A value-at-risk approach to mis-estimation risk (Discussion)

[Institute and Faculty of Actuaries, Sessional Event, Monday 10 May 2021]

This discussion relates to the paper by Stephen J. Richards presented at the IFoA sessional event held on Monday 10 May 2021.

The Moderator (Mr R. J. Kairis, F.I.A.): Good morning everyone. Thank you everyone for joining. I am Robert Kairis, longevity lead at Lloyds Banking Group. Today's sessional meeting is "A Value-at-risk Approach to Mis-estimation Risk". It gives me great pleasure to introduce Dr Stephen Richards, who is the managing director of Longevitas, a specialist provider of mortality analysis software for actuaries. Stephen (Richards) is an honorary research fellow at Heriot-Watt University, and, as many of us know, he has made big contributions to the modelling of longevity risk over the last decade. With that, I will hand over to Stephen (Richards).

Dr S. J. Richards, F.F.A.: Thanks, Rob (Kairis). Welcome to the presentation for "A Value-at-Risk Approach to Mis-estimation Risk". This paper is a follow-up to Richards (2016) that addresses the specific question of how to look at mis-estimation risk through a value-at-risk approach. Today's presentation is relatively technical and detailed. I am not going to assume that people have read all of the paper. I will first look at some of the background on mis-estimation risk. I will then look at some features of annuity portfolios, albeit many of these features also apply to a lot of other portfolios actuaries work with. I will then recap some basics of parameter estimation, specifically maximum-likelihood estimation, as this is very relevant to mis-estimation risk. I will look at a number of critical preconditions for a mis-estimation assessment to produce valid results. I will then recap what we call run-off mis-estimation, which is a mis-estimation assessment without a value-at-risk element. Then, I will contrast this with the value-at-risk approach to mis-estimation, i.e. how we have to modify our algorithm to look at the problem from a value-at-risk perspective. I will then contrast the results and look at some expected – and unexpected – aspects of the value-at-risk view. I will do a number of detailed comparisons, so that we can get some insight into how this methodology behaves.

We begin with some definitions of mis-estimation risk. A good place to start is with the Prudential Regulation Authority: "the PRA considers that longevity risk includes at least two sub-risks (. . .) namely base mis-estimation risk and future improvement risk" (Woods, 2016). In the PRA's view of the world, longevity risk decomposes into at least two sub risks, namely current mortality rates and future projections. Burgess *et al.* (2010, page 11) described mis-estimation risk as "the risk that the base mortality estimate is incorrect (i.e. the mortality estimate based on actual experience in the portfolio)". An important subsidiary point here is that the mortality basis is presumed to be set based on the actual experience of the portfolio in question. If you are not using the actual experience of the portfolio, or are mixing it with other data, then you need an additional capital allowance for basis risk, i.e. the risk that the data set that you have used to calibrate your mortality basis is not truly representative of the underlying mortality of your

portfolio. In the paper and this presentation, we critically assume that a mis-estimation assessment is based solely on the actual experience data of the portfolio. Armstrong (2013, slide 10) posed the mis-estimation question differently: "how wrong could our base mortality assumptions be, or: what if our recent historical experience did not reflect the underlying mortality?" Armstrong (2013, slide 12) further observed that "mis-estimation risk lends itself to statistical analysis, if there is sufficient accurate data". Makin (2008) further pointed out a crucial actuarial aspect, namely that "the impact of uncertainty should always be quantified financially". It is not enough to look at mis-estimation risk as a statistician, you have to quantify the financial impact of this statistical uncertainty. This background survey gives us four key aspects to quantifying mis-estimation risk: (i) we consider uncertainty over current mortality rates, (ii) we must use actual portfolio experience, (iii) we must model statistically and (iv) we must quantify the financial impact of uncertainty. We will next consider some portfolio features that might pose particular challenges.

We first consider a medium-sized pension scheme with 15,698 lives of pensions in payment, with 2,076 deaths observed in 2007–2012. We sort the lives by pension size and split them into deciles and look at lives- and pension-weighted measures:

Liability concentration in a medium-sized pension scheme. Source: data from Richards (2016, Table A2).
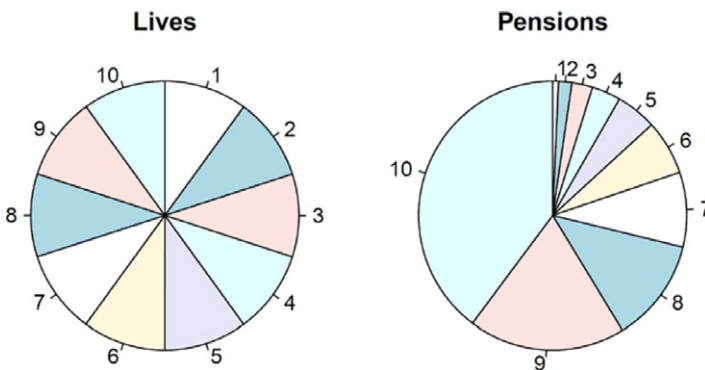
Measured by lives, the deciles are of course equal. But if we weight by pension size, and look at where the liabilities sit, they are radically unequal. We have a considerable degree of concentration risk – decile 10 has 1,567 members of the scheme with the largest pensions, and their total pensions exceed those received by deciles 1–7 of the membership (10,991 lives). This is directly relevant to actuaries because the largest pensions are where most of the financial liability is. The top decile of pensioners here receives about 40% of all total pensions, and the next two deciles receive a further 30%. Totally, 70% of the pension scheme's liabilities are concentrated in just 30% of lives.

