DIFFICULT RISKS AND CAPITAL MODELS

A preliminary report from the Extreme Events Working Party

R. Frankland, L. Hewitt, P. Jakhria, A. D. Smith, S. Sharp, J Sharpe, and others.

1. Introduction

A financial firm assesses that €10bn is sufficient to absorb losses with 99.5% probability. By the end of the year, the firm has lost €20bn. Many things have gone wrong at once, including the following:

- A large fall in equity markets. Although equity returns had been modelled using a distribution fitted to historic data, that distribution is no longer compatible with the data once the latest year's observation is included, and so the original model is now considered inappropriate.

- A complex change in yield curve shape, where very short and medium term rates rose, while short and long rates fell. The yield curve model was based on three factors, and the firm had taken care to hedge these. However, the actual outcome exposed a potential for loss that had not been captured within the scenario tests on which the liability forecast distribution was based.

- Some new risks had emerged that had not previously been modelled explicitly. Specifically, a loss on derivative positions arose from the widening of spreads between swaps based on LIBOR and overnight index swaps. A euro government defaulted on its bonds, and the annuity portfolio took a one-off hit because of a mis-estimation of proportion married.

- A change in the shape of the credit spread curve meant that the existing market consistent ESG could not calibrate the year end market conditions exactly. A new ESG from a different provider was put in place, but for reasons that are still unclear, this led to a doubling of the stated time value of liability options.

- Early in the year, the firm participated in some securitised AAA investments which exploited market anomalies to provide yields closer to those on junk bonds. The capital model had been based on the portfolio prior to this participation. The participation turned out to be disastrous, and virtually all the investment had to be written off following an avalanche of unforeseen defaults in the underlying assets.

- A problem arose with a set of policies that, for calculation convenience, had been modelled together. It turned out that within a group of policies assumed to be homogenous, there was in fact a variety of investment choices; as a result, a subset of these policies had guarantees that came into the money. Policyholders selectively took advantage of these guarantees, producing costs far beyond those projected from the capital model.

- Losses from a March earthquake in the Middle East were substantially greater than the maximum possible loss derived from a third party expert model. The reasons for this seem to be a combination of a disproportionate number of January policy inceptions (the model assumed uniformity over the year), inadequate coverage of the

DRAFT

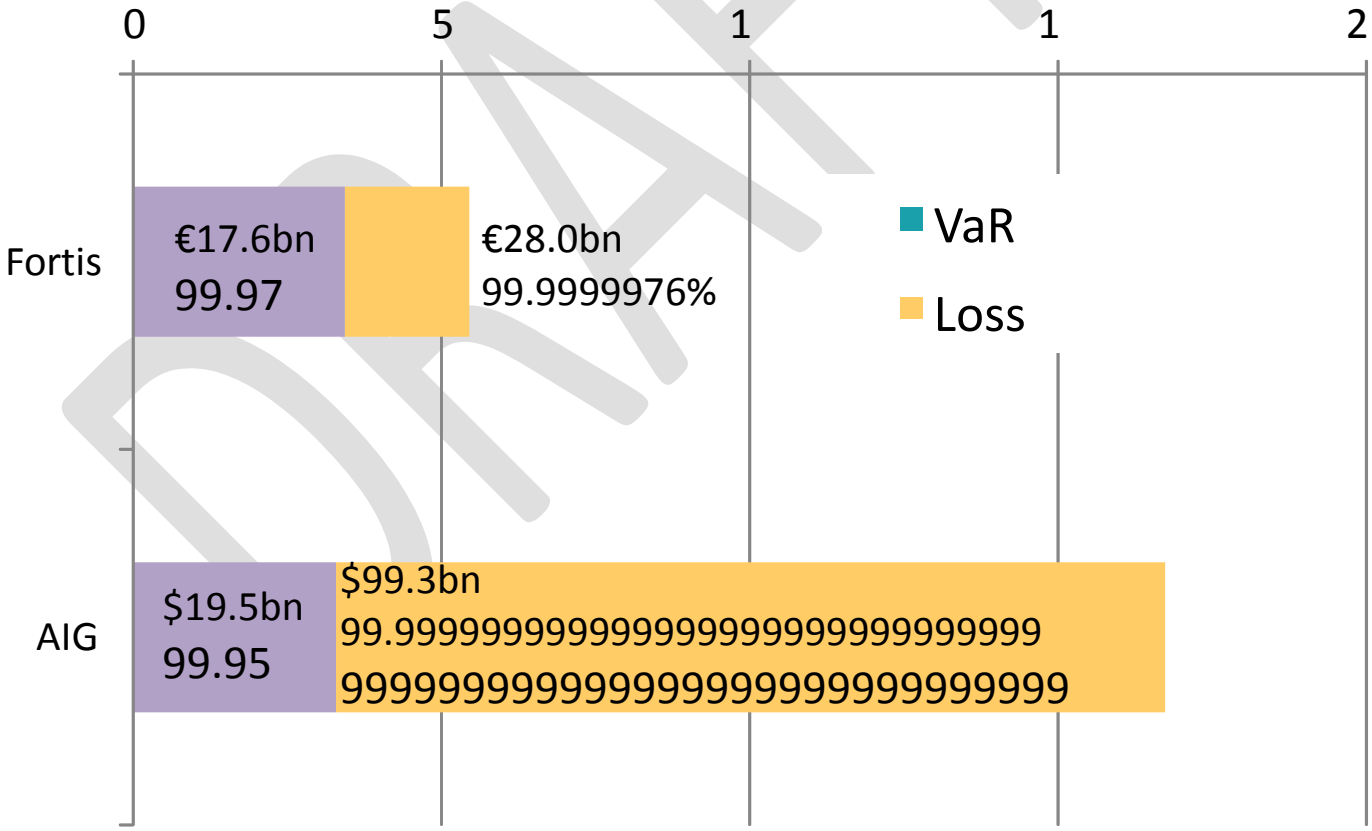
region in the external catastrophe model, and poor claims management exacerbated by political instability and corruption.

- Many life insurance claims became due following an industrial accident. The insurer was confident that a proportion would be recovered from reinsurers as assumed within the capital model. However, the reinsurance contract contained a clause not captured in the capital model, which limited payouts in the event of large losses from a single event.

- A clarification in the regulatory calculation of the illiquidity premium meant that the insurer could no longer use the yield on certain illiquid assets to discount the liabilities. The liabilities rose as a result of this change, leading to an accounting loss. At the same time, a change in tax rules meant that a deferred tax asset previously considered recoverable had to be written off.

