat{p}$, and $\beta = \hat{p}/p^T$. Then $p = (1 - \beta)p^T$ and we can rewrite the equation for W as

$$\frac{d}{ds}W(s) = -p^T[(\mu + \beta\Delta\mu)F + e - g].$$

We will call the quantity β a margin variable because the quantity $(\mu + \beta\Delta\mu)$ is starting to look like a risk-loaded mortality assumption.

The quantity p^T is a risk-loaded discount factor that we can show by calculating

$$\begin{aligned}\frac{d}{ds}p^T &= \frac{d}{ds}(p + \hat{p}), \\ &= -p^T[r + \mu + \beta\Delta\mu],\end{aligned}$$

so that $p^T(s) = p^T(t)e^{-\int_t^s (r + \mu + \beta\Delta\mu)dv}$ for $s > t$.

Now choose initial conditions $p(t) = 1, \hat{p}(t) = 0$, so that $p^T(t) = 1$ and $W(t) = V(t)$ at the valuation date.

The differential equation for W can now be written as

$$\frac{d}{ds}W(s) = -e^{-\int_t^s (r + \mu + \beta\Delta\mu)dv}[(\mu + \beta\Delta\mu)F + e - g].$$

Integrate this equation from the valuation date $s = t$ to a maturity date $s = T$ where $V(T) = \hat{V}(T)$, i.e., $W(T) = p^T(T)V(T)$. We find

$$p^T(T)V(T) - W(t) = -\int_t^T e^{-\int_t^s (r + \mu + \beta\Delta\mu)dv}[(\mu + \beta\Delta\mu)F + e - g] ds,$$

or

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

We have shown that solving the primal system, at time t , is equivalent to using a dynamic margin assumption of the form $\mu + \beta\Delta\mu$. The dynamics of the margin variable are determined by the initial condition $\beta(t) = 0$ and the evolution equation, for $s > t$ of

$$\begin{aligned}\frac{d\beta}{ds} &= \frac{d}{ds}\hat{p}/p^T \\ &= \frac{\dot{\hat{p}}}{p^T} - \beta\frac{\dot{p}^T}{p^T}, \\ &= \beta(\beta - 1)\Delta\mu + \pi(1 - (1 - \alpha)\beta).\end{aligned}$$

An important point to note here is that the interest rate r has dropped out of the margin analysis. The risk-loaded mortality assumption is therefore independent of the economic scenario as long as $\Delta\mu$ doesn't depend on the economic scenario.

If we look at this result through the eyes of a financial engineer we might say that we have a risk-neutral mortality assumption of the form $\mu + \beta^i \Delta\mu$ where the implicit margin variable β^i is zero in the real world (P measure), so $\beta^i(t) = 0$, but for $s > t$ evolves according the dynamical law

$$\begin{aligned} d\beta^i &= \{ \beta^i(\beta^i - 1)\Delta\mu + \pi(1 - (1 - \alpha)\beta^i) \} ds, \\ &= \{ (\pi - \beta^i \Delta\mu)(1 - (1 - \alpha)\beta^i) + \alpha\Delta\mu(\beta^i)^2 \} ds, \end{aligned}$$

in the valuation measure.

The two ways of writing the dynamics given above allow us to compare the implicit method to the explicit method and simple mean introduced earlier. Relative to the explicit method whose dynamics are given by $d\beta^e = (\pi - \beta^e \Delta\mu)(1 - (1 - \alpha)\beta^e) ds$ it is clear that if $\Delta\mu > 0$ then $\beta^e \leq \beta^i$ depending on α . If $\Delta\mu < 0$ then the relationship is reversed and $\beta^e \geq \beta^i$.

Since the dynamics of the simple mean approximation are given by

$$d\bar{\beta} = \pi[1 - (1 - \alpha)\bar{\beta}] ds$$

we conclude that if $\Delta\mu > 0$ then $\beta^i \leq \bar{\beta}$ as long as we don't go too far into the future where we might have $\beta^i > 1$. Similarly, if $\Delta\mu < 0$, we see that $\beta^i \geq \bar{\beta}$ in the short run. We summarize below the relationships between the methods discussed for far.

$$\begin{aligned} \Delta\mu > 0 &\rightarrow \beta^e \leq \beta^i, \beta^e \leq \bar{\beta} \text{ and } \beta^i \leq \bar{\beta} \text{ in the short run,} \\ \Delta\mu < 0 &\rightarrow \beta^e \geq \beta^i, \beta^e \geq \bar{\beta} \text{ and } \beta^i \geq \bar{\beta} \text{ in the short run.} \end{aligned}$$

Since the implicit margin method usually produces results that lie in between the explicit and simple mean results, it makes sense to think of the implicit method as a pragmatic approximation to the first-principles geometric shock hierarchy approach.

To calculate a shocked value \hat{V} in the dual approach we simply choose to solve the dual equations with different initial conditions. If $p(t) = 0$ and $\hat{p}(t) = 1$ then $W(t) = \hat{V}(t)$.

Write the new margin variable in the form $1 + \alpha\hat{\beta} = \hat{p}/(p + \hat{p})$, $\hat{\beta}(t) = 0$ then we can represent the shocked value as

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

The dynamics of the shocked beta can be derived by noting that $1 + \alpha\hat{\beta}$ must satisfy the same evolution equation as β , i.e.,

$$\frac{d}{ds} (1 + \alpha\hat{\beta}) = (1 + \alpha\hat{\beta})(1 + \alpha\hat{\beta} - 1)\Delta\mu + \pi \left[1 - (1 - \alpha) \left(1 - (1 + \alpha\hat{\beta}) \right) \right].$$

If $\alpha > 0$ this simplifies down to

$$\frac{d}{ds}\hat{\beta} = (1 + \alpha\hat{\beta})\hat{\beta}\Delta\mu + \pi[1 - (1 - \alpha)\hat{\beta}].$$

Using this equation, it is possible to show that $\Delta\mu > 0 \rightarrow \hat{\beta}(s) > \beta(s)$ for $s > t$ with the inequality reversing if $\Delta\mu < 0$.

There is a second way to estimate economic capital when using the dual model. In the valuation measure, we can think of the reserve value as a function $V = V(s, \beta)$, that satisfies the partial differential equation¹⁴

$$\frac{\partial V}{\partial s} + \{\beta(\beta - 1)\Delta\mu + \pi[1 - (1 - \alpha)\beta]\}\frac{\partial V}{\partial \beta} + (\mu + \beta\Delta\mu)(F - V) = rV + g - e.$$

What this equation says is that the total expected rate of change, in the valuation measure, is equal to the risk-free rate plus the impact of premiums and expenses. At the valuation date, when $s = t$ and $\beta = 0$, this simplifies down to

$$\frac{\partial V}{\partial s} + \pi \frac{\partial V}{\partial \beta} \Big|_{s=t} + (\mu_0 + \pi\Delta Q)(F - V) = rV + g - e.$$

With respect to the real-world measure, the expected rate of change is

$$\frac{\partial V}{\partial s} + \mu_0(F - V) = rV + g - e - \pi\Delta Q(F - V) - \pi \frac{\partial V}{\partial \beta} \Big|_{s=t}.$$

Since this model is equivalent to the primal system it must be true that the “greek” $\frac{\partial V}{\partial \beta}$ associated with the margin variable β is the economic capital for parameter risk, i.e., at the valuation¹⁵ date,

$$\frac{\partial V}{\partial \beta} \Big|_{s=t} = \hat{V}(t) - V(t).$$

This result can be generalized to any financial engineering approach that uses a risk-neutral mortality assumption of the form $\mu + \beta\Delta\mu$ and the margin variable evolves according to a rule of the form $d\beta = B(s, \beta)ds$. As long as $B(s, 0) = \pi$ we can say that margin for parameter risk is being released in a way that is consistent with holding an implied economic capital amount equal to the greek $\frac{\partial V}{\partial \beta}$ calculated on the valuation date. We will use this theoretical result several times in what follows.

¹⁴ This is fairly intuitive, if you think like a financial engineer. If you don't, then you can derive the equation by repeating the $E[\Delta L] = E[\Delta A]$ argument developed earlier in this paper. In this more general setting, $\Delta L = (F - V)$ if the life dies, and $\Delta L = [V(t + \Delta t, \beta + \Delta\beta) - V(t, \beta)] \approx [\frac{\partial V}{\partial t} + \frac{\partial V}{\partial \beta} \frac{\Delta\beta}{\Delta t}] \Delta t$ if the life survives. The probability of death in the time interval is $(\mu + \beta\Delta\mu)\Delta t$.

¹⁵ A quick way to confirm this result with algebra is to consider the solution to the dual equations with initial conditions $p(t) = 1 - \varepsilon$ and $\hat{p}(t) = \varepsilon$, which implies $W(t) = V(t) + \varepsilon[\hat{V}(t) - V(t)]$ and $\beta(t) = \varepsilon$.

One practical application of this result is that we now have two different ways to estimate capital. Estimating greeks is a well-developed topic in financial engineering,¹⁶ and many of that discipline's tools could be applied to this problem.

The Prospective Method—Dual Version

In this section we apply the same analytical process to the prospective method that we used for the implicit model. We present less detail here because the process is basically the same.

For this method there are three primal variables, V_1, V_0, M , which satisfy the following system of differential equations:

$$\begin{aligned}\frac{dV_0}{dt} + \mu(F - V_0) &= (r + \vartheta)V_0 + g - e \\ \frac{dV_1}{dt} + (\mu + \Delta\mu)(F - V_1) &= (r + \vartheta)V_1 + g - e \\ \frac{dM}{dt} + \mu(0 - M) &= rM - \pi[V_1 - V_0 - (1 - \alpha)M].\end{aligned}$$

In writing down this system we are including a deterministic illiquidity spread $\vartheta = \vartheta(s) \geq 0$ that Solvency II currently allows in the calculation of V_1, V_0 , but we have not set $\alpha = 1$.

