

## Difficult Risks and Capital Models: *A report from the Extreme Events Working Party*

---

### Abstract of the London Discussion

[Institute and Faculty of Actuaries, 22 April 2013]

#### Contact

Parit Jakhria, F.I.A., CFA, E-mail: [parit.jakhria@mandg.co.uk](mailto:parit.jakhria@mandg.co.uk)

**Dr D. J. P. Hare, F.I.A. (Chairman):** I am going to invite Parit Jakhria, who is one of the authors of tonight's paper, to introduce it. He is an experienced actuary who works for the Prudential in the capital modelling team and has worked on a number of working parties for the profession. He has helped to produce what I think is an excellent paper.

**Mr P. C. Jakhria, F.I.A., CFA:** Welcome to the discussion of our professional paper entitled 'Difficult Risks and Capital Models'. We talk about modelling in general and capital models in particular. Before I proceed any further, I should like to acknowledge and thank the contributions of all the members of the Extreme Events Working Party. It is on behalf of the working party that I will be summarising the paper this evening.

My intention is to provide a very high level summary of the themes of the paper. The paper itself has quite a number of ideas, so I will try to capture the key ideas and also use three examples to highlight some of the ideas that we think are important and/or new.

As background, it is worth talking a little about modelling in general. Models permeate every corner of the actuarial world. Although our paper discusses capital modelling in particular, many of the insights may apply to actuarial modelling in general.

Right at the start of the paper there are some real, as well as hypothetical, examples of where capital models have gone wrong.

Before discussing the paper, I will give a brief summary of our understanding of capital models. We gather information from the past, plus some insight into present conditions. We combine that with the knowledge of a particular problem and model the future, with the ultimate aim of making decisions about the future, for example how much capital to put aside.

We would like to highlight that a model is a simplified representation of the real world. Thinking about it philosophically, if one wanted a perfectly accurate model of the universe it would need to be as big as the universe. In fact, we want much more than an accurate model of the real world in capital modelling. We want to be able to project faster than real-time, so as to extrapolate to 'tail' scenarios. You have to make some choices in terms of how to simplify the model to match the real world. It is really this process of simplifying which requires a huge amount of choice and judgement. Later we will go into one of the examples in a little more detail.

We try to list the different areas in which you can encounter choices when modelling. These are not exhaustive, but we try and cover a lot of different things. The single biggest decision, for example, is simply what risk factors to model or, more precisely, what risk factors to model stochastically.

Other than that, you can choose your overall framework (e.g. whether you want to use a building block approach, a risk factor approach, etc.). For each model component you could choose which model to use. You could choose: how to calibrate the model; what data to use; whether there is any judgement within the data; and, finally, the parameters for the chosen model. It is probably much more interesting to look at some examples.

If you take the average company balance sheet, for example, we can see that the number of random factors easily adds up to over 1 million. Say you have a company which has 30 products in three countries, each product having a number of different settings. Underlying each product, there may be 5000 individual customers. In practice, it is likely to be many more than that. What you can see is that the number of items that can vary in a non-deterministic way goes up very rapidly. What you also find is that, using current technology, you simply cannot have so many stochastic factors in the model. Most companies use some form of reduction technique: grouping of policy data; assuming certain factors (e.g. lapses) are deterministic or functions of other variables; or use stock indices rather than modelling individual stocks; etc. You may still end up with a large number of factors. You could ultimately use your judgement, for example, to choose the different factors, or you could use statistical reduction techniques. We talk about that section 3 of the paper.

One of the important points that we make is that, irrespective of how you come down from a very large number of factors to a very small number that you are modelling, one thing you have to remember is that there is a lot of choice involved. To summarise, there are a large number of choices inherent in building the model. Another important thing to remember is that we cannot ignore the risk factors that we identified as being variable but did not model. So we need a way to gross-up factors. A very naive example would be say you had 1 million initial factors which you boil down to 1000, you could multiply your capital by 1000, assuming that you picked those factors randomly. Such an approach is going to raise eyebrows, and it should do so because it is to be hoped that most companies use some science to choose the factors, and do not do so at random.

The point remains that you need to have some capital for the factors that you have identified but not modelled. This is also good practice from a risk management point of view because it provides an incentive to have better modelling of risks and better risk management in order to then reduce the capital you put aside for those risks that you have not fully modelled.