Although our example is hypothetical, there are real examples of firms who proclaimed they held economic capital to withstand a loss equivalent to a 1-in-2000 year event (or rarer) before losing a multiple of that amount and being rescued in state bail-outs.

The chart below show the published economic capital figures for AIG and Fortis, with the actual loss in 2008. We have also calculated the confidence level associated with those losses, assuming normal distributions.



It is theoretically possible that such an outcome was a case of exceptionally bad luck, but with hindsight we have seen how risks that ultimately proved to be important were either overlooked or deliberately scoped out of the capital models.

DRAFT

Capital Model Scope

Our thought experiment highlights many possible sources of loss. Some of these may be captured within a stochastic internal model, but many will not. While perhaps more of the risks could be modelled stochastically than is currently common practice, other losses are attributed to lack of knowledge rather than anything explicitly stochastic. It is debatable whether probability theory is the right tool to address such risks.

If the non-stochastic risks are simply disregarded, then firms are likely to see frequent exceptions, that is, experienced losses worse than the previously claimed 1-in-200 event. Mounting evidence of such exceptions could undermine firm's claims of financial strength, and call into question the advertised degree of protection (for example, 1-in-200) that a supervisory regime offers. Furthermore, the firm's internal models may themselves be discredited in the eyes of the market, likely leading to significant loss of investor confidence.

It is therefore desirable that risks are identified and some attempt is made at quantification, even if a variety of techniques are needed to address different elements. A less satisfactory solution is to assert that some risks are out of scope. Then, the case has to be made for not including the corresponding losses towards the exception count. This case is more persuasive if the out-scoping has been signalled in advance of the losses occurring.

The Remainder of this Paper

In the remainder of this paper, we focus on three specific types of error

- Error in statistically estimated model structures and parameters
- Errors due to human judgement
- Fitting error in liability models

DRAFT

2. Error in Statistically Estimated Models

With a correct model and, accurate assumptions, we can justify statements such as "with €100m of available capital, there is a 99.5% probability of sufficiency one year from now". This paper considers how such a statement may be modified if a firms has concerns about the correctness of models or accuracy of parameter estimates

We examine three special cases, as follows:

- Case G (Green): Models and parameters are known to be correct
- Case A (Amber): Past and future observations are samples from a given model, for example normal distributions, but the model parameters are uncertain.
- Case R (Red): Both the applicable model and the parameters are uncertain. For example, there may be some dependence between observations and the observations may be drawn from one of a family of fatter tailed distributions. In each case, limited data is available to test the model or fit the parameters.

Possible Statements: Case A

In case A (amber), with uncertain parameters, we can still use the concept of a "true" model, with an associated "true" 99.5%-ile, but we do not know what that true model is. We take as an example the prediction of a future observation from a normal distribution, given a set of past observations from the same distribution.

We can make statements as follows:

A1. We estimated the parameters using the [method of moments]. If these estimates are exact then €100m of capital is 99.5% certain to be sufficient. This calculation ignores the possibility of parameter estimation error.

A2. We estimated the 99.5%-ile using the [method of moments]. If this method had been applied on many alternative historic scenarios, then on average the estimated 99.5%-ile is equal to the true 99.5%-ile (this is Fisher's concept of an unbiased estimate).

A3. As A2, but in addition, our chosen method produces an estimate whose variance is lower than other methods (Fisher's concept of an efficient estimate)

A4. We estimated the 99.5%-ile using the [method of moments]. We used the finite data available, but if we had unlimited data then our method would produce estimates that converge to the true 99.5%-ile (Fisher's concept of a consistent estimate)

A5. We estimated upper and lower bounds for the 99.5%-ile using [chosen method]. In a large number of trials of alternative historic scenarios, this interval contains the true 99.5%-ile in 95% of the time, regardless of what the true parameters are. (a classical 95% confidence interval for the 99.5%-ile)

A6. We generated values for the parameters according to a prior distribution, and generated historic scenarios. We kept only those combined scenarios where the simulated data was close to our own history. Out of those scenarios, the average true 99.5%-ile was €100m and the most likely 99.5%-ile was €90m. (Bayesian mean or modal prediction)

A7. We generated values for the parameters according to a prior distribution, and generated historic scenarios. We kept only those combined scenarios where the simulated data was close to our own history. Restricting attention to those close historic outcomes, the true 99.5%-ile was between €65m and €150m for 95% of cases. (Bayesian confidence interval)

DRAFT

A8. We estimated the 99.5%ile using [chosen method]. Using our method, and regenerating both past and future data, our estimated percentile exceeds the next observation 99.5% of the time, regardless of the true parameters. (Geissner's prediction interval)

A9. As A8, but in addition our chosen method, on average, produces lower estimated percentiles than other methods. (efficient prediction interval)

A10. We generated values for the parameters according to a prior distribution, and generated linked historic and future scenarios. In each scenario we used [chosen method] to estimate the 99.5%-ile of the next observation. We kept only those combined scenarios where the simulated data was close to our own history, and out of those scenarios, our estimated percentile exceeds the next observation 99.5% of the time. [Bayesian prediction interval].

We note the following about these different statements:

- We believe that market practice is currently closest to A1, in that a single, validated, model is used for estimating capital.

- While A1 is a statement of the process that has been followed, statements A2 to A10 describe properties of the outcome. The claimed properties can be checked, for example, by Monte Carlo, and robustness to model mis-specification investigated.

- Statements embedded in existing legislation (such as Solvency 2 directive or Basel 3) make sense under case G (Green) of model and parameter certainty. All the statements A1-A10 are possible generalisations to case A (amber). Thus, statements A1-A10 are all answers to different question – the appropriate answer requires an interpretation of which question is being asked.

- When talking about probabilities, care needs to be taken in respect to the set over which averages are calculated. All statements from A1-A10 average over possible future outcomes. The Bayesian statements A6, A7 and A10 average over possible alternative parameter sets but not over counter-factual past data sets. The frequentist statements allow mixing of the given past data with alternative counter-factual histories to construct probabilities, while requiring that the resulting inference is uniformly valid across all possible parameters.

- The Bayesian methods appear to require a prior distribution, and so require more judgement than the other (frequentist) methods. However, all of this is in the unrealistic context of a known model family.