To illustrate the importance of this concentration risk, we look at a simple time-varying model of the mortality hazard, $\mu_{x,y}$:

$$\mu_{x,y} = \frac{e^{\epsilon} + e^{\alpha + \beta x + \delta(y - 2000)}}{1 + e^{\alpha + \beta x + \delta(y - 2000)}}$$

where x is the exact age, y is the calendar time and -2000 is an offset to scale the parameters appropriately. This equation represents the force of mortality varying by age x and calendar

time y. We have a Makeham-like constant term, and a Gompertz-like age structure with α and β δ allows for changes in mortality levels in calendar time.

We build a model with an individual mortality rate for each life i as follows:

$$\alpha_i = \alpha_0 + \alpha_{male}z_{i,male} + \alpha_{decile\ 8\ or\ 9}z_{i,decile\ 8\ or\ 9} + \alpha_{decile\ 10}z_{i,decile\ 10}$$

where α_i is the level of mortality. This is built up from, α_0, the intercept, which is common to everyone. We then add a series of additional components, depending on whether or not a given life has a particular risk factor. The indicator variable, $z_{i,j}$, takes the value 1 when life i has risk factor, j and zero otherwise. Other risk factors can be added, such as membership of a postcode group, whether or not they retired early, etc. Using this approach, we can build up an individual mortality rate for each individual life using any number of risk factors. Below is a table showing the results when we fit this particular survival model to this portfolio:

Parameter estimates. Source: Richards ([2016](), Table 6).

Parameter	Estimate	Standard error	Contributing lives
β	0.148	0.005	15,698
α_{male}	0.479	0.060	5,956
α_0	−14.731	0.491	15,698
$\in$	−5.420	0.154	15,698
$\alpha_{decile\ 8\ or\ 9}$	−0.180	0.078	3,140
$\alpha_{decile\ 10}$	−0.313	0.108	1,567
δ	−0.046	0.016	15,698

α_0 is the intercept: all 15,698 lives contribute to that estimate. We also have α_{male}, which is the effect of being male. That estimate is driven by the 5,956 male lives, with the remainder being female. Of particular interest are the parameters for the effect of being wealthy (deciles 8 and 9) or very wealthy (decile 10). The estimate of −0.313 for $\alpha_{decile\ 10}$ is driven by just 1,567 lives. Positive values increase the mortality level, while negative values decrease the level of mortality. The effect of being male adds 0.479 to the base level of mortality on a log scale, whereas being in the top pension decile deducts 0.313. These are large differentials. The level of mortality for the lowest-income individuals is represented by α_0, while people in the 8th and 9th deciles have around 20% lower mortality and the people in the top decile have 30% lower mortality than the baseline (or 10% lower mortality still compared to those in deciles 8 and 9). This is actuarially relevant because deciles 8, 9 and 10 form a disproportionate share of total liabilities.

We next look at the coefficient of variation, which is the standard error of a parameter divided by its estimate:

Coefficients of variation. Source: calculated from Richards (2016, Table 6).

Parameter	Coefficient of variation
β	0.03
α_{male}	0.13
α_0	0.03
$\in$	0.03
$\alpha_{decile\ 8\ or\ 9}$	0.43
$\alpha_{decile\ 10}$	0.35
δ	0.35

We have a small coefficient of variation where we have a small standard error, i.e. a high degree of confidence in an estimate. However, the coefficients of variation for the parameters for the largest pensions are quite large. This is of interest because we now have what seems to be a perfect storm for actuarial work: we have liabilities that are highly concentrated; we have sub-groups with the largest liabilities that also have the lowest mortality and the parameters describing that low mortality also have the highest relative uncertainty. Our liabilities are concentrated in small subgroups where we have the least confidence over their mortality levels.

Let us now quickly refresh some basics of parameter estimation. We assume that we have a statistical model with m parameters. We assume that we have a parameter vector, θ, that corresponds to the parameter estimates column in the previous tables. We have some estimate of, θ denoted, over which there is a degree of uncertainty. This uncertainty is our estimation risk. The more data we have, the less uncertainty we would expect to have; the less data we have, the more uncertainty there should be around $\hat{\theta}$. We make a number of further assumptions. We assume we have a log likelihood function, l, which depends on θ and the data. An example is given below for $m = 1$:

An example log-likelihood function. Source: based on Richards (2016, Figure 1).
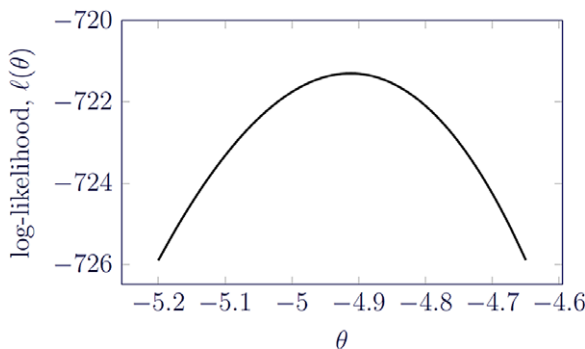
Log-likelihood functions usually have this quadratic, upside-down U shape. We further assume that all first partial derivatives of the log-likelihood exist, and so we can look at the gradient of the log-likelihood, $l'(\theta)$:

**Log-likelihood gradients. Source: based on Richards (2016, Figure 1).**
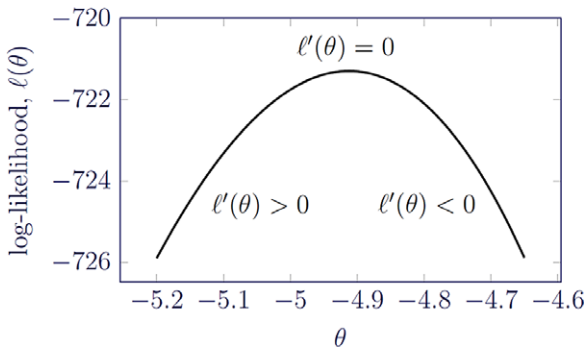
On the left, as the log likelihood is increasing, the first derivative is greater than zero. On the right, the first derivative is less than zero and at the peak the first derivative is zero; this is the point where we would find our maximum-likelihood estimate of the parameter of interest.

Our third assumption is that we have the Hessian matrix, H, which contains all second partial and cross partial derivatives. $\hat{\theta}$ is then the maximum-likelihood estimate of θ if $l'g(\hat{\theta}g) = 0$, and the Hessian matrix is negative and semi-definite. In the one-dimensional example above, $\hat{\theta} = -4.9$ and we can then evaluate the Hessian to get the standard error of the parameter estimate. This extends to multiple dimensions, i.e. $m > 1$.

The interesting thing about $\hat{\theta}$ is that, because it is uncertain, it can itself be viewed as a random variable. According to the maximum likelihood theorem, $\hat{\theta}$ has an approximate multivariate normal distribution with mean vector of $\hat{\theta}$ and covariance matrix $\hat{\Sigma}$, which is estimated as the negative inverse of the Hessian. This is a useful property of maximum likelihood estimates, and we can make considerable use of it for mis-estimation assessments. This multivariate normal assumption means that the log-likelihood has an essentially quadratic form, as demonstrated below:

Log-likelihood and quadratic approximation. Source: based on Richards (2016, Figure 1).
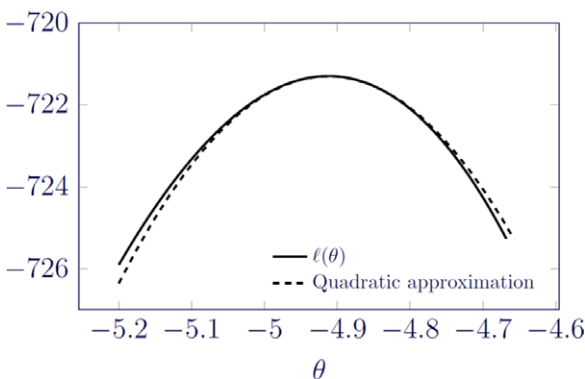
Here we see that, in this one-dimensional case, the solid black line is the actual likelihood function (which happens to be a Poisson likelihood in this example), and the dashed line shows the quadratic approximation that is inherent in the multivariate normal assumption for the

distribution of $\hat{\theta}$. We can see that the quadratic approximation is very close. In general, the multivariate normal assumption for the distribution of $\hat{\theta}$ is a good one for most models with statistically significant parameter estimates. This means that if $\hat{\theta}$ has a multivariate normal distribution, all of our estimation risk is summarised in $\hat{\Sigma}$. In particular, the leading diagonal of $\hat{\Sigma}$ is the variance of the individual components of $\hat{\theta}$, and all the off-diagonal entries of $\hat{\Sigma}$ contain the covariances of the various components of $\hat{\theta}$ with each other.

To summarise, estimation risk is about statistical parameter uncertainty, whereas mis-estimation risk is the financial impact of that parameter uncertainty. To put it more formally, we could say that the estimation risk is the uncertainty over, $\hat{\theta}$, whereas mis-estimation risk is the uncertainty over, $V(\hat{\theta})$, where V is our liability function. In the examples considered here, V is the reserve for the pensions in payment, but it could be any other actuarial liability.

There are some preconditions before we look at some mis-estimation results. Armstrong (2013, slide 13) asked "what assumptions are you making, e.g. independence? Duplicate policies? Amounts versus lives?" In UK annuity portfolios, people tend to have multiple annuity policies, and indeed, wealthier people are more likely they are to have multiple policies (Richards & Currie, 2009, Figure 1 and Table 1). Then comes the question of amounts- or lives-based mortality. Statistical models are inherently lives-based, but actuaries are very attuned to the concentration risk that comes with amounts. A number of preconditions therefore apply for any mis-estimation risk assessment. The first is that you need to de-duplicate your records; you need to turn a data set of policies into a data set of lives by identifying people with multiple benefits and aggregating them; see Chapter 2 of Macdonald *et al.* (2018). We will use a lives-based statistical model in order to get valid statistical estimates with covariance matrices, but we cannot forget the "amounts effect" of mortality, i.e. both concentration risk and the tendency for wealthier people to have lower mortality rates.