As before, we introduce dual variables, p_0, p_1, m , and consider a linear combination

$$W = p_0V_0 + p_1V_1 + mM.$$

If the dual variables satisfy the appropriate dual system of equations

$$\begin{aligned}p_0 + p_0(r + \vartheta + \mu) + \pi m &= 0, \\ p_1 + p_1(r + \vartheta + \mu + \Delta\mu) - \pi m &= 0, \\ \dot{m} + m(r + \mu) + \pi(1 - \alpha)m &= 0,\end{aligned}$$

Then

$$\dot{W} = -[p_0(\mu F + e - g) + p_1((\mu + \Delta\mu)F + e - g)].$$

Again we introduce new variables $p^T = p_0 + p_1, \beta = p_1/p^T, \omega = m/p^T$. Then the risk-loaded value $V = V_0 + M$ can be calculated using the dynamic margin assumption $\mu + \beta\Delta\mu$ where the dynamics of the margin variables are defined by the system

$$\begin{aligned}\dot{p}^T &= -p^T[r + \vartheta + \mu + \beta\Delta\mu], \quad p^T(t) = 1 \\ \frac{d\beta}{ds} &= \beta(\beta - 1)\Delta\mu + \pi\omega, \quad \beta(t) = 0 \\ \frac{d\omega}{ds} &= \omega(\vartheta + \beta\Delta\mu - \pi(1 - \alpha)), \quad \omega(t) = 1, \\ \frac{d}{ds}W(s) &= -p^T[(\mu + \beta\Delta\mu)F + e - g].\end{aligned}$$

¹⁶ See, for example, Chapter 7 of Glasserman, P. "Monte Carlo Methods in Financial Engineering." Springer (2005).

The first of these four equations shows that the risk-loaded cash flows should be discounted using the risk-free rate plus the illiquidity spread, if it is being used, i.e.,

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

The second and third equations in the margin system constitute a closed system of evolution equations for the pair (β, ω) , which does not depend on the interest rate r unless $\Delta\mu$ does. The risk loads do depend on the illiquidity spread ϑ , which is usually assumed to be independent of the economic scenario. Again the mortality margins are independent of the economic scenario.

The financial engineering partial differential equation for $V = V(s, \beta, \omega)$ associated with this model is

$$\begin{aligned} \frac{\partial V}{\partial s} + \{ \beta(\beta - 1)\Delta\mu + \pi\omega \} \frac{\partial V}{\partial \beta} + [\omega(\vartheta + \beta\Delta\mu) - \pi(1 - \alpha)\omega] \frac{\partial V}{\partial \omega} + (\mu + \beta\Delta\mu)(F - V) \\ = (r + \vartheta)V + g - e. \end{aligned}$$

On the valuation date, when $\beta = 0$ and $\omega = 1$, we see that the margin release rate for parameter risk appears to be

$$\pi \left[\frac{\partial V}{\partial \beta} - (1 - \alpha) \frac{\partial V}{\partial \omega} \right] \Big|_{\beta=0, \omega=1} + \vartheta \frac{\partial V}{\partial \omega} \Big|_{\beta=0, \omega=1}.$$

The quantity above in square brackets looks like it should be the economic capital $V_1 - V_0 - (1 - \alpha)M$. This strongly suggests, and more detailed analysis can prove,¹⁷ that on the valuation date

$$\begin{aligned} \frac{\partial V}{\partial \beta} \Big|_{\beta=0, \omega=1} &= V_1(t) - V_0(t), \\ \frac{\partial V}{\partial \omega} \Big|_{\beta=0, \omega=1} &= M(t). \end{aligned}$$

A financial engineering approach based on the dual model, that was used to compute the three quantities $V, \frac{\partial V}{\partial \beta}, \frac{\partial V}{\partial \omega}$, could then be used to recover the three primal variables, V_1, V_0, M , by using

$$\begin{aligned} M(t) &= \frac{\partial V}{\partial \omega}(t), \\ V_0(t) &= V(t) - M(t), \\ V_1(t) &= V_0(t) + \frac{\partial V}{\partial \beta}(t). \end{aligned}$$

¹⁷ It is possible to show that $V = V_0 + \omega \frac{\partial V}{\partial \omega} + \beta \frac{\partial V}{\partial \beta}$ for all $s \geq t$.

The second term of the margin release rate $\vartheta \frac{\partial V}{\partial \omega} \big|_{\beta=0, \omega=1}$ can be thought of as an additional margin release needed to make up for the fact that the prospective method assumes the assets backing V_0 earn $(r + \vartheta)$ but the assets backing the present value of margins $M = \frac{\partial V}{\partial \omega} \big|_{\beta=0, \omega=1}$ only earn the risk-free rate r .

In order to compare the prospective method to the other methods introduced here we rewrite the dynamics for the prospective margin variables as

$$\begin{aligned} \frac{d\beta^p}{ds} &= \beta^p(\beta^p - 1)\Delta\mu + \pi\omega, \\ &= \beta^p(\beta^p - 1)\Delta\mu + \pi(1 - (1 - \alpha)\beta^p) + \pi[\omega - (1 - (1 - \alpha)\beta^p)], \\ &= (\pi - \beta^p\Delta\mu)(1 - (1 - \alpha)\beta^p) + \alpha\beta^{p^2}\Delta\mu + \pi[\omega - (1 - (1 - \alpha)\beta^p)], \\ \frac{d\omega}{ds} &= \omega(\vartheta + \beta^p\Delta\mu - \pi(1 - \alpha)). \end{aligned}$$

We can show that if $\Delta\mu > 0$ then $\omega - (1 - (1 - \alpha)\beta^p) > 0$. This is done by explicitly calculating

$$\frac{d}{ds} [\omega - (1 - (1 - \alpha)\beta^p)] = (\omega\vartheta + \alpha(\beta^p)^2\Delta\mu) + \beta^p\Delta\mu[\omega - (1 - (1 - \alpha)\beta^p)].$$

Solving this differential equation, and using $\beta^p(t) = 0, \omega(t) = 1$, we see that for $s > t$

$$[\omega - (1 - (1 - \alpha)\beta^p)](s) = \int_t^s e^{\int_v^s \beta^p \Delta\mu du} [\omega\vartheta + \alpha(\beta^p)^2\Delta\mu] dv.$$

Since $\vartheta \geq 0$ this quantity will be positive if $\Delta\mu > 0$. If $\Delta\mu < 0$, we can't draw a definitive conclusion, unless $\vartheta = 0$, in which case the inequality reverses.

Based on these results we can definitely conclude that $\Delta\mu > 0 \rightarrow \beta^e \leq \beta^i \leq \beta^p$, and if $\vartheta = 0$ we can also conclude that $\Delta\mu < 0 \rightarrow \beta^e \geq \beta^i \geq \beta^p$. The equations above don't suggest a definitive relationship between β^p and $\bar{\beta}$.

We have done enough to justify the conclusion that the prospective method can, like the implicit method, be thought of as a pragmatic approximation to the first-principles model with geometric shock hierarchy. In the author's opinion, this puts both actuarial shortcuts on a solid theoretical footing. The results will be reasonable as long as the key inputs $\Delta\mu$ and α are chosen appropriately.

Discrete Time Implementation of the Dual Methods

The dual methods are easier to implement than the theoretical developments above suggest. It is possible to solve the dynamical system for the margin variables numerically and then compute $\mu + \beta\Delta\mu$ explicitly. But it is easier to use the discrete time version of the primal model to value a sequence of simple pure endowments assuming zero interest. The result is a sequence of margined persistency factors $p^T(s) = \exp[-\int_t^s (\mu + \beta\Delta\mu)dv]$ that

can easily be converted to a set of discrete time margined mortality rates. The dual theory given here justifies the use of this relatively simple procedure.

Table 3 shows the results of applying this idea to both the implicit and prospective models with $\alpha = 1, \pi = 0.06$ and $\vartheta = 0$ for the prospective model. These results are based on the same discretization scheme that was used to illustrate the primal versions of the models.

Table 3: Dual Approach to Margined Mortality

	1000 qx					
	Inputs		Implicit Model Margined Mortality		Prospective Model Margined Mortality	
	Base μ	Shocked $\mu+\Delta\mu$	Base $\mu+\beta\Delta\mu$	Shocked $\mu+\Delta\mu+\alpha\beta\Delta\mu$	Base $\mu+\beta\Delta\mu$	Shocked $\mu+\Delta\mu+\alpha\beta\Delta\mu$
1	1.01499	1.11649	1.02108	1.12258	1.02108	1.12258
2	1.10634	1.21698	1.11962	1.23025	1.11962	1.23025
3	1.20784	1.32862	1.22958	1.35037	1.22958	1.35037
4	1.31949	1.45144	1.35115	1.48311	1.35115	1.48312
5	1.44128	1.58541	1.48451	1.62866	1.48452	1.62867
6	1.57323	1.73055	1.62985	1.78721	1.62986	1.78723
7	1.71533	1.88686	1.78735	1.95894	1.78735	1.95897
8	1.86757	2.05433	1.95719	2.14404	1.95719	2.14408
9	2.04012	2.24413	2.15024	2.35439	2.15025	2.35446
10	2.22281	2.44509	2.35612	2.57860	2.35614	2.57870

We see the prospective model is slightly conservative relative to the implicit model for this example, but we have had to use five decimal places to see the difference. This is what the theory suggested should happen. If the decrement shock went down rather than up, this relationship would reverse.

Two Financial Engineering Shortcuts

The models outlined in the previous section were motivated by shortcuts designed to make the primal problem look simple. The reader may have noticed that the dual version did not necessarily look “simple.” In this section we focus on simplifying assumptions designed to make the dual form of the model look simple. We could try to reverse the process and work back to the corresponding primal problem, but we will not do that explicitly. Instead, we ask if this simple approach to dynamic margins leads to a “reasonable” implied capital requirement.

The first model we look at, called the simple mean, starts by looking at the dynamical rule $d\bar{\beta} = \pi[1 - (1 - \alpha)\bar{\beta}]ds$ that arose as an approximation to the first-principles geometric shock hierarchy model. We then develop a tool that allows us to understand, and essentially correct, the theoretical error in this model. The corrected model is called the explicit margin model and represents, in the author’s opinion, the best compromise between theoretical rigor and practical issues of all the methods presented here.