- One limiting case of the Bayesian framework is the uninformative prior. This produces the same numerical output as the classical intervals for parameters (confidence intervals) or the next observation (prediction intervals). However, the corresponding statements remain different, and in the Bayesian case have now become untestable as a test would require generation of random parameters from a non-integrable density. This does not mean that the Bayesian and frequentist approaches agree, but rather we have found two problems whose solutions coincide in this special case of independent normal observations.

DRAFT

Possible Statements: Case R (Red)

We now consider the case where both the model and the parameters are uncertain. We suppose a set of possible models has been constructed. We call this the “robustness set”. We can expect that this set could be very large, consisting of a wide variety of different model types and parameters. However, there may be some restrictions, for example on how fat the tails can be, or to exclude “hopeless” cases where the future observation is drawn from a distribution unconnected with the past. The robustness set is a judgementally specified range of models under which we might reasonably require statistical techniques to work. In some applications, the robustness set may be infinite dimensional (for example, allowing random samples from any distribution whose kurtosis does not exceed 4). In other cases, it might be a union of several finite dimensional sets (for example, several model families each of which has four parameters).

In Case R (Red) we see a greater divergence between the Bayesian and classical approaches.

The Bayesian approaches require a prior distribution over the robustness class. This may be more difficult to specify than a prior parameter distribution; in particular, specification as a probability density makes sense only if the robustness class is finite dimensional. However, having chosen the prior distribution, the wordings are the same as in case A (amber).

For the frequentist approaches, a large robustness set may longer allow us to achieve exactly the required degree of confidence. Instead, we use bounds on the probability. In the red case, it is clear that both the Bayesian and frequentist approaches require a large degree of judgement, in the selection of the robustness set and (in the Bayesian case) the prior distribution.

Red statements corresponding to the amber cases are as follows (same numbering)

R1. We estimated the parameters using the [method of moments] and chose the model family within the robustness set based on [least Kolmogorov Smirnov difference]. If these estimates are exact then €100m of capital is 99.5% certain to be sufficient. This calculation ignores the possibility of parameter estimation error or model mis-specification error.

R2. We estimated the 99.5%-ile using the [method of moments] and chose the model family within the robustness set based on [least Kolmogorov Smirnov difference]. If this method had been applied on many alternative historic scenarios, then on average the estimated 99.5%-ile is equal to, or greater than, the true 99.5%-ile depending on which model applies from the robustness set.

R3. As A2, but in addition, other methods produce more extreme percentiles than our method for some models in the robustness set.

R4. We estimated the 99.5%-ile using the [method of moments] and chose the model family within the robustness set based on [least Kolmogorov Smirnov difference]. We used the finite data available, but if we had unlimited data then our method would produce estimates that converge to the true 99.5%-ile, for any model in the robustness set and any parameters.

R5. We estimated upper and lower bounds for the 99.5%-ile using [chosen method]. In a large number of trials of alternative historic scenarios, this interval contains the true 99.5%-ile at least 95% of the time, depending on the model from the robustness set. (a classical 95% confidence interval for the 99.5%-ile)

DRAFT

R6. We generated models according to a prior distribution over the robustness set, and generated historic scenarios. We kept only those combined scenarios where the simulated data was close to our own history. Out of those scenarios, the average true 99.5%-ile was €100m and the most likely 99.5%-ile was €90m.

R7. We generated models according to a prior distribution over the robustness set, and generated historic scenarios. We kept only those combined scenarios where the simulated data was close to our own history. Restricting attention to those close historic outcomes, the true 99.5%-ile was between €65m and €150m for 95% of cases.

R8. We estimated the 99.5%ile using [chosen method]. Using our method, and regenerating both past and future data, our estimated percentile exceeds the next observation at least 99.5% of the time, depending on the model from the robustness set.

R9. As A8, but in addition our chosen method, on average, produces lower estimated percentiles than other methods for specified models in the robustness set.

R10. We generated models according to a prior distribution over the robustness set, and generated linked historic and future scenarios. In each scenario we used [chosen method] to estimate the 99.5%-ile of the next observation. We kept only those combined scenarios where the simulated data was close to our own history, and out of those scenarios, our estimated percentile exceeds the next observation 99.5% of the time.

We now see the clear distinction between the Bayesian statements that work on average across parameters and models, compared to the frequentist statements that are exact for some of the robustness set and are prudent elsewhere. The numerical values from Bayesian and frequentist have diverged so we can no longer argue that choice between Bayesian and frequentist approaches is moot.

It may appear that frequentist approaches are inherently more prudent, as Bayesian approaches achieve the stated confidence level on average over the robustness set, while frequentists achieve at least the stated confidence level. However, this comparison supposes that the same robustness set would apply in each case. It might alternatively be reasonable to use a smaller robustness set for frequentist methods than for Bayesian methods, in order to equate the perceived strength of either method.

DRAFT

3. Errors in Judgement:

Manifestations of Judgment

Firstly, it is important to acknowledge that judgment is a necessary part of Actuarial modelling, and it is very difficult (in fact, one could argue impossible with the exception of pathological examples) to simply avoid the use of any human judgement.

A model is necessarily a simplified representation of the real world, and as such needs to be stripped down to its most relevant components, such that the representation is a useful analogy of the real world for the specific purpose that we have in mind. The process of stripping down to the bare useful components and 'calibrating' the resultant model (necessarily) has a large amount of judgement associated with it.

This judgement can manifest itself in various different ways. Broadly speaking, some of the ways we encounter judgement over the process of modelling (in the more limited context of Actuarial modelling) are:

- Choice of overall framework for the model
- Choosing individual parts of the model
- Choice of calibration methodology
- Choice of parameters, overriding certain parameters if necessary

One can imagine that these have (broadly) decreasing levels of significance to the end results. However, the industry appears to have the greatest focus on the final (and perhaps second to last) elements of parameter choice and calibration methodology, often to the detriment of the overall choice of framework and model fitting. For example, companies may focus most of the documentation and rationale of expert judgment in the final two categories, potentially at the expense of reduced oversight and attention paid to the substantial implied judgments involved in the first two categories.

DRAFT

Case Study – Manifestation within a life insurance context

The next few sections aim to illustrate the different broad areas where judgments are used. These are primarily in the context of a life insurance company, although the key results could easily generalise into different areas.