We can handle the amounts effect as a categorical factor, as we did earlier with the various pension deciles, or alternatively we can model mortality that varies continuously with the exact pension size; see Richards (2021b). In particular, it is important to note that any assessment of mis-estimation capital will lead to an underestimate if (i) you fail to de-duplicate the records before the analysis, (ii) you ignore the amounts effect on mortality or (iii) your model does not acknowledge any time trend.

Let us next look at what we will call run-off mis-estimation risk before we come to the value-at-risk approach, and then we can contrast the two. Run-off mis-estimation risk could be formulated as the question "What is the uncertainty over V caused by the uncertainty over $\hat{\theta}$?" We have two potential approaches to addressing the question of the uncertainty over our reserve. We will first look at the delta method. We said earlier that $\hat{\theta}$ can be viewed as a random variable, and that V, our reserve function, is a function of a random variable. We can apply the delta method for functions of a random variable as follows:

$$\mathrm{Var}\left[V\left(\hat{\theta}\right)\right] \approx a^T \hat{\Sigma} a$$

where $\hat{\Sigma}$ is the covariance matrix vector and a is the first partial derivative of the reserve with respect to, $\hat{\theta}$, i.e. $a = \frac{\partial V(\hat{\theta})}{\partial \hat{\theta}}$ a has m elements and each element is the first partial derivative of the reserve with respect to the corresponding member of $\hat{\theta}$. We can estimate a using central differences of the reserve.

An alternative to the delta method is to generate a set, S of liability valuations subject to parameter risk. Then we can calculate percentiles of, S, for example a 99.9th percentile, the median or the average. To generate parameter variation, we perturb $\hat{\theta}$ consistent with the estimated covariance matrix. We use the maximum likelihood theorem, and we generate an alternative parameter vector, θ' by sampling from the multivariate normal distribution using Monte Carlo simulation. Each time we sample an alternative parameter vector, we calculate our liability function and then

add it to our liability set, S. This sampling approach to parameter risk is illustrated in the following flowchart:
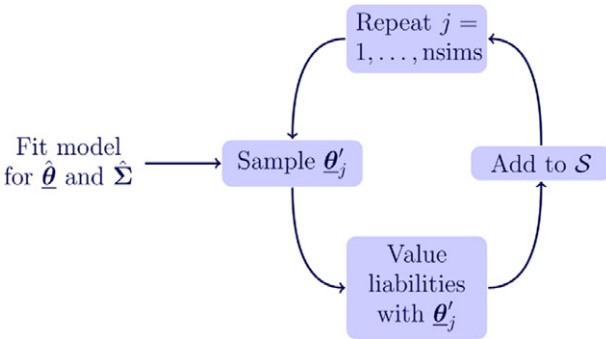
For the sampling method, we start off by fitting one model to estimate $\hat{\theta}$ and $\hat{\Sigma}$. Using the multivariate normal distribution, we then sample, θ'_j value the liabilities with, θ'_j add the resulting liability value to, S and then repeat, say, 10,000 times. This gives us a set, S of liability values subject to parameter uncertainty caused by the sampling procedure from the multivariate normal distribution.

We will now look at some results for these two alternative approaches applied to a large pension scheme with 44,616 lives. As with the medium-sized pension scheme earlier, the data are for pensioners only and here we have 10,663 deaths observed in 2001–2009. This time the model is richer in the number of factors considered; see Table 4 of the main body of the paper. The distribution of, S our sampled set of valuations placed on the pensions in payment, as shown in Figure 4 of the main body of the paper. The best estimate of the liability value is about £2.03 billion, but there is a fair degree of uncertainty caused by the uncertainty of the parameter estimates (we have gender, pension size, early retirement and other risk factors included). Figure 4 shows that the uncertainty over some of these parameter estimates causes a reasonably wide spread of up to £100 million in the reserve estimate. It is important to note that this is a mis-estimation assessment only, so it only considers uncertainty over current rates of mortality. If we were to put in uncertainty over future improvements, for example, then the reserve values would increase and would also be more spread out as there would be greater overall uncertainty.

One question of interest to insurers in particular is what relative capital percentage (RCP) is required to cover a proportion, p of mis-estimation risk? We denote this, RCP_p and it is the additional capital expressed as a percentage of the best-estimate reserve. We can contrast the delta method and sampling approach to run-off mis-estimation, and they each have different ways of calculating this relative capital percentage. Under the delta method, we have the following:

$$RCP_p = \Phi^{-1}(p) \frac{\sqrt{a^T \hat{\Sigma} a}}{V(\hat{\theta})} \times 100\%$$

where Φ^{-1} is the inverse of the normal distribution function and $\frac{a^T \hat{\Sigma} a}{V(\hat{\theta})}$ is the coefficient of variation of V. Under the sampling approach, we use the appropriate percentile of S:

$$RCP_p = \left[\frac{Q_p(S)}{V(\hat{\theta})} - 1 \right] \times 100\%$$

where $Q_p(S)$ is the p-quantile of S. Below is a plot of the two alternatives for quantiles 90–99.5%:

Quantiles for run-off mis-estimation capital.
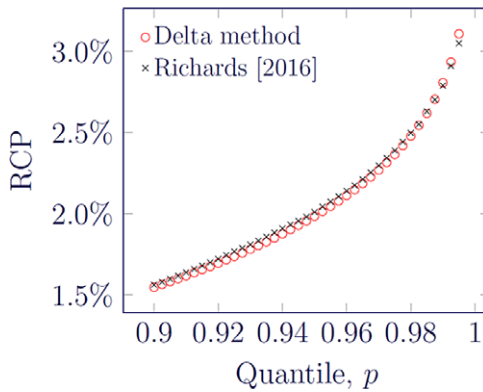
We can see that there is a high degree of agreement between the two approaches. There are slightly higher capital requirements for p levels in the range 90–98% under the sampling approach, but by and large it is a close approximation. Even quite far into the tail, at the 99.5% level, we have still got reasonably good agreement, despite capital requirements increasing semi-exponentially as the quantile increases. For run-off mis-estimation, both the delta method and the set-based sampling method agree quite closely for this particular liability, even quite far into the upper tail. The delta method is quicker, but the set-based approach is better if the liability values are skewed. Thus, if liability values are symmetric around a single peak, then you could use the delta method. However, if liability values are skewed in response to parameter uncertainty, then the sampling approach would be better.

What we have discussed above is the run-off approach to mis-estimation risk, and we now need to contrast it with a value-at-risk approach, such as might be required for Solvency II. The run-off mis-estimation question is essentially "What is the uncertainty over my reserve, caused by the uncertainty over my parameter estimates?" In contrast, the value-at-risk approach answers a different question, namely "What is the uncertainty over my reserve that could be caused by re-estimating parameters, based on n years of additional data?" This is a quite different question, and it is essentially about recalibration risk (Cairns, 2013). Although both are billed as mis-estimation questions, they are asking different things, so we expect them to produce different answers.

An important point to note about the run-off approach is that it has no "new experience" element. From the mortality experience you estimate the parameters, and then you can estimate the uncertainty over your reserve. However, we need to make some changes to the previous flow-chart to make it a proper value-at-risk approach appropriate for the likes of Solvency II. We therefore make two changes to the sampling approach of Richards (2016). The first change is that we need to simulate n years of experience, and the second is that we need to refit our model to the combined real data and simulated data. We can then proceed as before to generate our set, S, of liability values. The flowchart of the value-at-risk approach to mis-estimation risk is then as follows:
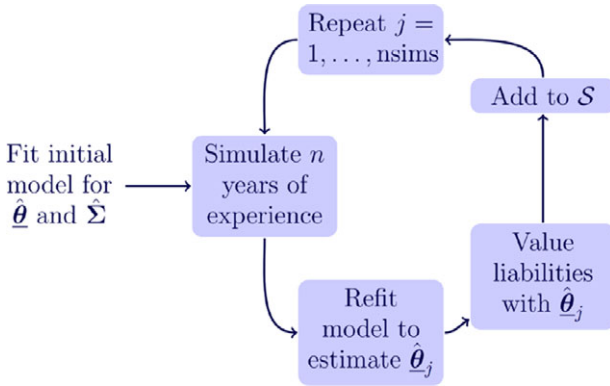
As in the run-off flowchart, we fit an initial model for $\hat{\theta}$ and $\hat{\Sigma}$. We then repeatedly simulate n years of further mortality experience amongst the survivors. We refit the model to the combined real data and simulated experience and we re-estimate $\hat{\theta}_j$. We then use $\hat{\theta}_j$ to value the liabilities as before, adding each new value to S. This produces a set of valuations subject to the recalibration risk from n years of new experience data. This gives us a true n-year, value-at-risk approach to mis-estimation risk. We can calculate the percentiles of S as before, and we can contrast the run-off and value-at-risk approaches:

Distribution of liability values under run-off and VaR approaches to mis-estimation risk. Source: Figures 4 and 6 in main body of paper.
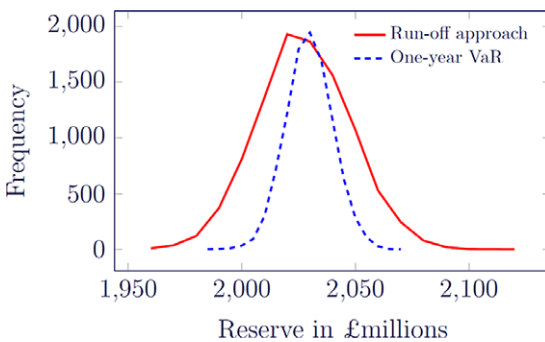
The 1-year, value-at-risk approach to the same portfolio produces a similar peak in liability estimates, but has a much narrower spread. The run-off approach looks at parameter risk over the entire term of the liabilities, but the 1-year value-at-risk approach is driven by the recalibration risk over a 1-year horizon.