The Simple Mean Margin Method

In the first principles section of this paper we derived the approximation $\tilde{\mu}_{t+s} \approx \mu + \bar{\beta}\Delta\mu$ where

$d\bar{\beta} = \pi[1 - (1 - \alpha)\bar{\beta}]ds$. In this section we forget where this came from and simply ask what the implied economic capital is for this shortcut. To do this we write down the financial engineering version of this model as

$$\frac{\partial V}{\partial s} + \pi[1 - (1 - \alpha)\bar{\beta}] \frac{\partial V}{\partial \bar{\beta}} + (\mu + \bar{\beta}\Delta\mu)(F - V) = rV + g - e.$$

As we saw earlier, the margin release rate at the valuation date is $\pi \frac{\partial V}{\partial \bar{\beta}}$, so $\frac{\partial V}{\partial \bar{\beta}}$ is the implied economic capital. We can calculate this quantity by differentiating the equation above with respect to $\bar{\beta}$ and deriving a valuation equation for $\delta = \frac{\partial V}{\partial \bar{\beta}}$. Carrying out the calculus we find

$$\frac{\partial \delta}{\partial s} + \pi[1 - (1 - \alpha)\bar{\beta}] \frac{\partial \delta}{\partial \bar{\beta}} - \pi(1 - \alpha)\delta + (\mu + \bar{\beta}\Delta\mu)(0 - \delta) + \Delta\mu(F - V) = r\delta.$$

On rearranging this becomes

$$\frac{\partial \delta}{\partial s} + \pi[1 - (1 - \alpha)\bar{\beta}] \frac{\partial \delta}{\partial \bar{\beta}} = (r + \mu + \bar{\beta}\Delta\mu + \pi(1 - \alpha))\delta - \Delta\mu(F - V).$$

This equation tells us that the total rate of change of δ , in the valuation measure, is to grow at the rate $r + \mu + \bar{\beta}\Delta\mu + \pi(1 - \alpha)$ while releasing an experience gain amount of $\Delta\mu(F - V)$. We can therefore write the solution at time t as

$$\delta(t) = \int_t^\infty e^{-\int_t^s (r + \mu + \bar{\beta}\Delta\mu + \pi(1 - \alpha)) dv} \Delta\mu(s)(F - V(s)) ds.$$

The expression makes more sense if we rewrite it using $e^{-\int_t^s \pi(1 - \alpha) dv} = [1 + \alpha\bar{\beta}(s) - \bar{\beta}(s)]$. We then have

$$\delta(t) = \int_t^\infty e^{-\int_t^s (r + \mu + \bar{\beta}\Delta\mu) dv} [1 + \alpha\bar{\beta} - \bar{\beta}] \Delta\mu(s)(F - V(s)) ds.$$

In this form, the expression for the implied economic capital is almost reasonable. If we value on the assumption that $\mu + \bar{\beta}\Delta\mu$ is correct, but mortality experience actually turns out to be $\mu + \Delta\mu + \alpha\bar{\beta}\Delta\mu$, the experience loss rate at any future time is $[1 + \alpha\bar{\beta} - \bar{\beta}]\Delta\mu(s)(F - V(s))$. Economic capital should be the present value of these losses, discounted using the capital assumption, i.e.,

$$EC = \int_t^\infty e^{-\int_t^s (r + \mu + \Delta\mu + \alpha\bar{\beta}\Delta\mu) dv} [1 + \alpha\bar{\beta} - \bar{\beta}] \Delta\mu(s)(F - V(s)) ds.$$

The difference between these two capital expressions is the mortality rate used for discounting future losses. The simple mean model gets the correct experience gain term but uses an incorrect discounting method. The discounting error is conservative if $\Delta\mu > 0$ and is immaterial if $\Delta\mu$ is small and the contract is relatively short.

The Explicit Margin Method

The explicit margin method starts by asking whether we can modify the margin variable dynamics so as to correct the theoretical error in the simple mean model. Assume the margin dynamics are given by $d\beta = B(s, \beta)ds$ for some unknown function B such that $B(s, 0) = \pi$.

The financial engineering equation is then

$$\frac{\partial V}{\partial s} + B(s, \beta) \frac{\partial V}{\partial \beta} + (\mu + \beta \Delta\mu)(F - V) = rV + g - e,$$

and the corresponding equation for $\delta = \frac{\partial V}{\partial \beta}$ is given by

$$\frac{\partial \delta}{\partial s} + B \frac{\partial \delta}{\partial \beta} = \left(r + \mu + \beta \Delta\mu + \frac{\partial B}{\partial \beta} \right) \delta - \Delta\mu(F - V).$$

An expression for the implied economic capital on the valuation date is then

$$\delta(t) = \int_t^\infty e^{-\int_t^s (r + \mu + \beta \Delta\mu + \frac{\partial B}{\partial \beta}) dv} \Delta\mu(s)(F - V(s)) ds.$$

The simplifying assumption we now make is that the shocked world risk-loaded scenario is given by $\mu + \Delta\mu + \alpha\beta\Delta\mu$ i.e. $\hat{\beta} = \beta$. This assumption can be motivated by its own intuition. A second motivation for this assumption is the observation that it is actually true for the geometric shock hierarchy if $\alpha = 1$ or $\alpha = 0$.¹⁸ If $0 < \alpha < 1$, this is an approximation to the geometric shock hierarchy, i.e., a shortcut.

Given this assumption, we want the economic capital to be the difference between valuing on the shocked and the base fair value scenarios, i.e.,

$$EC = \int_t^\infty e^{-\int_t^s (r + \mu + \Delta\mu + \alpha\beta\Delta\mu) dv} [1 + \alpha\beta - \beta] \Delta\mu(s)(F - V(s)) ds.$$

This will be the case, provided we can engineer the function $B(s, \beta)$ to satisfy, for all $s > t$

$$e^{-\int_t^s (r + \mu + \beta \Delta\mu + \frac{\partial B}{\partial \beta}) dv} = e^{-\int_t^s (r + \mu + \Delta\mu + \alpha\beta\Delta\mu) dv} [1 + \alpha\beta(s) - \beta(s)].$$

This is a solvable calculus problem.¹⁹ The answer is $B(s, \beta) = [\pi - \beta\Delta\mu(s)][1 - (1 - \alpha)\beta]$.

The dynamics for the explicit margin model are then given by

$$d\beta = [\pi - \beta\Delta\mu(s)][1 - (1 - \alpha)\beta] ds.$$

¹⁸ This is fairly clear if $\alpha = 0$, and if $\alpha = 1$ it follows from the translational symmetry of the assumed shock hierarchy.

¹⁹ The basic idea is to differentiate the equation with respect to s and derive a simple first order partial differential equation for B . Together with the boundary condition $B(s, 0) = \pi$, this determines the unknown function.

The qualitative behavior of the explicit margin variable is very similar to that of the simple mean model especially near the valuation date when the margin variable is small. Over longer time frames there can be a difference as the first factor above slows down the growth if the decrement shock is positive and speeds it up if negative.

If $\Delta\mu < 0$, the quantity $1/(1 - \alpha)$ is still the upper bound. If $\Delta\mu > 0$ the upper bound could be lower if $\pi/\Delta\mu < 1/(1 - \alpha)$.

As before, the danger zone is a situation where $\Delta\mu < 0$ and $\alpha = 1$ because there is no upper bound on β in this case. The loaded decrement assumption $\mu + \beta\Delta\mu$ could become negative.

As mentioned earlier, if $\alpha = 0$ or 1 , the explicit margin model is equivalent to the first-principles approach using the geometric shock hierarchy. One way to confirm this statement, and also get a sense of the remaining theoretical error, when $0 < \alpha < 1$, is to calculate the economic capital that the model implies we should hold in a shocked world. To do this we write down the financial engineering equation for the shocked value \hat{V} and then compute $\hat{\delta} = \frac{\partial \hat{V}}{\partial \beta}$ at the valuation date.

The first step is

$$\frac{\partial \hat{V}}{\partial s} + [\pi - \beta\Delta\mu(s)][1 - (1 - \alpha)\beta] \frac{\partial \hat{V}}{\partial \beta} + (\mu + \Delta\mu + \alpha\beta\Delta\mu)(F - \hat{V}) = r\hat{V} + g - e.$$

Now differentiate with respect to β to find

$$\frac{\partial \hat{\delta}}{\partial s} + [\pi - \beta\Delta\mu(s)][1 - (1 - \alpha)\beta] \frac{\partial \hat{\delta}}{\partial \beta} = (r + \mu + \Delta\mu(2 - 2\beta + 3\alpha\beta) + \pi(1 - \alpha))\hat{\delta} - \alpha\Delta\mu(F - \hat{V}).$$

At the valuation date we then have

$$\hat{\delta}(t) = \int_t^\infty e^{-\int_t^s (r + \mu + \Delta\mu(2 - 2\beta + 3\alpha\beta) + \pi(1 - \alpha)) dv} \alpha\Delta\mu(s)(F - \hat{V}(s)) ds.$$

From the evolution equation $d\beta = [\pi - \beta\Delta\mu(s)][1 - (1 - \alpha)\beta] ds$ we can write

$$\frac{d\beta}{[1 - (1 - \alpha)\beta]} = [\pi - \beta\Delta\mu(s)] ds.$$

Integrating this equation from t to s we get the useful identity

$$1 - (1 - \alpha)\beta(s) = e^{-(1 - \alpha) \int_t^s [\pi - \beta\Delta\mu] dv}, \quad (*)$$

which can be used in the expression above for $\hat{\delta}(t)$ to rewrite it as

$$\hat{\delta}(t) = \int_t^\infty e^{-\int_t^s (r + \mu + \Delta\mu(2 - \beta + 2\alpha\beta)) dv} [1 - (1 - \alpha)\beta] \alpha\Delta\mu(s)(F - \hat{V}(s)) ds.$$

If there were no theoretical error, the expression we would expect to see here is the difference between valuing on the double shocked assumption $\mu + \Delta\mu + \alpha\Delta\mu + \alpha^2\beta\Delta\mu$ versus the single shock $\mu + \Delta\mu + \alpha\beta\Delta\mu$, i.e.,

$$EC = \int_t^\infty e^{-\int_t^s (r+\mu+\Delta\mu+\alpha\Delta\mu+\alpha^2\beta\Delta\mu)dv} [1 - (1 - \alpha)\beta]\alpha\Delta\mu(s)(F - \hat{V}(s))ds.$$

We have the right gain/loss term $[1 - (1 - \alpha)\beta]\alpha\Delta\mu(s)(F - \hat{V}(s))$, but there appears to be an error in the discounting again. To see the discounting error more clearly we write the expression for $\hat{\delta}(t)$ as

$$\hat{\delta}(t) = \int_t^\infty e^{-\int_t^s (r+\mu+\Delta\mu+\alpha\Delta\mu+\alpha^2\beta\Delta\mu+(1-\alpha)\{1-(1-\alpha)\beta\}\Delta\mu)dv} [1 - (1 - \alpha)\beta]\alpha\Delta\mu(s)(F - \hat{V}(s))ds.$$

From this expression it is clear that if $\alpha = 1$ there is no error in the discounting so we get the right answer. If $\alpha = 0$, there is still an error in the discounting, but it doesn't matter because the gain/loss term is zero and we still get the right answer, i.e., $\hat{\delta}(t) = 0$.

If $0 < \alpha < 1$ the sign of the discounting error $(1 - \alpha)\{1 - (1 - \alpha)\beta\}\Delta\mu$ depends on the sign of $\Delta\mu$ since $\beta < 1/(1 - \alpha)$. We conclude that if $\Delta\mu > 0$ the explicit margin model is slightly liberal relative to the exact geometric shock hierarchy, and it is slightly conservative if $\Delta\mu < 0$.