Choice of overall framework:

This is quite possibly the single most significant judgement to be made within a (capital) modelling context, although it is not always appreciated as such. Of course, the materiality of the different choices depends on the particular problem at hand. We try to illustrate using a very simple case study with two risks (described as two products):

- Each individual risk to be akin to a simple product with a guaranteed £100m liability. Not all the risks are hedge-able, so there is a residual 1 in 200 risk that the assets (and capital) would lose half their value. Thus the extra capital required at time 0 such that the product has a 99.5% chance of meeting its guarantees at time 1 is (an extra) £200m (£100m for each product).
- The two products are assumed to be uncorrelated

Let us now consider two common methods of calculating the joint capital requirement:

1. Using an "external correlation matrix" approach, the answer is relatively simple. The joint capital requirement can be calculated as $\sqrt{Capital_A^2 + Capital_B^2}$, which is £141.6m in total and £70.8m per product.

2. Use an alternative approach of simulating the two products such that the risk of loss is lognormal (again, a popular choice amongst practitioners) with the same 1 in 200 probabilities on an individual basis. This produces a different number of £121m in total and £60.5m per product!

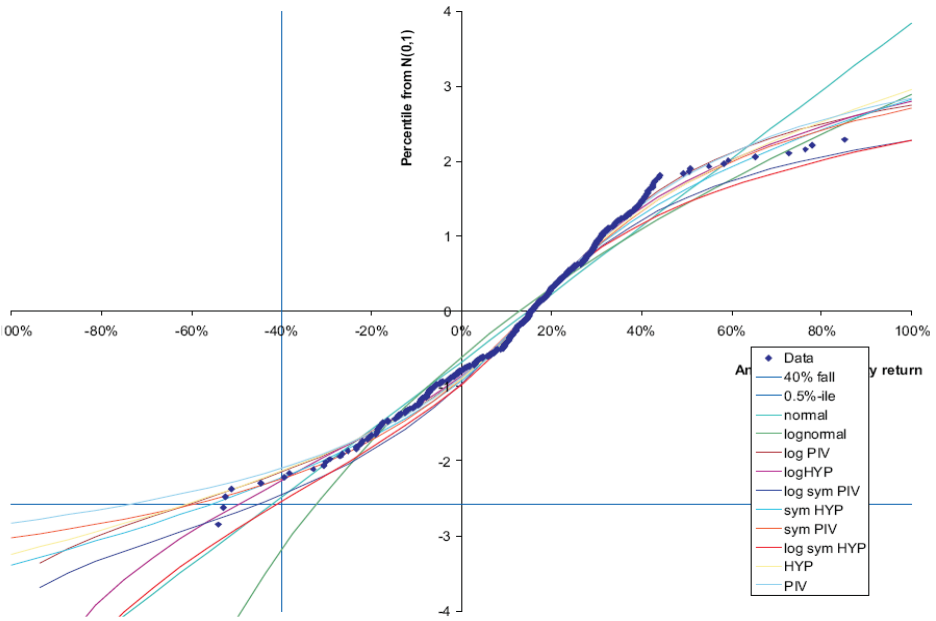
Thus, a simple product with two different commonly used "aggregation" methods produces significantly different answers, amounting to 20% of the liabilities. This example is not special in any sense, in that the two products could represent pretty much any two risk factors. Moreover, most companies have a significantly larger number of risk factors so the 'framework' judgement would be used multiple times.

One can appreciate that there are already multiple choices in simply aggregating different risk factors, even before delving into more exotic copula structures, non-linearities, etc. Finally, this example only touches on a very specific aspect of framework choice; there are many other implicit judgements necessitated by trying to create a simplified representation of the real world.

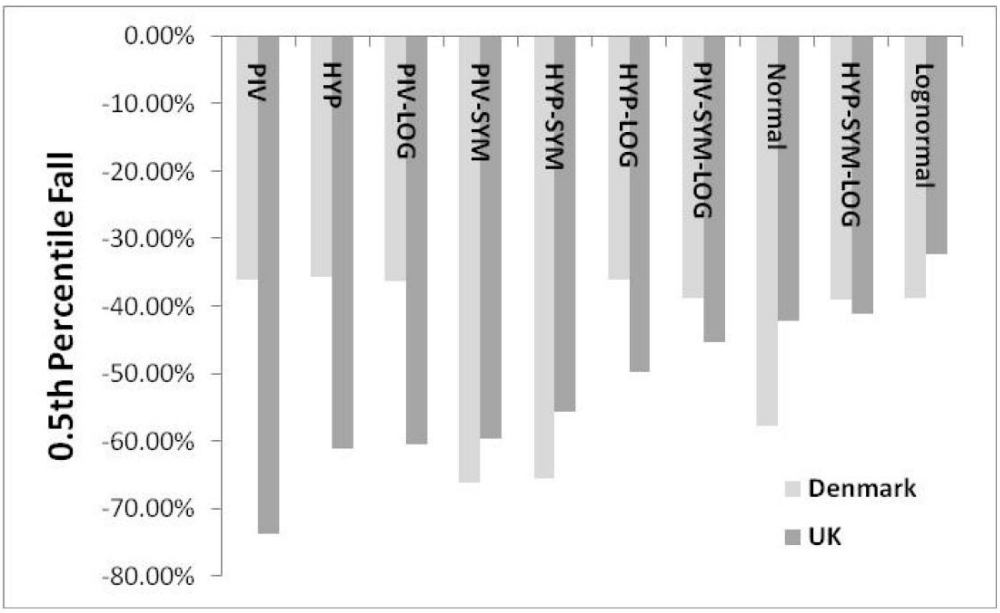
DRAFT

Choosing individual models

For example, previous work done by this working party showed that fitting different models to the same historical data can have a range of different results, even when using very long term data.



"Even after settling on a single data set, the fitted curves for U.K. produce a wide range of values for the 1-in-200 fall. The most extreme results are from a Pearson Type IV, applied to simple returns, which implies a fall of 75% at the 1-in-200 probability level. At the other extreme is the lognormal distribution, with a fit implying that even a 35% fall would be more extreme than 1-in-200 event. Other distributions produce intermediate results."



The next two sections describe the more documented aspects of human judgement, including the choice of calibration methodology and / or parameters.

Choice of calibration methodology

There are other, more commonly acknowledged calibration choices that should be noted in the same context – for example, the choice of data period is crucial. Very simply, if we are using x years of data, then the model would accept the worst event within this data window as being a 1 in x event. Thus the choice of data window makes an important contribution to the results, and it is important to note that even picking any available dataset comes with a default assumption regarding the data period.