In this example, we used a survival model with a force of mortality varying with age. The additional risk factors in the model are age and gender, normal versus early retirement, first life versus surviving spouses, pension size and a time trend. It is important to include a time-varying parameter to allow for mortality improvements, otherwise the mis-estimation uncertainty will be under-estimated.

The 99.5% value-at-risk capital requirement is the reserve that would cover 99.5% of all 1-year recalibrations. In the examples above, we calculated the 99.5th percentile of S and then divided it by the mean of S. However, it does not make any difference if you use the mean or median of

S as the denominator, as each produces the same result for this portfolio. A comparison of various choices for the denominator is provided in section 7 of the main body of the paper.

There are two options when simulating the new mortality experience for the value-at-risk approach to mis-estimation risk. We can ignore parameter risk and use $\hat{\theta}$ each time we simulate the experience and recalibrate the model. Or we can include parameter risk by first sampling from the multivariate normal distribution, then simulating the new mortality experience, and then carrying out the recalibration. This choice makes a difference, as shown in Figure 5 of the main body of the paper, which shows the capital requirements as a percentage of the best-estimate reserve by various value-at-risk horizons. If we simulate and recalibrate with parameter risk, we have steadily increasing capital requirements with the value-at-risk horizon, n. However, if we simulate and recalibrate without parameter risk, we get capital requirements that are broadly constant. Of particular interest is the 1-year mis-estimation capital requirement in Figure 5, where there is surprisingly little difference from including or excluding parameter risk. Although run-off mis-estimation risk is driven solely by parameter risk, for the value-at-risk approach most of the 1-year capital requirement is not driven by parameter uncertainty at all. Here, most of what is called mis-estimation risk is actually driven by the idiosyncratic risk underlying the recalibration. While we think of mis-estimation risk as being primarily driven by parameter uncertainty, for short horizons the value-at-risk capital is largely driven by recalibration risk alone. Adding parameter risk will obviously increase mis-estimation capital requirements, but, at least in this example, most of the 1-year capital requirement is not due to parameter risk at all. In fact, Figure 5 shows that, even at a 5-year horizon, only about half of the 5-year mis-estimation capital is actually driven by parameter uncertainty. This means that the result for a 1-year solvency approach is qualitatively different from the longer horizons that one would use for, say, an ORSA assessment. The fact that most of the 1-year mis-estimation capital requirement is not driven by parameter risk prompts the question: how correlated is value-at-risk mis-estimation capital with the idiosyncratic risk?

Most insurers would hold more than the minimum two capital sub-elements for longevity risk. In addition to mis-estimation risk and longevity trend risk, they would usually have an adverse deviation element, also known as idiosyncratic risk. But how correlated is mis-estimation risk with idiosyncratic risk? Section 12 of the main body of the paper looks at three different possible metrics that one might use to describe idiosyncratic risk. For each horizon period, n, Table 9 shows the correlation of the reserves with (i) the time lived by the survivors over the period, (ii) the new deaths occurring in the period and (iii) the annuity payments made to the survivors over the period. Regardless of which metric we pick for idiosyncratic risk, there is a high degree of correlation with the reserve. The value-at-risk mis-estimation capital is therefore strongly correlated with the idiosyncratic risk over a 1-year horizon, and this must be reflected in any aggregation matrix used for Solvency II.

Let us additionally look at the impact of features like the discount rate. We can calculate liability values with different discount rates, and, as shown in Figure 9 in the main body of the paper, capital requirements rise as the discount rate falls.

Another interesting aspect is the choice of mortality law used in the model, of which the Gompertz (1825) model is the oldest and simplest. However, there are alternatives, as shown in equations (10)–(12), in the main body of the paper. The Perks (1932) model in equation (10) is a variant of the Gompertz model with the same number of parameters; it is largely linear on a log scale between ages 60 and 90, but it features late-life mortality deceleration. The Beard (1959) model in equation (11) is similar but has more flexibility at the oldest ages, while equation (12) features a Makeham term to allow for the fact that mortality is not (log−)linear between ages 50 and 60. We also have a relatively recent addition to the model pantheon, namely the Hermite-spline model (Richards, 2020).

The Hermite-spline model is designed specifically for post-retirement mortality, and it is different from other models because it allows for the automatic convergence of mortality

differentials with age. For example, if you have a differential for pension size at age 60, then it usually narrows with increasing age to the point where it largely vanishes by age 95. We use Hermite splines to give us smoothness and flexibility, but also for their ability to provide this mortality convergence. Figure 10 in the main body of the paper shows a basis of four Hermite splines, of which the most important are the h_{00} and h_{01} splines. To turn Hermite splines into a mortality model we define x_0 and x_1 as our minimum and maximum ages, respectively. We then map the age range onto the interval (0,1). The logarithm of the force of mortality is then just a linear combination of these Hermite splines, as shown in equation (13), where just three of the splines are used. As with the other models in equations (10)–(12), α is the level of log(mortality) at the youngest age and ω is the limiting rate of log(mortality) at the oldest age. We can vary α with different risk factors, but in equation (13) the impact of this will decrease with age because of the shape of the h_{00} spline in Figure 10. We can estimate the parameters in equation (13) from the data, just as we do with the other models in equations (10)–(12). The key difference for the Hermite-spline model is that the impact of α reduces automatically with age, so we do not need the interactions with the β term that are required for the models in equations (10)–(12). As a result, a Hermite-spline model usually needs fewer parameters.