In the author's practical experience one usually starts the assumption-setting process by putting $\alpha = 1$ and then modifying that position if warranted by the results. In practice, this usually means that we only use $\alpha < 1$ when $\Delta\mu < 0$. The end result is that using the explicit margin method is either exactly equivalent to the geometric shock hierarchy or a conservative approximation to it.

Discrete Time Implementation of the Financial Engineering Shortcuts

Both of the financial engineering shortcuts 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 $x + t$; 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-loading process begins on the valuation date.

Define the decrement shock $\Delta\mu_s$ by

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

For the simple mean approximation, we can go from one margined 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 loaded mortality rate is then given by

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

This is obviously easy to implement. A corresponding margined mortality rate for the shocked scenario is given by

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

Table 4 shows a spreadsheet implementation of the above logic.

Table 4: Simple Mean Approximation to Margined Mortality

	1000 qx		α = 100%	π = 5.83%	
	Inputs			Loaded Results	
s	Base	Shocked	ks	Base	Shocked
	q ^{[x]+t+s}	q ^{^[x]+t+s}		q ^{{[x]+t}+s}	q ^{^{[x]+t}+s}
0	1.01499	1.11649	2.91%	1.01795	1.11945
1	1.10634	1.21698	8.74%	1.11601	1.22664
2	1.20784	1.32862	14.57%	1.22543	1.34622
3	1.31949	1.45144	20.39%	1.34640	1.47834
4	1.44128	1.58541	26.22%	1.47908	1.62320
5	1.57323	1.73055	32.05%	1.62365	1.78097
6	1.71533	1.88686	37.87%	1.78030	1.95182
7	1.86757	2.05433	43.70%	1.94919	2.13594
8	2.04012	2.24413	49.53%	2.14117	2.34516
9	2.22281	2.44509	55.36%	2.34586	2.56812

Since this is a continuous time model we have used a continuously compounded cost of capital rate $\pi = \ln(1.06) = 0.0583$ to make the results comparable to those given in Table 3. Even

with that adjustment, the results of the two actuarial shortcuts reported in Table 3 are both more conservative than what we see here. This is due to the discretization approach taken in Table 3, which has added an element of conservatism that diminishes over time.

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

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

In the previous section we derived the identity $1 - (1 - \alpha)\beta(s) = e^{-(1-\alpha)\int_t^s [\pi - \beta\Delta\mu] dv}$. Using this expression we can write, assuming $0 \leq \alpha < 1$

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

One way to move forward 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

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

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

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

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

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

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

Table 5 shows a spreadsheet implementation of the logic summarized above for the explicit margin method. It uses the same inputs that were used in Table 4.

Table 5: Explicit Margin Method Example

	1000 qx		$\alpha = 100\%$	$\pi = 5.83\%$	Loaded Results	
	Inputs				Base	Shocked
	Base	Shocked				
s	$q[x]+t+s$	$q^{\wedge}[x]+t+s$	$\pi(1-\alpha)-\Delta\mu_s$	J_s	$q\{[x]+t\}+s$	$q^{\wedge}\{[x]+t\}+s$
0	1.01499	1.11649	-0.01%	0.00%	1.01795	1.11945
1	1.10634	1.21698	-0.01%	5.83%	1.11601	1.22664
2	1.20784	1.32862	-0.01%	11.65%	1.22543	1.34621
3	1.31949	1.45144	-0.01%	17.48%	1.34639	1.47834
4	1.44128	1.58541	-0.01%	23.30%	1.47907	1.62319
5	1.57323	1.73055	-0.02%	29.12%	1.62363	1.78095
6	1.71533	1.88686	-0.02%	34.95%	1.78027	1.95179
7	1.86757	2.05433	-0.02%	40.77%	1.94915	2.13589
8	2.04012	2.24413	-0.02%	46.59%	2.14110	2.34509
9	2.22281	2.44509	-0.02%	52.40%	2.34576	2.56802

The main point to be emphasized here is that the actual implementation is not onerous. The results in Tables 4 and 5 are fairly close, and, since $\Delta\mu > 0$ we see that the simple mean method is a bit conservative relative to the explicit margin method as theory suggests.

When comparing Table 5 to Table 3 we do not see the relationships suggested by theory. This is due to the different approach taken to discretization and the annual time step used. The good news here is that all four shortcuts are so close that a practical issue, like the choice of discretization method, can swamp the theoretical differences.

Summary and Comparison of the Four Shortcuts

The previous sections of this paper have introduced a number of margin calculation methods and derived some of their theoretical properties. The numerical examples given so far suggest that, in practice, there is not much to choose between them. In this section we provide a compact summary of the theoretical results obtained and provide some additional numerical examples to point out where there could be material differences between the various approaches.

The table below summarizes the five approaches introduced in this paper.

Method	Simplifying Assumption	Margin Dynamics	Comment/Greeks
First principles	Geometric shock hierarchy	No closed form for β^{fp} unless $\alpha = 0$ or 1 when $\beta^{fp} = \beta^e$	Needs Monte Carlo simulation
Implicit	$\hat{V}^{(2)} - \hat{V} \approx \alpha(\hat{V} - V)$	$\frac{d\beta^i}{ds} = \beta^i(\beta^i - 1)\Delta\mu + \pi(1 - (1 - \alpha)\beta^i)$	$\frac{\partial V}{\partial \beta} _{s=t} = \hat{V}(t) - V(t)$

Prospective (Solvency II)	$\hat{M} \approx \alpha M$	$\frac{d\beta^p}{ds} = \beta^p(\beta^p - 1)\Delta\mu + \pi\omega$ $\frac{d\omega}{ds} = \omega(\vartheta + \beta^p\Delta\mu - \pi(1 - \alpha))$	$\frac{\partial V}{\partial \beta} _{s=t} = V_1(t) - V_0(t),$ $\frac{\partial V}{\partial \omega} _{s=t} = M(t).$
Simple mean	ESS \approx Mean	$d\bar{\beta}/ds = \pi[1 - (1 - \alpha)\bar{\beta}]$	$\frac{\partial V}{\partial \beta} _{s=t} = EC \approx \hat{V}(t) - V(t)$
Explicit	$\mu + \beta\Delta\mu$ $\rightarrow \mu + \Delta\mu + \alpha\beta\Delta\mu$	$d\beta^e/ds = (\pi - \beta^e\Delta\mu)$ $\times (1 - (1 - \alpha)\beta^e)$	$\frac{\partial V}{\partial \beta} _{s=t} = \hat{V}(t) - V(t)$

By analyzing the margin dynamics we were able to derive the following relative relationships between the various approaches

$$\Delta\mu > 0 \rightarrow \beta^e \leq \beta^i \leq \beta^p, \beta^e \leq \bar{\beta} \text{ and } \beta^i \leq \bar{\beta} \text{ in the short run,}$$

$$\Delta\mu < 0 \rightarrow \beta^e \geq \beta^i \geq \beta^p, \beta^e \geq \bar{\beta} \text{ and } \beta^i \geq \bar{\beta} \text{ in the short run.}$$

We now present a numerical example that is deliberately chosen to exaggerate the differences between the four shortcuts. The product is a simple pure endowment that pays \$1,000 to a life that survives for n years. The interest rate assumed is zero. We have very little information about the life but we are told a contagion-loaded, best-estimate decrement rate of $q = 0.01$ for all years is reasonable. Given the lack of good information we also assume a fairly large parameter shock $\Delta q = -0.005$. The cost of capital rate is 6.0 percent compounded annually.

If we set $\alpha = 1$, the underlying geometric shock hierarchy is the very unreasonable

$$0.01 \rightarrow 0.005 \rightarrow 0.000 \rightarrow -0.005 \rightarrow -0.01 \rightarrow \dots$$

Since this is clearly an inappropriate choice of input assumptions we should not be surprised if we see inappropriate results (garbage in, garbage out).

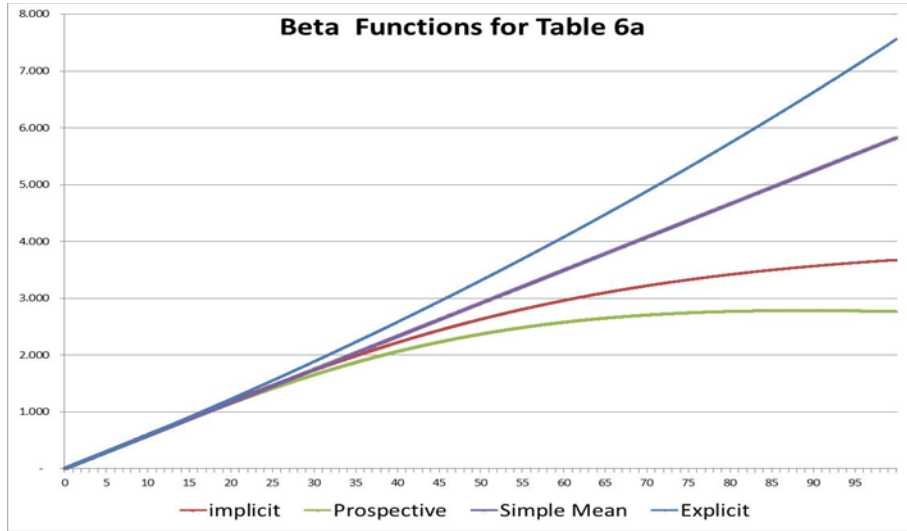
Table 6a: Pure Endowment Values									
		$\alpha = 100\%$		$q = 1.00\%$		$\Delta q = -0.50\%$			
				$i = 0.0\%$		$\pi = 6.0\%$			
Maturity		Implicit		Prospective		Simple Mean		Explicit	
n	V_0	V_M	$V_M + EC$	V_M	$V_M + EC$	V_M	$V_M + EC$	V_M	$V_M + EC$
1	990	990	995	990	995	990	995	990	995
5	951	955	980	955	980	955	979	955	979
10	904	920	967	920	966	918	964	918	965
25	778	859	967	858	962	855	963	856	971
50	605	876	1,072	859	1,033	883	1,105	902	1,161
75	471	1,005	1,287	933	1,149	1,101	1,518	1,205	1,758
100	366	1,226	1,606	1,032	1,272	1,659	2,497	2,104	3,481

The first column in the table is a sequence of contagion-loaded, best-estimate values

$$V_0(n) = 1,000 (1 - q)^n.$$

This a reasonable sequence of *decreasing* values as the maturity gets longer.

For each margin method we then show the fair value V_M and the total balance sheet requirement $\hat{V}_M = V_M + EC$. For each method, the endowment values start *increasing* somewhere between years 25 and 50. We can explain these results by looking at the beta functions, out to 100 years, for each method in the chart below.



Once the beta function exceeds 2.0 the resulting margined decrement rate $q + \beta\Delta q$ becomes negative. This happens somewhere between years 30 and 40 depending on the details of the method. The good news is that the beta functions are very similar up to that point.