This context results in an obvious place where one may want to exercise judgement i.e. one may want to impose some views on the extremity of actual events observed. An obvious example can be constructed by using very short data series that included the recent credit crisis, which would naively overestimate the resulting extreme percentile calculated by assuming such an extreme event would occur regularly within such a short time period. Of course, it goes without saying that the reverse would also have been true when looking at short term data prior to the recent financial crisis!

Choice of parameters, overriding certain parameters if necessary

Judgement in the form of choice of parameters can be explicit or implicit. For example, at the most explicit level, one may override the 1 in 200 stress itself, by super-imposing the views of investment experts. In many cases, the paucity of data makes this a necessary and important part of the capital calculation exercise. One advantage of explicit judgement is that it is extremely transparent, and openly recognises that models and data can only go so far in terms of predicting future distributions.

Alternatively, judgement can be more implicit in the structure of the model. For example, conditional on the form of the model, one may have prior views on the certain parameters. (e.g. we may have a prior view on the sigma parameter of a lognormal distribution).

DRAFT

4. Errors in Proxy Liability Models

This section describes some simplified insurance assets and liabilities for the purpose of testing proxy models in capital aggregation calculations.

Proxy Models
Insurers' assets and liabilities are complicated functions of millions of inputs whose future values are uncertain. The inputs include market prices of assets and financial instruments, incidence of events giving rise to insured losses, as well as vast tables of experienced and projected actuarial decrements such as mortality or attrition. Insurers use "heavy models" to compute assets and liabilities as functions of the long list of inputs.

In theory, insurers are supposed to perform full stochastic projection in order to demonstrate a sufficiently high probability that, in future, assets will exceed the value of liabilities. This means: simulate millions of joint scenarios, in each of which the asset and liability functions are evaluated using a heavy model fed from the millions of jointly simulated input variables.

Despite the advances in computer calculation speed and storage capacity over the last few decades, a full stochastic projection remains beyond the reach of most insurers. Instead, there is a widespread use of simplified "proxy" models. These are simple functions of a small number of variables intended to approximate the heavy model output, usually expressed as a sum of terms whose coefficients are estimated by fitting to heavy models. While running the heavy model millions of times is costly, this number of proxy model evaluations is easily feasible.

Spanning Error
The method of proxy functions assumes that the true assets and liability functions are of the chosen form, or can be approximated sufficiently accurately by functions of that form. The assumption could fail severely, for example if:

- Assets and liabilities depend on inputs that have been excluded in the dimension reduction and are not substantially explained by the retained inputs
- The true function has a "cliff-edge" (i.e. a discontinuity) but all the basis functions are continuous.
- The assets and liabilities have "sink holes" that is, regions of parameter values where assets collapse or liabilities explode, while none of the basis functions exhibit this behaviour.

These are all examples of "spanning failure", where the true asset and liabilities are not in the linear span of a proposed set of basis functions. The "spanning error" is any mis-statement in the required capital that arises from spanning failure.

DRAFT

Selected Test Models

We construct test valuation formulas based on three risky asset classes: government and corporate bonds, equities and a risk-free cash asset. We consider three lines of business: regular premium term assurance, a life annuity and a single premium guaranteed equity bond. We use simplified policy models so that exact valuation is possible with closed formulas.

Risk driver	Term Assurance	Annuity	Guaranteed Equity	Government bond	Corp Bond	Equity
Risk free discount rate	X	X	X	X	X	
Equity price			X			X
Equity volatility			X			
Corp bond portfolio spread					X	
Liquidity premium		X				
Mortality(Term)	X					
Mortality (annuity)		X				
Lapses (term)	X					
Lapses (Guaranteed equity)			X			

This is a simplified set of risk drivers, for ease of calculation. We use a single discount rate for risk-free cash flows of all maturities. The equity volatility is required for valuing the guarantee according to option pricing theory. We assume that annuities cash flow will be discounted at a discount rate that includes a return premium for holding illiquid assets. This is assumed to be assessed with respect to the observed yield spread on a class of illiquid asset subject to credit rating criteria.

Our firm also holds corporate bonds as investments. The change in value of those bonds is driven by changes in bond spreads, but in this case we need to follow the fate of a fixed portfolio of bonds rather than tracking typical spreads for a particular grade. This distinction implies that a portfolio of bonds cannot perfectly hedge the liquidity premium assessed in relation to a particular grade.

Term assurance: Valuation Formula

We assume lapses are a constant proportion of policies in force during the year, while mortality is a constant number of deaths each year (thus, the q-factor increases with age) We construct the policy count as follows between t-1 and t:

In force at balance sheet t-1		$(1 - \text{Lapse})^{t-1} * (1 - \text{Mort} * (t-1))$
Premiums collected		
Deaths from t-1 to t	$(1 - \text{Lapse})^{t-1} * \text{Mort}$	
In force after deaths		$(1 - \text{Lapse})^{t-1} * (1 - \text{Mort} * t)$
Lapses at time t	$\text{Lapse} * (1 - \text{Lapse})^{t-1} * (1 - \text{Mort} * t)$	
In force at balance sheet t		$(1 - \text{Lapse})^{t} * (1 - \text{Mort} * (t))$

DRAFT

We can then calculate:

$$PVclaims = SA * \sum_{t=1}^{T} \frac{(1-lapse)^{t-1} * mort}{(1+disc)^t}$$

$$= \frac{SA * mort}{disc + lapse} \left\{ 1 - \left(\frac{1-lapse}{1+disc} \right)^T \right\}$$

And

$$PVprems = annprem * \sum_{t=0}^{T-1} \frac{(1-lapse)^t * (1-mort*t)}{(1+disc)^t}$$

$$= \frac{annprem(1+disc)}{(disc+lapse)^2} \left[\begin{array}{l} \{disc - mort + lapse * mort + lapse\} \\ - \left(\frac{1-lapse}{1+disc} \right)^T \{disc - mort + lapse * mort + lapse - (disc+lapse) * mort * T\} \end{array} \right]$$

It is easy to find the break-even premium at which the value of claims is equal to the value of premiums; this occurs when:

$$\frac{breakeven_annprem}{SA}$$

$$= \frac{\frac{mort}{disc+lapse} \left\{ 1 - \left(\frac{1-lapse}{1+disc} \right)^T \right\}}{\frac{1+disc}{(disc+lapse)^2} \left[\begin{array}{l} \{disc - mort + lapse * mort + lapse\} \\ - \left(\frac{1-lapse}{1+disc} \right)^T \{disc - mort + lapse * mort + lapse - (disc+lapse) * mort * T\} \end{array} \right]}$$

We can consider two variants of the valuation formula.