Which models fit best? And does having fewer parameters in the model mean less mis-estimation risk? We have (at least) two ways of looking at model fit. There is the statistician's view, which would be a likelihood-based approach, such using an information criterion like the AIC (Akaike, 1987). But as actuaries we know the crucial importance of amounts-related mortality, so we also need to look at, for example, a pension-weighted resampling approach like the procedure described in section 6.7 of Macdonald *et al.* (2018).

What counts as the best fit? In terms of lives, the best fit is usually the model with the lowest information criterion. For our test of fit by amounts, having a percentage as close as possible to 100% would be best. Table 7 in the main body of the paper lists the five models sorted by descending AIC, together with the number of parameters estimated. The Hermite-spline model has the fewest parameters because it does not need the age interactions. The other models need interactions with β to achieve narrowing mortality differentials with age. In terms of the information criterion, the Gompertz model is by far the poorest fit, while the Hermite-spline model is the best; the Makeham–Perks model is a close second. In terms of the bootstrapping percentage, the Gompertz model is the furthest away from 100%, while the Perks and Hermite-spline models are the closest to 100%. The Gompertz model therefore has the worst fit in terms of both lives and amounts. In contrast, the Hermite-spline model has the fewest parameters, the best fit in terms of the information criterion, and the second-best fit in terms of explaining amount-based variation. A natural next question is what impact this all has on the mis-estimation capital requirements?

Figure 12 in the main body of the paper shows the mis-estimation capital requirements across various horizons using the different models. The Gompertz and Hermite-spline models, having the fewest parameters, have some of the lowest capital requirements, with the Hermite-spline model having the lowest capital requirements of all. It appears that models with fewer parameters produce lower mis-estimation capital requirements. However, the poor fit of the Gompertz model does not justify its use in a mis-estimation assessment.

In summary, we have a concept of mis-estimation risk, which stems from having a limited quantity of experience data and thus some uncertainty over current mortality rates. Quantification of mis-estimation risk needs to be statistical to get proper standard errors for our estimates and to account for correlations between parameters. The impact of mis-estimation risk must be quantified financially to account for concentration risk. We have two alternative views of mis-estimation risk: (i) the run-off approach, which is useful for pricing block transactions like longevity swaps, bulk annuities, reinsurance treaties and any large transactions involving a large number of policies or lives and (ii) the value-at-risk approach, which would be suitable for Solvency II reporting and ORSA work. For the value-at-risk mis-estimation approach, we have seen that parameter risk is a surprisingly minor driver of 1-year capital requirements.

Instead, most of the 1-year, mis-estimation capital requirements are, in fact, driven by recalibration risk. As a result, there is a strong correlation between 1-year mis-estimation VaR capital and idiosyncratic risk.

Before concluding, I'd like to thank to Gavin Ritchie of Longevitas, Patrick Kelliher of Crystal Rock Consulting and Professor Andrew Cairns of Heriot-Watt University for helpful comments on some earlier drafts of this presentation.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrica*, **52**: 317–333. ISSN: 0033–3123. doi: 10.1007/BF02294359.

Armstrong, K. (2013). Calibrating longevity stresses: a practical life office perspective, presentation, available at https://www.actuaries.org.uk/documents/rising-challenge-generating-longevity-stresses

Beard, R. E. (1959). Note on some mathematical mortality models, in *The Lifespan of Animals* (ed. G.E.W. Wolstenholme and M. O'Connor), Boston, Little Brown, 302–311.

Burgess, S., Kingdom, J. & Ayton, J. (2010). Longevity risk: issues to be considered, presentation, available at www.actuaries.org.uk/system/files/documents/pdf/burgesskingdomayton.pdf

Cairns, A.J.G. (2013). Robust hedging of longevity risk. *Journal of Risk and Insurance*, **80**, 621–648. doi: 10.1111/j.1539-6975.2013.01525.x.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies, *Philosophical Transactions of the Royal Society of London*, **115**(1825), 513–583.

Macdonald, A.S, Richards, S.J. & Currie, I.D. (2018). *Modelling Mortality with Actuarial Applications*, Cambridge, Cambridge University Press. ISBN: 978-1-107-04541-5.

Makin, S. (2008). Response to discussion paper on actuarial mortality assumptions.

Perks, W. (1932). On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries*, **63**, 12–40.

Richards, S.J. & Currie, I.D. (2009). Longevity risk and annuity pricing with the Lee-Carter model. *British Actuarial Journal*, **15**(2), 317–343. doi: 10.1017/S1357321700005675.

Richards, S.J. (2016). Mis-estimation risk: Measurement and impact. *British Actuarial Journal*, **21**(part 3), 429–457. doi: 10.1017/S1357321716000040.

Richards, S.J. (2020). A Hermite-spline model of post-retirement mortality. *Scandinavian Actuarial Journal*, **2**, 110–127. doi: 10.1080/03461238.2019.1642239.

Richards, S.J. (2021b). *Modelling mortality by continuous benefit amount*. Longevitas Ltd, available at www.longevitas.co.uk.