The next example is designed to correct the obvious flaw in the above by setting $\alpha = 1/2$. The decrement shock hierarchy is now bounded below by zero, i.e.,

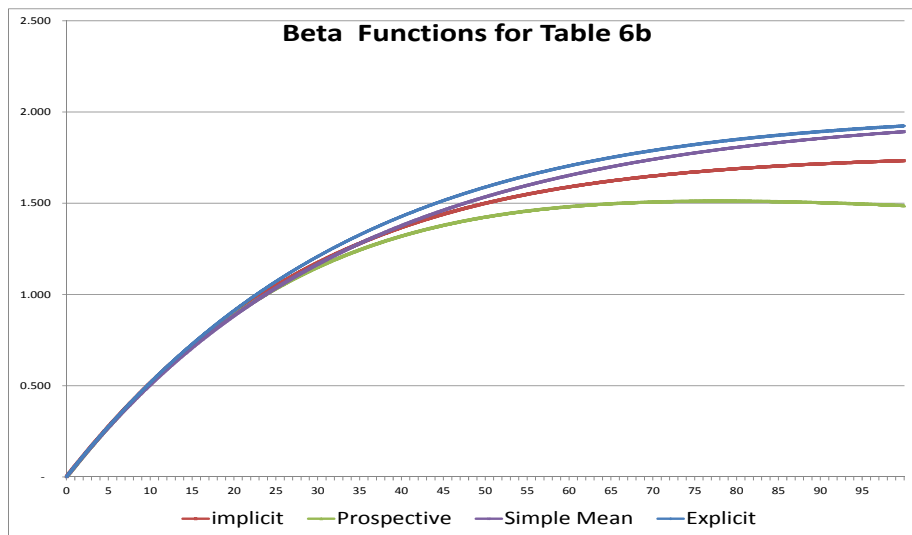
$$0.010 \rightarrow 0.005 \rightarrow 0.0025 \rightarrow 0.00125 \rightarrow \dots \rightarrow 0.00.$$

We are also using a positive interest rate of 4.00 percent.

Table 6b: Pure Endowment Values									
		$\alpha = 50\%$		$q = 1.00\%$ $i = 4.0\%$		$\Delta q = -0.50\%$ $\pi = 6.0\%$			
Maturity		Implicit		Prospective		Simple Mean		Explicit	
n	V_0	V_M	$V_M + EC$	V_M	$V_M + EC$	V_M	$V_M + EC$	V_M	$V_M + EC$
1	952	952	957	952	957	952	957	952	957
5	782	785	803	785	803	784	803	784	803
10	611	620	647	620	647	619	646	619	647
25	292	315	343	314	342	314	342	314	344
50	85	108	122	108	121	109	123	109	124
75	25	39	44	38	43	39	45	39	46
100	7	14	16	13	15	14	17	15	17

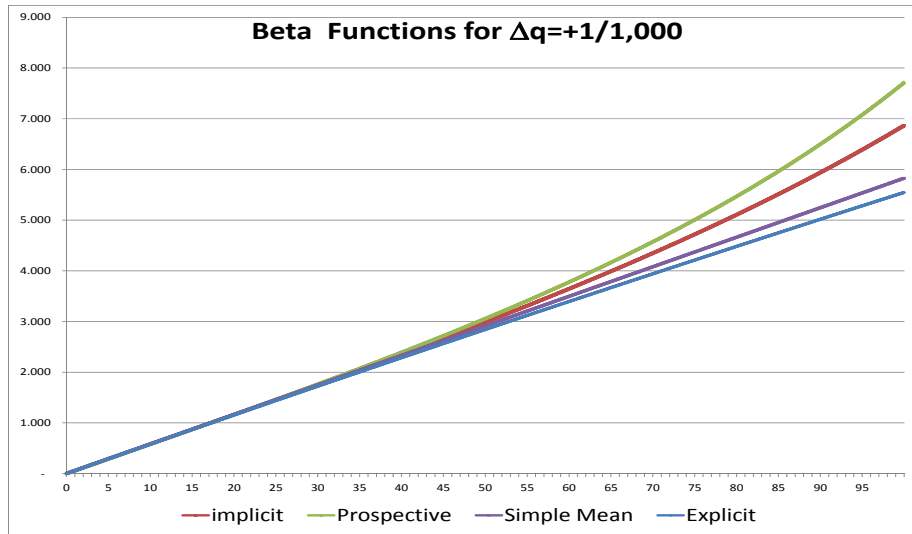
These results are not only closer together but much more reasonable. Any long-term differences between the four margin methods are now discounted with both interest and persistency. This is clearly a better set of assumptions.

The chart below shows the four beta functions, out to 100 years, for this modified example.



As expected, beta is less than $2 = 1/(1 - \alpha)$ for all four methods. There are no negative decrements (resurrections) going on.

As a final example, we show the beta functions that would apply in something more like a life insurance situation where $\Delta q = +1/1,000$. Again we use $\alpha = 1$. The interest rate does not affect beta.



We see the beta functions are very similar out to about 40 years. For many practical problems, all four methods will therefore produce very similar results.

We believe that the work presented here provides a solid theoretical foundation for all four shortcut methods as long as they are used with appropriate inputs.

One critique of the current Solvency II specification is that it implicitly assumes $\alpha = 1$ without considering the appropriateness of that assumption. Assuming $\alpha = 1$ is often reasonable, but not always, as the lapse-supported example in this section shows.

The Solvency II issue is exacerbated by the use of an illiquidity premium since this speeds up the grading process in the risk-loading calculation.

The stated Solvency II rationale for not allowing the use of the illiquidity premium in the margin calculation is that the margined cash flows are not as predictable as the best estimates. This makes sense if you look at the primal problem but not if you look at the issue through the eyes of the dual model. As the dual calculation shows, the margined cash flows are just as predictable as the best estimate. The author therefore takes the view that an illiquidity premium, if used at all, should apply to the entire process.

Finally, it should be noted that nowhere in this paper has a prima-facie case been made for using the geometric shock hierarchy in practice. It has been introduced here because it is a reasonable starting point and it is a useful conceptual tool for understanding how the more pragmatic shortcuts behave. If a good argument could be given for using an alternative shock hierarchy, there is no technical reason not to use it. Such a model might have to be implemented using Monte Carlo simulation or some alternative shortcut.

Risk Interaction and Diversification

Risk interaction deals with the issue of whether margins in the assumptions for one risk should, or should not, affect the assumption margins for another risk. Risk diversification deals with the more statistical issue of whether the capital required to back a mix of risks is more, or less, than the sum of the individual required capital amounts.

Risk Interaction

A simple example of the risk interaction issue was already encountered earlier when we added parameter risk to the first-stage contagion risk model. When we wrote down the basic equation

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

we were implicitly assuming that the present value of margins set up for parameter risk would be available, if we had to deal with a contagion event. This allowed us to hold a slightly lower amount of economic capital for contagion risk than would otherwise be the case. Making this assumption did not change the resulting contagion-loaded, best-estimate mortality. It is still $\mu_0 + \pi\Delta Q$. We will call this the *natural interaction* approach where margined assumptions are determined one by one and then used simultaneously to get a fair value.

An alternative to natural interaction is to deal with each risk in isolation, compute capital and margins for each risk individually and then add the results. This kind of approach is being used in Europe and has an appealing element of simplicity when viewed from a practical point of view.

Simplicity, like beauty, is somewhat in the eye of the beholder. Natural interaction looks simple when viewed from the dual or financial engineering point of view and we take that view in this paper.

As an example, we extend our simple life insurance example to include a deterministic cash value $CV(t)$ payable on surrender. To value this benefit we introduce a new margined force of decrement $w + \beta^w \Delta w$ where, for definiteness, we have chosen the explicit margin approach to evolve β^w , i.e., $d\beta^w = (\pi - \beta^w \Delta w)(1 - (1 - \alpha^w)\beta^w)$. The financial engineering equation for $V = V(s, \beta, \beta^w)$ can now be written as

$$\begin{aligned} \frac{\partial V}{\partial s} + (\pi - \beta \Delta \mu)(1 - (1 - \alpha)\beta) \frac{\partial V}{\partial \beta} + (\pi - \beta^w \Delta w)(1 - (1 - \alpha^w)\beta^w) \frac{\partial V}{\partial \beta^w} \\ + (\mu + \beta \Delta \mu)(F - V) + (w + \beta^w \Delta w)(CV - V) = rV + g - e. \end{aligned}$$

The solution to this equation is a traditional actuarial calculation using the two risk-loaded decrement assumptions:

$$V(t) = \int_t^\infty e^{-\int_t^s (r + \mu + \beta \Delta \mu + w + \beta^w \Delta w) dv} [F(\mu + \beta \Delta \mu) + CV(s)(w + \beta^w \Delta w) + e - g] ds.$$

We can also present the value as

$$V(t) = \int_t^\infty e^{-\int_t^s (r + \mu + w) dv} [\{(F\mu + CV w) + e - g\} + \beta \Delta \mu(F - V) + \beta^w \Delta w(CV - V)] ds.$$

In this expression, all discounting is being done using decrements without any margin for parameter risk. The present value of the cash flows in the curly brackets {}, represents the value

with no margin for parameter risk, while the last two terms represent the present value of margins for parameter risk.

The first term would be the same whether we used natural interaction or a Solvency II style add-up approach to interaction. The assumed natural interaction means that the present value of all margins is available when calculating the net amounts at risk $F - V, CV - V$. For the problem discussed above this leads to a lower present value of margins and capital requirements, relative to the alternative add-up approach. In the next section we will see an example where natural interaction makes the present value of risk margins go up.

When calculating capital we should be consistent with the approach taken to margins. If we use the natural interaction approach then the best way to ensure consistency is to estimate the capital required for parameter risk using the “greek” method, i.e., estimate the derivatives $\frac{\partial V}{\partial \beta}, \frac{\partial V}{\partial \beta^w}$ on the valuation date.

Another approach is to calculate two shocked scenarios. The first would estimate capital for mortality risk using assumptions $(\mu + \Delta\mu + \alpha\beta\Delta\mu, w + \beta^w\Delta w)$ and the second scenario would use assumptions $(\mu + \beta\Delta\mu, w + \Delta w + \alpha^w\beta^w\Delta w)$ to estimate the capital required for lapse risk.

This is equivalent to the greek method when using explicit margin dynamics and is usually a good approximation for the other margin models.

Risk Diversification

This is an independent issue from risk interaction. The issue here is whether any potential diversification benefits between underwriting risks or between underwriting risks and other risks on the balance sheet, such as market risk and credit risk, should be taken into account when estimating fair value margins.