- Variant 1 permits negative provisions if premiums are sufficiently large: Technical provisions = PVclaims – Pvprems
- Variant 2 assumes all policies immediately lapse if provisions were to become negative, so that: Technical provisions = max{0, PVclaims – Pvprems}

In our example, in the base case, we assume the actual premium is less than the breakeven premium, so that the value of future benefits exceeds future premiums and the technical provisions are positive under Variant 1. This could reflect the situation of a seasoned policy in which mortality was previously higher. However, we make the inequality close enough so that, under the downward mortality stress, the value of benefits is now less than the premiums. This ensures we obtain different results with Variant 1 and Variant 2.

DRAFT

Annuities: Valuation Formula

Assume a constant rate of mortality 1/*T* in each future year, expressed as a fraction of the initial number of lives, until all have died. The proportion of lives in force at time *t* is

$\max\left\{1-\frac{t}{T},0\right\}$. All policyholders have died by time *t*=*T*.

Thus, the cash flows are (with LCF – liability cash flow per annum before mortality):
LCF * (1-1/T) at the end of year 1
LCF * (1-2/T) at end of year 2

We use discount rate = risk-free + liquidity premium.
The liability (for unit annual cash flow) is:

$$PV = LCF\sum_{t=1}^{\text{int}(T)}\left(1-\frac{t}{T}\right)(1+disc)^{-t}$$
$$= LCF\left\{\frac{1}{disc}-\frac{1+disc}{T.disc^2}\left[1-\frac{1}{(1+disc)^{\text{int}(T)}}\right]-\frac{1-\text{int}(T)/T}{disc.(1+disc)^{\text{int}(T)}}\right\}$$

Here, int(*T*) is the integer part of *T*, that is, the largest integer not exceeding *T*.

Annuities are single premium products. As the premium has already been paid, we do not need to model it here. There is no lapse risk with annuities.

Linked with Guarantee: Valuation Formula

We consider a fund with initial value Unit_Fund, invested in risky assets but with a guaranteed amount Gtee_AMT at maturity. We assume that the annual management charge (AMC) is deducted annually in advance. Lapses occur during the year, after the AMC is deducted (with 0 < AMC < 1.00). The total technical provision is the unit fund minus the value of future charges, plus the guarantee.

The value of charges is as follows (assuming an integer term *T*≥1)

$$Unit_Fund*\left[AMC+(1-AMC)(1-lapse)AMC+\ldots+AMC(1-AMC)^{T-1}(1-lapse)^{T-1}\right]$$
$$=Unit_Fund*AMC*\frac{1-(1-AMC)^T(1-lapse)^T}{AMC+lapse-AMC*lapse}$$

The guarantee is assumed to apply only to policies that reach maturity, and is given by the formula:

$$(1-lapse)^T*BS\left(\frac{Gtee_AMT}{(1+disc)^T},(1-AMC)^T*Fund,volatility*\sqrt{T}\right)$$

Here, the Black-Scholes formula is:

$$BS(PVget,PVpay,\alpha)=PVget*\Phi\left(\frac{\ln(PVget/PVpay)}{\alpha}+\frac{\alpha}{2}\right)-PVpay*\Phi\left(\frac{\ln(PVget/PVpay)}{\alpha}-\frac{\alpha}{2}\right)$$

DRAFT

As with the term assurance, we can consider whether the value of the guarantee is more or less than the value of charges. However, we have no simple formula for the breakeven rate of annual management charge. This because increasing the charge causes a rise in the value of charges (clearly) but also reduces PVpay in the option formula, leading to a rising guarantee cost because more is taken out of the fund before the guarantee is met.

Indeed, in some cases, there is no break-even management charge, as attempts to increase the value of charges also increase the guarantee cost, with the value of charges never catching up. To avoid this, we have to insist that:

Unit_fund > $(1\text{-lapse})^T / (1+\text{disc})^T$ * Gtee_AMT

Equivalently, expressing the guarantee as an annual return Gtee_AMT = $(1+\text{gtee_ret})^T$ * Unit_Fund, a break-even management charge exists provided the guaranteed return is not too onerous; specifically

1 + gtee_ret < (1+disc) / (1 –lapse)

As with the term assurance, we can investigate this product on two bases:
Variant 1 permits negative provisions if charges are sufficiently large: Technical provisions = guarantees - charges
Variant 2 assumes all policies immediately lapse if provisions were to become negative, so that: Technical provisions = max{0, guarantees - charges}

Assets
Equities: Responds only to equity stress
Cash: Unaffected by any stress
Risk free bond (term): Affected only by risk free rate, assuming an annual coupon and term *T*.
Corporate bond (term); affected by risk free rate plus credit spread, assuming an annual coupon and term *T.*

The formula for the risk free bond and corporate bond is:

$$Market_Value = Face_Value * \left\{ \frac{1}{(1+disc)^T} + coupon \sum_{t=1}^{T} \frac{1}{(1+disc)^t} \right\}$$

$$= Face_Value * \left\{ \frac{1}{(1+disc)^T} + \frac{coupon}{disc} \left[1 - \frac{1}{(1+disc)^T} \right] \right\}$$

Curve Fitting Methodology

We now develop tools for fitting curves to the actuarial formulas:

We fit curves to a series of stress tests. The stress tests are constructed using median values for each risk driver, as well as high (α-quantile) and low (1-α quantile) values. For example if α = 0.995 then we would calibrate to the 0.5%-ile, median (50%-ile) and 99.5%-ile for each risk driver.

We look at all combinations of high, median and low for each risk driver. That implies that with 9 risk drivers, we may have to examine 3^9 = 19 683 stress tests. Thankfully, this is an overestimate, as our example assets and liabilities depend individually on at most 4 risk drivers. The maximum number of stresses is then 3^4 = 81 risk drivers, in the case of the

DRAFT

guaranteed equity bond. The minimum number of stresses is 3, for the risk-free bond and the equity, which depend on only one risk driver.