Woods, S. (2016). *Reflections on the 2015 Solvency II Internal Model Approval Process*. Bank of England Prudential Regulation Authority, available at http://www.bankofengland.co.uk/pra/Pages/publications/ps/2016/ps3316.aspx.

The Moderator: Thank you, Stephen (Richards). That was a really interesting presentation.

Questions and Responses from the Webinar: *the following questions were raised during and immediately after the webinar.*

Questioner: A question on the first approach you presented, the run-off approach. There are two ways you look at modelling it. You have got the delta approach and the sampling approach, whereas for the VaR approach, we are only looking at the sampling. Is there a way we could adapt the delta approach to be used on the VaR methodology?

Dr Richards: No, not that I can think of. The difference lies in the recalibration aspect, i.e. the simulation of experience and the model refitting. I cannot think of a way whereby you could take advantage of, say, the delta method for that. I think you just have to do this through simulation and recalibration, unfortunately.

Ms S O'Sullivan F.I.A.: We said earlier that the parameter distribution was a Poisson but yet close enough to the normal maximum likelihood distribution. So, how do we assess that it was a Poisson distribution and, subsequently, how would any of the results change if we used that real distribution rather than assuming that it was normal? Can we just substitute in a Poisson?

Dr Richards: There is no Poisson assumption anywhere in the mis-estimation methodology. In Richards (2016, Figure 1), I illustrated the general quadratic shape of a log-likelihood function and I happened to pick a Poisson random variable.

Questioner: Why are you doing the refitting of the parameter at all in the VaR approach? Are you not fitting a parameter to your already known uncertainty in the parameter?

Dr Richards: The model refitting answers the value at risk question: "By how much could my liabilities change in the following year from new knowledge from an additional year's experience?" In order to do that, we need to fit the model

to the new experience in that year. It is unavoidable that the value-at-risk approach involves the refitting of the model because, in essence, that is what the VaR question is "By how much can my liabilities change from refitting the model?"

Questioner: Is the simulation shown under the recalibration approach including a stochastic period effect?

Dr Richards: There is no stochastic period effect. We simulate next year's experience based purely on the parameters in the model, which cover age, gender, pension size and other differentials. You would include a stochastic period effect if you were calculating, say, an adverse experience effect, so that might be done separately. The results in the paper are purely based on the recalibration effect.

Questioner: The VaR calculation we have seen in the example in this paper is based on a relatively modestly sized book of lives. If we had a much larger book of lives, as a large insurance company might have, how do you think the results would look, particularly the parameter part and the recalibration part?

Dr Richards: This is covered in section 9 of the paper. I took the portfolio and I duplicated it ten times to make the size closer to what a large insurer would have. It was still the same experience with the same risk factors and, when you fit the model, you get the same parameter estimates but with much smaller standard errors. In Figure 8 of the paper, you can see that there are considerably lower mis-estimation capital requirements for a larger portfolio.

Questioner: Don't you think a 1-year VaR horizon is insufficient given that policyholders' reasonable expectations are that the company will pay their claims when they fall due, which may be many years hence?'

Dr Richards: That is an interesting question because in annuity business the risks are largely in run-off. It is very unlikely that an annuity portfolio will be unable to pay policyholders over a 1- or even a 10-year time horizon. Arguably, longevity risk and annuity business are a poor fit for the value-at-risk regime because most of the risks are long-term, run-off risks. This is in contrast to, say, term assurance, where a pandemic could potentially drive a large spike in claims that could lead to the inability to pay. Annuity business is an example of a class of insurance business that is not a good fit for this 1-year, value-at-risk approach. But that is the way the regulations are framed, so that is the way that we need to calculate our capital and present our results.

Professor A Cairns F.F.A.: The simulation approach assumes multivariate normality. Suppose you do n simulation scenarios under this assumption and then calculate for each simulated θ. Could you calculate the true likelihood for each simulated θ and compare this to the multivariate likelihood?

Dr Richards: Yes, you could. I do this in the paper when I look at the shape of the profile log-likelihood. It is important that profile likelihoods are quadratic for the multivariate normal assumption to be valid. This is worth checking when you are fitting the model and carrying out these calculations.

Questioner: The quadratic approximation might be very good around the parameter estimate, but would it be less good as we get further away from the estimate?

Dr Richards: Figure 1 in the paper shows that the profile log-likelihood function is usually similar to the quadratic shape of a normal log-likelihood. As long as we are dealing with a modest range, say a 95% confidence interval, there is next to no difference between the quadratic approximation and the log-likelihood function's actual shape. However, as you get further away from the central estimate, the closeness of the approximation becomes poorer. If you are looking at really extreme parameter deviations, then the quadratic approximation behind the multivariate normal assumption might be less good.

Ms S O'Sullivan: What if the parameter distribution is not multivariate normal in reality?

Dr Richards: If that was the case, then I think you would need to refit your model. It does happen with some parameters that you do not get a quadratic shape and there is an example for the Makeham–Perks model in Table 7 of the paper, where one of the parameters does not have the proper quadratic shape. However, Table 8 suggests that this does not seem to have had any material impact on the mis-estimation assessment because the results are consistent across all the models.