If we answer no, then we are done. If we answer yes, then we have a potentially thorny issue to deal with. The actual diversification benefit available at any point in time will depend on an entity’s mix of risks and approach to capital aggregation. Both of these could change over time.

Furthermore, it is not clear that the entity-specific diversification benefit is what matters. If we want to be able to pay another insurer to take on our liabilities then that entity’s diversification structure could also be relevant.

The method outlined here is designed to allow an entity-specific approach to diversification for ongoing day-to-day risk management while recognizing that we may need to change the approach to actually exit the business. We do this by applying the three-step risk modeling process outlined earlier.

A practical way to begin is to ask what the marginal increase in total required capital is if we take on \$1 of risk capital. Most aggregation models can come up with a factor D , for each risk,²⁰ such that the marginal increase in required capital is $\$D$. We will call this a diversification factor

²⁰ The factor D gets larger as the amount of any given risk grows. Since most life insurance balance sheets are dominated by credit risk the diversification factor for underwriting risk is often on the order of 50 percent.

for the given risk. Furthermore, most aggregation models also have the property that the sum of all marginal contributions is the same as the aggregate capital amount.²¹

A simple example is to assume we use a correlation matrix ρ_{ij} to aggregate capital by

$$C = \sqrt{\sum_{ij} \rho_{ij} c_i c_j}$$

where the c_i are the undiversified capital requirements by risk. If we now set

$$D_i(c_1, c_2, \dots, c_n) = \frac{\partial C}{\partial c_i} = \frac{\sum_j \rho_{ij} c_j}{C}$$

then we can model the diversified capital requirement as a sum

$$C = \sum_i D_i c_i,$$

where the diversification factors D_i depend only on the relative mix of risks, i.e., if $\lambda > 0$ then $D_i(\lambda c_1, \lambda c_2, \dots, \lambda c_n) = D_i(c_1, c_2, \dots, c_n)$.

For our combined mortality/lapse example we assume the entity's internal aggregation model has produced two factors— D_m, D_w —for mortality and lapse risk, respectively. These factors reflect the entity's specific risk situation at a point in time. We can also imagine similar factors— \bar{D}_m, \bar{D}_w —being available, which reflect an industry standard mix of risks. We'll assume that these industry factors are effectively constant.

At the valuation date, the entity-specific marginal cost of holding capital for mortality and lapse parameter risk can then be written as

$$\pi[D_m(\hat{V}^m - V) + D_w(\hat{V}^w - V)] = \pi_m(\hat{V}^m - V) + \pi_w(\hat{V}^w - V).$$

We can write down a similar expression using industry standard diversification factors:

$$\pi[\bar{D}_m(\hat{V}^m - V) + \bar{D}_w(\hat{V}^w - V)] = \bar{\pi}_m(\hat{V}^m - V) + \bar{\pi}_w(\hat{V}^w - V).$$

At this point there are a number of reasonable options.

1. Ignore any entity-specific issues and value using an industry standard approach. This has the advantages of simplicity and consistency with the idea that, if the business were sold to another insurer, the entity-specific diversification benefit could be irrelevant to the buyer. The disadvantage of this approach is that the actual margin release, built into the insurance liabilities, may not line up with the real-world cost of capital, i.e., this is not what we want for “going concern” risk management. For a large multiline insurer the difference between entity-specific and industry-standard diversification factors may not be material.

²¹ This is true if the aggregation model is homogeneous. See the reference in footnote 5 for more detail.

2. Take a more sophisticated approach where we apply our three-step risk analysis process to determining a set of dynamic diversification factors. For example:
 - a. Develop a best-estimate model for the time evolution of the entity-specific diversification factors D_i . A simple example would be to assume they remain constant.
 - b. A plausible shock is that we have to revalue the liabilities using industry-standard diversification factors $D_i(t) \rightarrow \bar{D}_i(t)$. This means we hold capital for the difference in margins that would be required if industry-standard diversification was used.
 - c. Develop a reasonable parameter shock ΔD and choose an α factor to reflect the fact that the actual mix of risks, etc., will not stay constant. This is the standard approach to parameter risk used in this paper.

A simple model, based on the above considerations, would be to develop a set of shocked diversification factors \hat{D}_i that capture both issues (b) and (c) above and then use

$$D_i(t + s) = D_i(t) + \bar{\beta}(t, s, \alpha)(\hat{D}_i - D_i(t)),$$

$$d\bar{\beta} = \pi(1 - (1 - \alpha)\bar{\beta}).$$

The implied economic capital $\frac{\partial V}{\partial \bar{\beta}}$ associated with the margin variable $\bar{\beta}$ is then an appropriate amount of capital to hold for diversification parameter risk.

More sophisticated approaches are possible.

A common denominator of all these approaches is that they can be implemented by allowing the cost of capital rates to be dynamic and vary by risk. While this is more complex than a simple approach, it has the advantage of treating the issue of diversification benefits as just another risk model, subject to the continuous model improvement process.

Practical Pros and Cons—Primal vs. Dual Approach

For the simple actuarial problems discussed so far we have introduced both the primal and dual approaches to implementing the cost of capital model and shown that they are theoretically equivalent. The main purpose of this section is to outline some of the practical issues that can arise as the complexity of the risk model application increases.

The primal or actuarial approach has the advantage of being intuitive to most actuaries so it is not hard to explain. The disadvantage is that it can lead to “projection within projection” or nested projection issues as the complexity of the application increases. This is in addition to any nested stochastic issues raised by the economic model. We use the specific example of modeling mortality improvement to illustrate the problem.

The dual or financial engineering approach may be less intuitive to some people but it can be formulated so as to avoid the nested projection issue. It also has the advantage of greater transparency because it can produce explicit risk-loaded or margined assumptions that can be reviewed for reasonableness.

The dual can also be used to produce risk-loaded cash flows that could be used in a replicating portfolio or A/LM process. This is typically not an option when using the primal approach which usually forces the insurer to do their A/LM on a best-estimate cash flow basis.

The biggest current disadvantage of the dual approach appears to be that it is new, so it is not supported on insurance industry standard actuarial platforms at the time of writing (June 2014).

Mortality Improvement

This section overviews the extension of the mortality risk model to include mortality improvement. The product considered is a 20-year term insurance with no lapses, premiums or expenses. The interest rate is 4.00 percent, and the cost of capital rate is 6.00 percent compounded annually.

We present three versions of the model. The first one uses the actuarial version of the prospective method with no interaction among the risks. This is fairly easy to explain, but the actual calculations require an element of nested projection to get answers for mortality improvement.

The second approach illustrates a financial engineering method without natural interaction. This method is not theoretically equivalent to the first approach, but the results are very similar anyway. A key advantage of this approach is that there is no need for any nested projections, and we can explicitly see what the margined assumptions look like. A disadvantage is that the method is harder to explain.

The third approach extends the financial engineering method by allowing for natural interaction. For this example, the assumed interaction makes the results more conservative and we are able to explain why.

For each of the three calculations we use the following common assumptions:

1. Best-estimate mortality is the same as was used in previous examples except that we now have a best-estimate mortality improvement assumption of $\lambda_0 = 1.5\%$ per year. More precisely, the new best-estimate force of mortality is $\Lambda_0(s)\mu_0(s) = e^{-\lambda_0(s-t)}\mu_0(s)$ for $s > t$.
2. For contagion shocks we consider the usual idea of bad experience in a given year with $\Lambda_0(s)\Delta Q(s) = (\frac{1}{2})\Lambda_0(s)\mu_0(s)$ extra deaths, and we also consider the possibility of a contagion shock to the cumulative improvement factor $\Lambda_0(s) \rightarrow \Lambda_0(s)(1 + k)$. One reasonable choice for k might be $k = 3\lambda_0$. This could be justified by assuming mortality studies are only done once every three years so it is possible that the next study comes to the conclusion that the expected improvement has failed to materialize for three years. For the numerical examples we have assumed $k = 5.0\%$.

- Parameter shocks are assumed to be $\Delta\mu = 0.05(\mu_0 + \pi\Delta Q)$ and $\Delta\lambda = -0.5\%$. We set $\alpha = 100\%$ for mortality but use $\alpha = 50\%$ for the improvement model. This means the ultimate assumed improvement rate is $\lambda_0 + \Delta\lambda/(1 - 0.5) = 0.5\%$.

It is worth noting that a contagion shock to the cumulative improvement $\Lambda_0(s) \rightarrow \Lambda_0(s)(1 + k)$ is almost the same thing as a parameter shock to the base level mortality $\mu_0 \rightarrow \mu_0(1 + k)$ with $\alpha = 1$. This is because if $\Delta\mu = k\mu$ and $\alpha = 1 + k$ then

$$\Lambda(1 + k)(\mu + \beta_\mu\Delta\mu) = \Lambda(\mu + \Delta\mu + \alpha\beta_\mu\Delta\mu).$$

It could then be argued that using both shocks is double counting, or that the total shock is really the sum $5\% + 5\% = 10\%$. We will carry both risks forward in the analysis anyway expecting the improvement contagion cost to be roughly equal to the level parameter cost.

The Prospective Calculation

The calculation steps for this approach are:

- Develop a best-estimate projection $V_0(s)$ for $s \geq t$ using base assumptions.
- Develop a second projection $V_1(s)$ using base improvement but the shocked mortality level.
- Develop a third projection $V_2(s)$ that uses base mortality level and improvement up to time s but uses shocked improvement for future times. This requires a nested projection calculation.
- Develop a fourth projection $V_3(s)$ that uses base mortality up to time s but then applies an improvement contagion shock at that point in time. This could be very similar to item (3) above.
- Given these projections we can estimate the required margins M_j by discounting the costs of capital $\pi[V_j(s) - V_0(s)]$ at the rates $i + \pi(1 - \alpha_j)$ as appropriate. For mortality level contagion risk we have estimated the required capital as $(0.05)\Lambda_0(s)\mu_0(s)[F - V_0(s)]$.
- Economic capital at the valuation date is then calculated as

$$EC_j = [V_j(t) - V_0(t) - (1 - \alpha_j)M_j].$$

Some numerical results are in Table 7a.

Table 7a: Mortality Improvement Models					
Best Estimate Value		2,605.92	Face	100,000	
		Contagion Risk		Parameter Risk	
		Level	Improvement	Level	Improvement
				Total	
Prospective/ Actuarial No Interaction					
Margin	76.90	82.99	82.99	51.11	293.99
Capital	48.33	127.76	127.76	120.55	424.40

Under the prospective method an improvement contagion shock and a level % parameter shock are actually identical if we use the same discretization assumptions.