Having estimated constructed the stress tests, we then fit a polynomial. In each case we use ordinary least squares to fit the formula. In this note, we consider three fits:

- *Linear fit*: We estimate assets or liabilities as a constant term plus linear terms in each risk driver.
- *Separable Quadratic Fit*: We estimate assets or liabilities as a constant term, plus linear and squared terms in each risk driver. This formula is still separable, that is, basic own funds is a sum of functions each of which depends only on one risk driver. There is no mechanism for the level of one risk driver to affect the sensitivity of net assets to another driver.
- *Quadratic Fit with Cross Terms*: Assets and liabilities are a constant term, plus linear and squared terms in each risk driver, and also products of risk drivers taken two at a time.

We can consider extending these fits to cubic and higher order polynomials. However, at least in the one-dimensional cases (equities and government bonds), we have only three fitting points in which case an exact quadratic is uniquely defined and a cubic is not.

Risk Driver Distribution

For demonstration purposes we assume that risk drivers have shifted lognormal distributions. We construct these using formulas of the form:

$$riskdriver = median + b\frac{e^{cZ}-1}{c}$$

Here, *b* is a positive coefficient; while *c* is positive or negative. If *c* = 0 we replace the fraction $\frac{e^{cZ}-1}{c}$ by its limiting value *Z*. This is readily calibrated to quantiles. For example, suppose for some *k*, the Φ(-*k*) quantile is median–α and the Φ(*k*) quantile is median+β, we have to solve the equations:

$$b\frac{e^{-ck}-1}{c} = -\alpha$$

$$b\frac{e^{ck}-1}{c} = \beta$$

From this, it immediately follows that:

$$e^{-ck} = \frac{e^{-ck}-1}{1-e^{ck}} = \frac{\alpha}{\beta}$$

$$c = \frac{1}{k}\ln\left(\frac{\beta}{\alpha}\right)$$

$$b = \frac{c.\alpha\beta}{\beta-\alpha}$$

Once again, there is a special case when α=β, where *c* = 0 and *b* = α/*k*.
Within firm's internal models, there is no general rule relating the opening (t=0) risk drive values to their future distribution. However, for simplicity in the current curve fitting tests, we assume that the opening values are equal to the median.

DRAFT

5. Appendix - Bayesian Methods

This section discusses the use of Bayesian methods and how they might be used to provide a framework for Expert Judgement. The Bayesian method is described and an example case study is worked through.

General Description

Bayesian and frequentist statistical inference take very different approaches to statistical decision making.

- The frequentist view of probability, and thus of statistical inference, is based on the idea of an experiment that can be repeated many times
- The Bayesian view of probability and of inference is based on a personal assessment of probability and on observations from a single performance of an experiment

These different views lead to fundamentally different procedures of estimation, and the interpretations of the resulting estimates are also fundamentally different[1].

Bayes Theorem is the basis for Bayesian methods. For an observed event E and a partition {A1,A2,…,Ak} of the sample space S,

$$P(A_j \mid E) = \frac{P(A_j)P(E \mid A_j)}{\sum_{i=1}^{k} P(A_i)P(E \mid A_i)} \qquad \text{[Formula 1]}$$

The general version of Bayes Theorem involving data x and a parameter π is shown below:

$$p(\pi \mid x) = \frac{p(\pi)\,p(x \mid \pi)}{\int p(\pi)\,p(x \mid \pi)\,d\pi} \propto p(\pi)\,p(x \mid \pi) \qquad \text{[Formula 2]}$$

A posterior distribution of π is found from the prior distribution of π and the distribution of the data x given π. This is the basis for Bayesian methods. An initial prior view, based typically on expert judgment is updated with some data to give a new view that allows for both the data and expert judgment. Bayesian methods can be summarised as:

Posterior distribution $\propto$ Prior * Likelihood function [Formula 3]

[1] "Introduction of Probability Simulation and Gibbs Sampling in R" E.A Suess and B.E. Trumbo Page 195

Bayesian methods and expert judgment

Expert judgment in Bayesian methods arises when the prior distribution is defined by an expert. There are a number of ways of doing this and typically it depends on eliciting meaningful information from the expert that can be used to derive a probability distribution.

Case studies

In this section an example case study is considered. It considers how Bayesian methods might be used to calibrate a 1 in 200 equity stress calibration based on a specific data set. The data used is based on the Dimson Marsh & Staunton (DMS, 2002)[2] study. The data used includes updates to the original DMS data and covers the period 1900-2008. The case study below is based on logarithmic annual returns of the UK DMS data set.

Case Study

In this case study the simple case where the data is assumed to have a Normal distribution and the variance of the data is unknown is considered. These assumptions allow Formula 2 to have an analytic solution. More complex distributional assumptions mean that Formula 2 produces an analytically insolvable integral. Monte Carlo methods such as the Metropolis Hastings algorithm can be used in these cases. This case study has three different elements.

1. Two different expert judgements are considered and compared. These expert judgements are captured in two different prior distributions used. This presents the reasonably likely scenario in a life insurance company where the risk function might have a different expert judgement from the capital management team.

2. The posterior distribution is calculated for each of these expert judgements for a subset of the total data set (from 2008-1999); and secondly for the full data set (from 2008-1900). The impact on the posterior distribution from different expert judgements using the two different data sets is contrasted.

3. The calibration of the variance of the Normal distribution using Bayesian Methods is contrasted with the variance calculated using a frequentist method (the Maximum Likelihood Estimate).

Conjugate prior for variance of Normal distribution

In this case study the data i.e. the likelihood function, has a Normal distribution:

$$p(x \mid \pi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\}$$

For several (n) observations

$$p(x \mid \pi) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left\{ -\frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2 \right\}$$

[2] Dimson, E, Marsh, P & Staunton M (2002) Triumph of the Optimists, Princeton.