A Dual/Financial Engineering Approach

For this approach we let the risk-loaded force of mortality be written as $\Lambda(\mu + \beta_\mu \Delta\mu)$ where $\mu = \mu_0 + \pi\Delta Q$ as before and the cumulative improvement factor Λ has the dynamics

$$d\Lambda = -\Lambda[\lambda_0 - \pi k + \beta_\lambda \Delta\lambda]ds, \Lambda(t) = 1.$$

This is equivalent to risk-adjusting the best estimate improvement rate λ_0 by a static margin πk for contagion risk and a dynamic margin $\beta_\lambda \Delta\lambda$ for parameter risk. The margin variables β_μ, β_λ are assumed to evolve according to

$$\begin{aligned} d\beta_\mu &= (\pi - \beta_\mu \Lambda \Delta\mu)(1 - (1 - \alpha_\mu)\beta_\mu)ds, \alpha_\mu = 1, \beta_\mu(t) = 0, \\ d\beta_\lambda &= \pi(1 - (1 - \alpha_\lambda)\beta_\lambda)ds, \alpha_\lambda = \frac{1}{2}, \beta_\lambda(t) = 0. \end{aligned}$$

For the mortality level margin variable β_μ we have chosen explicit model dynamics with the wrinkle that the parameter shock is being driven by $\Lambda\Delta\mu$ rather than $\Delta\mu$. This is a deliberate modeling choice justified below.

We are using a simple mean dynamic for the improvement parameter because it is simple and there is not, at this time, a better theoretical alternative for this application.

In the valuation measure, the fair value is a function of time and the three dynamical variables, i.e., $V = V(s, \Lambda, \beta_\lambda, \beta_\mu)$. The financial engineering equation, based on the above dynamics, is given by

$$\begin{aligned} \frac{\partial V}{\partial s} - \Lambda[\lambda_0 - \pi k + \beta_\lambda \Delta\lambda] \frac{\partial V}{\partial \Lambda} + \pi(1 - (1 - \alpha_\lambda)\beta_\lambda) \frac{\partial V}{\partial \beta_\lambda} + (\pi - \beta_\mu \Lambda \Delta\mu)(1 - (1 - \alpha_\mu)\beta_\mu) \frac{\partial V}{\partial \beta_\mu} \\ + \Lambda(\mu + \beta_\mu \Delta\mu)(F - V) = rV + g - e. \end{aligned}$$

On the valuation date, when $s = t$, the P measure total expected rate of change in the liability is given by

$$\begin{aligned} \frac{\partial V}{\partial s} - \Lambda(t)\lambda_0 \frac{\partial V}{\partial \Lambda} + \Lambda(t)\mu_0(F - V) \\ = rV + g - e - \pi[\Lambda(t)\Delta Q(F - V) + k\Lambda(t) \left[\frac{\partial V}{\partial \Lambda} + \frac{\partial V}{\partial \beta_\mu} + \frac{\partial V}{\partial \beta_\lambda} \right]]. \end{aligned}$$

The four terms in the square bracket at right are clearly the implied capital requirements for this model. Three of the terms are familiar from previous work in this paper. The term for improvement contagion makes sense if we accept the modeling shortcut

$$V(s, \Lambda(1 + k), \beta_\lambda, \beta_\mu) - V(s, \Lambda, \beta_\lambda, \beta_\mu) \approx k\Lambda(s) \frac{\partial V}{\partial \Lambda}.$$

Using this shortcut has allowed us to model the margin for improvement contagion by using a static loading in the improvement rate.

If it works, doing one scenario with a dynamic margin in the improvement rate is much simpler than the nested projection required by the prospective method.

In practice, the greeks $\frac{\partial V}{\partial \Lambda}$, $\frac{\partial V}{\partial \beta_\mu}$, $\frac{\partial V}{\partial \beta_\lambda}$ will be estimated using numerical methods but it is worth looking at them analytically to see if they are reasonable for this application. We can get theoretical expressions for each of them by differentiating the financial engineering equation with respect to each dynamical variable. The result is a system of equations for the greeks that can be solved as follows:

$$\begin{aligned}\frac{\partial V}{\partial \beta_\mu}(t) &= \int_t^\infty e^{-\int_t^s (r + \Lambda(\mu + \Delta\mu + \alpha_\mu \beta_\mu \Delta\mu)) dv} \Lambda(s) (1 - (1 - \alpha_\mu) \beta_\mu) \Delta\mu (F - V) ds, \\ \Lambda(t) \frac{\partial V}{\partial \Lambda}(t) &= \int_t^\infty e^{-\int_t^s (r + \Lambda(\mu + \beta_\mu \Delta\mu)) dv} \left\{ \Lambda(s) (\mu + \beta_\mu \Delta\mu) (F - V) \right. \\ &\quad \left. + \Lambda(s) \beta_\mu \Delta\mu (1 - (1 - \alpha_\mu) \beta_\mu) \frac{\partial V}{\partial \beta_\mu}(s) \right\} ds, \\ \frac{\partial V}{\partial \beta_\lambda}(t) &= \int_t^\infty e^{-\int_t^s (r + \Lambda(\mu + \beta_\mu \Delta\mu)) dv} (1 - (1 - \alpha_\lambda) \beta_\lambda) \Delta\lambda \Lambda(s) \frac{\partial V}{\partial \Lambda}(s) ds.\end{aligned}$$

The first of these equations is exactly what we would want to see for mortality level risk. The dynamics for β_μ were chosen to achieve this result.

The first term in the second equation for $\Lambda \frac{\partial V}{\partial \Lambda}$ shows that this quantity is, almost, the present value of risk charges on the valuation scenario. This makes sense. The second term in the integrand is due to risk interaction and arises from our choice of simple mean dynamics for β_λ . Presumably, a more sophisticated dynamic for β_λ could make this term go away but, since it is of order $\Delta\mu^2$ it should be immaterial²² in most practical situations.

The final equation for $\frac{\partial V}{\partial \beta_\lambda}$ makes sense once we think of $\Lambda \frac{\partial V}{\partial \Lambda}$ as the present value of risk charges. One could argue that the discounting should use shocked improvement rather than base improvement. This could be considered a theoretical error that might be corrected by using a more sophisticated dynamic for β_λ .

The conclusion from this theoretical analysis is that, while the model's theory could probably be improved upon, it may well be good enough in practice. One way to test this conclusion is to work through the same numerical example that was used to illustrate the prospective method.

A discrete time spreadsheet implementation of the model described above was developed using the tools and concepts introduced earlier. Table 7b presents results comparable to Table 7a.

²² Note that the derivative term $\frac{\partial V}{\partial \beta_\mu}$ is also of order $\Delta\mu$.

Table 7b: Mortality Improvement Models					
Best Estimate Value	2,605.92		Face	100,000	
	Contagion Risk		Parameter Risk		Total
	Level	Improvement	Level	Improvement	
Prospective/ Actuarial No Interaction					
Margin	76.90	82.99	82.99	51.11	293.99
Capital	48.33	127.76	127.76	120.55	424.40
Dual/ Financial Engineering No Interaction					
Margin	76.79	83.93	82.34	50.70	293.76
Capital	48.29	131.90	127.78	119.24	427.22
Dual/ Financial Engineering Natural Interaction					
Margin	76.27	89.12	87.40	55.28	308.07
Capital	48.19	142.63	138.18	132.47	461.46
Dual/ Financial Engineering Natural Interaction					
Margin	76.28	-	171.15	55.20	302.63
Capital	48.19	-	267.30	132.13	447.62
Dual/ Financial Engineering Natural Interaction					
Margin	76.26	178.27	(0.00)	55.31	309.84
Capital	48.18	285.43	(0.00)	132.58	466.20

The first row of values above is a repetition of Table 7a. The next line shows what happens if we use the dual model, one risk at a time. The results are very similar. This is gratifying since the computational cost of running the dual model is lower. The setup is clearly more complex.

The cost of mortality contagion has dropped a small amount because the margins for other risks are being used to reduce the net amount at risk when calculating capital for this risk.

The numerical results for improvement contagion and level parameter risk are no longer identical, but they are still close as one would expect. This shows that a dynamic $k\%$ margin in the base level mortality, with $\alpha = 1$, is almost the same thing as a static $\pi k\%$ margin in the improvement rate. This is a useful insight.

The last result in this line shows that a nested projection calculation really can be replaced by a single scenario with a dynamic margin.

The third row of results shows what happens when we run three risks simultaneously. We now see interaction effects as follows:

- The cost of mortality level contagion risk has gone down because we can use the margins for other risks to reduce the net amount at risk. The effect is small.
- One reason the cost of mortality level parameter risk has gone up is because the 5 percent shock now applies to both the best estimate and the contagion shock. We would expect the margin to go up by 50% of 6% = 3%, and this explains about half the increase. The rest of the increase comes about because the natural interaction model is using margined mortality improvement for discounting.
- The cost of improvement parameter risk has gone up because the assumed improvement now applies to a base affected by the margins for mortality level risk. This is also where we would expect to see the impact of any theoretical error in the dynamics for β_λ .

The last two rows in Table 7b were calculated to show what happens if we decide to use only one of the improvement contagion or mortality level shocks. In each row we drop one risk and then double the size of the shock to 10 percent for the one we keep.

Since the introduction of interaction results in more material changes than the move from primal to dual calculation, it seems safe to conclude that the dual approach itself has not introduced any material errors.

In the author's opinion, natural interaction makes more sense from a theoretical perspective. From a practical point of view it can make the answers go up or down as this example shows.

There is also a new practical problem when using the dual method with natural interaction. The model really only produces a single present value of margins for all risks combined. If we want a breakdown we have to develop an allocation scheme. To get the results reported on the last line of Table 7b we did a full financial projection, on best-estimate assumptions, to get the required capital for each risk at all points in time. This gives us an accurate margin allocation but at the cost of having to do nested projections. At that point we have lost some of the comparative advantage of the dual approach.

Other, simpler, approaches to allocating the margin by risk are possible but are beyond the scope of this paper.

This section has used the example of a mortality improvement problem to illustrate some of the practical pros and cons of using a primal versus a dual approach to implementing the cost of capital method. The key takeaway is that a dynamic margin calculation can be used to replace the relatively expensive nested projection required by a more traditional approach.

Down but Not Out and Professional Standards

Earlier in this paper we made the statement that our “Down but Not Out” theory won't work in practice unless the shocked balance sheets presented by management to external investors have credibility. If the revised balance sheet is not credible to outside investors, no one will put up the new economic capital required to continue. At that point the risk enterprise is “Down and Out”—a situation all stakeholders, especially regulators, should want to avoid.

If an investor asks why he should believe management's choice of assumptions and calculations, an answer is that the professionals who put the shocked balance sheet together were a) properly

credentialed; and b) were following appropriate industry standards of practice for this kind of exercise. A history of rigorous audits and/or peer reviews would also go a long way toward increasing confidence in the process.