$p(x \mid \pi) \propto (\frac{1}{\sigma^2})^{-\frac{n}{2}} \exp\left\{-\frac{ns^2}{2\sigma^2}\right\}$, where $ns^2 = \sum_i (x_i - \mu)^2$

The conjugate prior for variance of the Normal distribution is the Inverse Gamma family, that is:

$p(y) \propto y^{-\alpha-1} \exp\left\{-\frac{\beta}{y}\right\}$, is the Inverse Gamma distribution with parameters α and β

(i.e. y ~ invGam(α, β))

Setting y= σ^2 gives:

$p(\sigma^2) \propto (\sigma^2)^{-\alpha-1} \exp\left\{-\frac{\beta}{\sigma^2}\right\}$ (i.e. σ^2 ~ invGam(α, β))

$p(\sigma^2 \mid x) \propto p(\sigma^2) p(x \mid \sigma^2)$

$p(\sigma^2 \mid x) \propto (\sigma^2)^{-\alpha-1} \exp\left\{-\frac{\beta}{\sigma^2}\right\}.(\frac{1}{\sigma^2})^{-\frac{n}{2}} \exp\left\{-\frac{ns^2}{2\sigma^2}\right\}$

$p(\sigma^2 \mid x) \propto (\frac{1}{\sigma^2})^{\frac{n}{2}+\alpha+1} \exp\left\{-\frac{1}{2\sigma^2}(ns^2 + 2\beta)\right\}$

Hence,

Prior: $p(\sigma^2)$ ~ invGam(α, β) =>

Posterior: $p(\sigma^2 \mid x)$ ~ invGam(α+n/2, β+ $n/2.s^2$)

That is given a prior distribution, the posterior distribution can be found with just n and s^2 from the data.

Prior distributions

In this case study two prior distributions are assumed for the calibration of a 1 in 200 year event. Two different experts (one from the firm's capital team and one from the risk team) are asked two questions about what they consider to be a 1 in 200 year event. The questions are:

1. What is your best estimate for a 1 in 200 year event for the firms equity risk
2. Please give a percentile you are fairly sure (i.e., 95% sure) the 1 in 200 event is lower than

As the 99.5th percentile is the object of interest in this case study, both experts agree that the expected equity return over the next year is 5%

The answers the experts give for annual simple equity returns are:

	Capital team expert	Risk team expert
Question 1	-40%	-50%
Question 2	-65%	-75%

Table A.1

These answers can be used to calibrate prior Inverse Gamma distributions. Converting these expert judgments to logarithmic returns (which the underlying data is based on) and calibrating α and β for each expert judgment gives:

DRAFT

Prior	Capital team expert	Risk team expert
α	2.074	2.074
β	0.051	0.089

Table A.2

Posterior distributions

Using the UK DMS data set two periods of data were considered:

- From 1900-2008 (109 annual non overlapping data points)
- From 1999-2008 (10 annual non overlapping data points)

The posterior distribution calibrated for the two different expert judgments and the two different data sets are shown in the table below:

Posterior	Capital team expert		Risk team expert	
	1900-2008	1999-2008	1900-2008	1999-2008
α	56.57	7.07	56.57	7.07
β	1.89	0.25	1.93	0.29

Table A.3

Having calibrated the prior and posterior distributions for the variance, it is now possible to present:

1. The best estimate of the 99.5th percentile using the mean of the distribution of the variance for the variance of the underlying data (i.e. using $E(\sigma^2)$ as the variance). This is the **best estimate 99.5th**
2. A 99.5th percentile calculated by taking the mean of the distribution of the variance plus two standard deviations of the variance as the variance of the underlying data (i.e. using $E(\sigma^2)$ +2* $\sqrt{Var(\sigma^2)}$ as the variance). This is the **high estimate 99.5th.**

The aim of the second point is to show changes in the uncertainty that the use of data brings relative to just the prior expert judgments.

	Capital team expert		Risk team expert	
99.5th percentile	1900-2008	1999-2008	1900-2008	1999-2008
Prior best estimate 99.5th	-40.0%	-40.0%	-50.0%	-50.0%
Prior high estimate 99.5th	-79.2%	-79.2%	-87.7%	-87.7%
Posterior best estimate 99.5th	-34.7%	-37.8%	-35.0%	-40.2%
Posterior high estimate 99.5th	-38.6%	-48.9%	-38.9%	-51.5%

Table A.4

There are two main results in Table A.4

- The impact of the differences between the two experts gets much smaller as more data is included. This is compared by looking at the 99.5th percentiles. The prior distributions have 40% vs 50% (a difference of 10%). With just 10 data points, this moves to 37.8% vs 40.2% (a difference of 2.4%). With 109 data points this moves to 34.7% vs 35.0% (a difference of 0.3%).

- The uncertainty around the estimate falls as more data is included in the analysis. For example, taking the capital team expert distributions, the prior difference between the best estimate 99.5th and high estimate 99.5th is 40% vs 79.2% (a difference of 39.2%). With 10 data points this moves to 37.8% vs 48.9% (a difference of 11.1%). With 109 data points this moves to 34.7% vs 38.6% (a difference of 3.9%)

Comparison to the Maximum Likelihood Estimate

The above results can be compared to frequentist approaches of distribution fitting such as the Maximum Likelihood Estimate (MLE). The two data sets above are calibrated to the Normal distribution using the MLE.

The variance is calibrated using the MLE and the 99.5th percentile is calculated from a mean of 5% in line with the method taken in the Bayesian approach above. This ensures the only difference between the two approaches is the calibration of the variance.

The MLE gives estimates of standards errors for the estimate of each parameter which give an indication of the uncertainty in the parameter calibrations. In the table below as similar approach as used in Table A.4 is given, with:

- The **best estimate 99.5th** taken from the variance parameter calibrated using MLE
- The **high estimate 99.5th** taken from the variance parameter plus two standard errors calibrated using MLE

	1900-2008	1999-2008
Best estimate 99.5th	-34.6%	-34.7%
High estimate 99.5th	-67.6%	-72.3%

Table A.5

The results seen in table A.5 are somewhat similar to those in table A.4. Looking at 109 years of data the 34.6% from the MLE compares to 34.7% (capital team expert posterior) and 35.0% (risk team expert posterior) using the Bayesian approach. The difference is larger with just 10 years of data, with 34.7% from the MLE compares to 37.8% (capital team expert posterior) and 40.2% (risk team expert posterior) using the Bayesian approach.

The biggest difference between the two approaches is the uncertainty in the calibration of the 99.5th percentile. For 109 years data the best estimate and high estimate for the MLE for the 99.5th percentiles is 34.6% vs 67.6% (a difference of 33.0%); compared to 34.7% vs 38.6% (a difference of 3.9%) for the Bayesian approach with the capital team expert judgment. This difference is less big for just 10 years data (i.e. 34.7% vs 72.3% (i.e. a difference of 37.6%) for MLE compared to 37.8% vs 48.9% (a difference of 11.1%) for the Bayesian approach. The reduced uncertainty in the Bayesian approach relative to the MLE might be attributed to the addition of expert judgment included in the Bayesian approach, but not present in the MLE.

DRAFT