One model for a standard-setting process is the one developed by the actuarial profession in Canada. In the late 1980s the Canadian Institute of Actuaries had to come up with a set of professional standards to support the implementation of Canadian GAAP for life insurers in 1992. This was a principles-based reserve system founded on the idea that each assumption for mortality, lapse, etc., was the sum of a best estimate and a Provision for Adverse Deviation (PfAD).

A very brief summary of their approach is as follows.

1. Appointed actuaries had the freedom to choose best-estimate reserve assumptions that were appropriate to the circumstances of the company. However, those assumptions had to be documented and confidentially reported to regulators once a year. The assumptions were also subject to peer review every three years.
2. Risk margins or PfADs were limited (in 1992) to what this paper would call static margins. The guidance for choosing the static margin involved a two-step process:
 - a. Each risk had to be classified into a risk bucket (high, medium, low) using qualitative criteria set out in the guidance.
 - b. Once a risk classification was made, the guidance gave a numerical range of acceptable risk margins. It was up to the appointed actuary to pick, and defend, a value in the acceptable range.
3. Actuarial reserves were supplemented by a rules-based capital requirement (MCCSR) similar to the U.S. RBC model of the time. The risk margins and capital requirements were not consistent with each other, so an embedded value calculation was necessary to put a value on the company. Some Canadian companies started publishing embedded value results in the early 2000s.
4. The final element in the Canadian model was Dynamic Capital Adequacy Testing or DCAT. This was an early version of the ORSA concept where the actuary was asked to model a number of stress scenarios over a five-year time horizon. The objective was to make sure that adequate capital resources would be available if needed.

Such an approach represented a compromise between giving appointed actuaries the freedom to do something reasonable, and the desire to see some uniformity of practice from one company to another. It clearly relied on the actuary's professionalism.

In principle, there is no reason why this kind of approach to standards could not be applied to the "Down but not Out" model developed in this paper. Guidance would have to be developed for both the static and dynamic margin components of the model. Once these assumptions are set, risk margins and capital requirements could then be calculated that are consistent with each other

in the sense that an actuary would not have to do an embedded value calculation to figure out what the enterprise was worth.

The static margins that come out of the current approach would likely be much smaller than current Canadian GAAP PfADs. One reason is that the current model asks the static margin to deal only with the contagion issue.

A starting point for the determination of dynamic margin shocks is the guidance currently used by European regulators for Solvency II. Their approach could be modified to allow the use of “liquidity” buckets where the size of a parameter shock varies by bucket.

Conclusion

This paper has developed a “Down but Not Out” risk management paradigm that is suited to a principle-based, fair-value approach to risk management. We have outlined the basic theory and also shown that there are several practical ways to implement the approach.

The basic idea is that all best-estimate models can be wrong in two different ways, and this can be addressed by two different kinds of risk margin structure.

1. Model errors in the short run are handled with static margins.
2. Model errors in the long run (assumption changes) are handled with dynamic margins.

We have shown that there are two different ways of approaching the practical calculations that we have called the primal and dual approaches, respectively.

The dual approach (dynamic margins) adds transparency and can also be a significant computational shortcut as the complexity of the application increases. It could also allow the A/LM process to focus on margined cash flows rather than best-estimate cash flows.

The main motivation of “Down but Not Out” is sound risk management by creating an environment where constant model improvement is encouraged. If the process is implemented with the appropriate professional integrity, then a risk enterprise should be able to withstand plausible adverse shocks and still be able to recover by going to the capital markets to replace any lost capital.

The paper has also presented a number of simple numerical examples to illustrate the ideas. The examples include the mortality issues of contagion, a shock to the base level of mortality, and a shock to future mortality improvement.

More complex examples have been developed but are not included here for reasons of space.

Two public sources that provide additional examples of the risk modeling paradigm in practice can be found in Manistre,²³ where the approach is applied to long-dated equity options, and Manistre,²⁴ where the focus is credit risk.

The model presented here is not focused on regulatory or accounting compliance. It is focused on risk management. However, the author believes this is a reasonable approach to ORSA modeling that could then be leveraged to meet other needs such as Solvency II or IFRS reporting by making a relatively small number of adjustments. In particular, the fact that “Down but Not Out” recognizes both current period and future (assumption change) gains and losses is consistent with IFRS disclosure requirements.

This model is not consistent with some current developments in U.S. GAAP that plan to bury the risk margin in with a “residual” margin used to prevent the recognition of profits at issue. In this author’s opinion the U.S. accounting profession appears to be more concerned about the possible manipulation of earnings than the transparency needed for sound risk management.

The model is consistent with the traditional actuarial concept of setting the present value of risk margins equal to the cost of holding economic capital. This means the surplus on the economic balance sheet really is surplus and any amount over and above required capital is economic free surplus. This is an important element of transparency that is missing in most other accounting or regulatory models.

Acknowledgments

The material in this paper has evolved over the last eight years of the author’s actuarial career. During that time I have benefited from many conversations with actuarial colleagues and other risk professionals. I would like to explicitly acknowledge a few brave souls who took the time to read an earlier draft of this document and provide useful feedback. They include Steve Strommen, FSA, MAAA; Mac Smith, FSA, FCIA; and Andrey Marchenko, Ph.D., FRM. I would also like to thank Nazir Valani, FSA, FCIA, and his actuarial team at KPMG Canada for their input and support.

Appendix—Systems of Linear Stochastic Equations

The previous sections of this paper assumed we were working in a deterministic economic environment. The primary purpose of this section is to show that nothing really changes when we go to a stochastic economic environment. The basic insight is that if we estimate capital, and hence margins, by using only the information available on a specific scenario, we get an unbiased estimate of the correct capital. This means that, when we average over all risk-neutral scenarios in a stochastic model, the margin errors made on individual scenarios average out to zero. This result is sometimes called the “Law of Iterated Expectations”.

²³ Manistre, B.J., “A Cost of Capital Approach to Extrapolating an Implied Volatility Surface.” A paper presented at the 2010 ERM Symposium in Chicago. This paper can be found on the Society of Actuaries website at www.soa.org. A summary can also be found in the 2010 CRO Forum Best Practice paper “Extrapolation of Market Data.”

²⁴ Manistre, B.J. “A Cost of Capital Approach to Credit and Liquidity Spreads.” This is a PowerPoint presented at a 2009 University of Waterloo QRISK seminar. The idea was to use static margins to model credit contagion risk and dynamic margins to capture liquidity issues. The resulting model is a cousin to the Jarrow, Lando & Turnbull approach.

The above statement may sound fantastic, but it is actually a standard result in the theory of linear stochastic differential equations and can be found in many textbooks on stochastic calculus. We present a short version of the argument here. It is a slightly more general version of the dual analysis presented earlier in this paper.

The importance of this result for applications is that if, for some practical reason, we modify the equations so that they become non-linear in some way, then we may invalidate this important result.²⁵

The system we will discuss here is general enough to include the shortcuts developed earlier as special cases. We will analyze the system

$$E[dV_k] = [-CF_k(t) + \sum_{j=0} H_{kj}(t)V_j(t)]dt, \quad k = 0, 1, \dots, n$$

where all of the quantities CF_k , H_{kj} could be stochastic processes.

To start we will rewrite the system above as

$$dV_k = [-CF_k + \sum_{j=0} H_{kj}V_j]dt + \sum_B \sigma_{kB} dz^B(t), \quad k = 0, 1, \dots, n, \quad B = 1, \dots, m$$

where the matrix σ_{kB} models the response of the variables to the random noise $dz^B(t)$.

Now for fixed t , and variable s , let $\Phi_{ij}(t, s)$ be a process that satisfies $\Phi_{ij}(t, t) = \delta_{ij}$ and $d\Phi_{ij}(t, s) = \dot{\Phi}_{ij} ds + 0dz$. We call $\Phi_{ij}(t, s)$ the deflator matrix.

Consider the quantity

$$\begin{aligned} d[\sum_j \Phi_{ij}(t, s)V_j(s)] &= \sum_j \dot{\Phi}_{ij}(t, s)V_j(s)ds + \sum_j \Phi_{ij}(t, s)dV_j(s) \\ &= \sum_j \dot{\Phi}_{ij}(t, s)V_j(s)ds + \sum_j \Phi_{ij}(t, s)\{-CF_j + \sum_{k=0} H_{jk}V_k\}ds + \sum_B \sigma_{jB} dz^B(s) \end{aligned}$$

After relabeling some dummy summation indices this can be written as

$$d[\sum_j \Phi_{ij}(t, s)V_j(s)] = \sum_j [\dot{\Phi}_{ij}(t, s) + \sum_k \Phi_{ik}H_{kj}]V_j(s)ds - \sum_j \Phi_{ij}(t, s)CF_j ds + \sum_j \Phi_{ij}(t, s) \sum_B \sigma_{jB} dz^B(s)$$

We now choose the stochastic process for the deflator matrix to be

²⁵ Some examples of non-linear systems would be the current regulatory models used in the United States and Canada that use conditional tail expectation (CTE) measures to estimate capital. Their approach makes sense if you think of an insurance enterprise as a closed system going forward (i.e., no future capital infusions considered). It does not make sense if you take the more realistic open system approach considered in this paper where future capital infusions are a normal part of the risk management process.

$$d\Phi_{ij} = \dot{\Phi}_{ij}(t, s)ds = -\sum_k \Phi_{ik} H_{kj} ds$$

this makes the first term in the equation above vanish so that

$$d[\sum_j \Phi_{ij}(t, s)V_j(s)] = -\sum_j \Phi_{ij}(t, s)CF_j ds + \sum_j \Phi_{ij}(t, s) \sum_B \sigma_{jB} dz^B(s).$$

Now perform a stochastic integration from t to T and take an expectation based on information available at time t . The result is

$$E \int_t^T d[\sum_j \Phi_{ij}(t, s)V_j(s)] = -E \int_t^T \sum_j \Phi_{ij}(t, s)CF_j ds + E \int_t^T \sum_j \Phi_{ij}(t, s) \sum_B \sigma_{jB} dz^B(s).$$

The left-hand side of this equation evaluates to $E \sum_j \Phi_{ij}(t, T)V_j(T) - V_i(t)$ (remember

$\Phi_{ij}(t, t) = \delta_{ij}$), and the second term on the right drops out since $E dz^B(s) = 0$. The end result is an expression for the system variables

$$V_i(t) = E \sum_j \Phi_{ij}(t, T)V_j(T) + E \int_t^T \sum_j \Phi_{ij}(t, s)CF_j(s)ds.$$

The key point here is that the deflator matrix has been calculated scenario by economic scenario. There is no need for any “stochastic on stochastic” calculations.