Having considered what risk factors to model, another part of the paper (section 4) is devoted to the concept of model and parameter risk. Let us start off with a very simple example. Suppose we know the true model, which means we can have an exact distribution of our capital and we can just read off, or numerically calculate, the percentiles. Suppose we know it is a normal distribution with a particular mean and standard deviation, and then everyone knows what to do. You find the critical value for a normal distribution which happens to be 2.58, and you say our capital is the mean +2.58 times the standard deviation. As one goes through more complicated distributions, the mathematics may become more complicated, but the idea is essentially the same. It is a known function with known parameters.

So, what happens when you do not know the model? The situation changes drastically when you have potential model errors. We highlight using an example that was published by our working party in 2008 as part of the analysis on equity stresses.

What we had was about 100 years of data for a country. In the context of other risks, that is a huge amount of data; it can be very difficult to obtain the same length of dataset for some of the other risks that are frequently modelled.

For that data we fitted a number of models. One thing we are going to try to explain is that although you may have a model that is the most likely given the dataset, you cannot reject all the other models. So you may have lots of models which you still, statistically, cannot reject, even after 100 years of data.

If we plot the fits for all those distributions, it looks like section 2.1.3 of the paper.

Say we had 100 years of data, but the regulations ask us to find the one in 200 year event. You can see, as you extrapolate, all the numbers fan out. One interesting point is that the difference between the capital requirement of the most prudent and the most optimistic model was as much as a factor of two. This is based on the biggest dataset that we have. You could only imagine what it looks like for some of the smaller datasets.

If I were to summarise this very quickly, you could say that there may be several models that adequately explain the data, and you may find that some models are better fitting than others, so you may be able to choose a particular model as the most credible explanation. However, that is not the same thing as saying that is the only credible explanation: there may be other models out there which are also plausible explanations that we cannot reject.

This begs the question: what does that mean and how do we try to allow for it? We have tried to allow for it in the case of parameter uncertainty. When the model and parameters are very certain, then the answer is, effectively, reading off a percentile, either calculated exactly or numerically, and it is fairly straightforward. That neatly fits into the traditional definition of value at risk.

When the models or parameters, and in this case we are focusing on parameters, are uncertain, even the meaning of value at risk has different possible interpretations. I will give some relevant background. Suppose  $K$  is the percentile that you estimated. That percentile is a function of three elements. It is a function of the data; the model you have used; and the parameters that you have used to fit to the model.

There are different ways of going about calculating the value at risk. The most common way is to do what we call a best estimate calculation. You take each parameter that you need to estimate. You find the best estimate value of those parameters and then you plug them into a model and assume that that is the true model and calculate the answer.

One thing to consider is whether, if you have all the unbiased estimates of the model, the answer itself is unbiased. This depends on how the models and parameters are set off. You may wish to remind yourself, at this stage, about Jensen's inequality, which says, if you have a convex function, the expectation of  $F(X)$  is not the same as the function of the expectation of  $X$ . In fact, it is greater

than that. So you need to be very careful about this inequality and ensure that you do not underestimate your capital. That is just the first stage.

What you could say is “let us carry out an unbiased estimate and that will give you your unbiased 99.5 scenario”. You could also acknowledge the risk of parameter error, and ask to be 95% sure that the parameters of our model capture that scenario.

The counterargument, potentially, is that it is very prudent because you would effectively be taking the 19<sup>th</sup> of 20 worst models. Say you have 20 possible models, you are taking the 19<sup>th</sup> worst choice of them. That is not going to be very palatable from a capital perspective. You also have to take into account that you may be picking from unlikely models. What you want to do is to try to have some sort of diversification between incorrectly estimating the model and extreme market events.

You can think of a concept of a prediction interval. What we are saying is simply “what is the risk value such that the probability that the next observation is less than your risk value is 99.5%?” You want to be 99.5% sure that your next observation will not be bigger than what you have calculated. What may be helpful is to look at it for different models (refer to the diagram in section 4.1), where we note that the prediction interval is less than the confidence interval, but greater than the unbiased interval. All of these are likely to be greater than the ‘best estimate’ method.

What we have not yet discussed is how to allow for model uncertainty. If one looks at the different models in section 4.1, you can see that we obtain substantially different answers (as a multiple of standard deviation) for different models. In the paper we explore some ideas on how to generalise the concepts. You could say we do not have the modelling resource to go through all the different options of the models so we are going to pick a reasonably prudent model out of the ones that we have identified and base the calculations on that model, hoping that they are close enough to the answer. You could say all models cannot be rejected unless you can statistically reject them and then build some kind of giant Monte Carlo hyper-model, which goes through two phases. The first phase is a random number generator that samples which model you are in; and the second phase is the random number that samples from within the model. That in itself is costly in terms of modelling capacity and also requires some judgement on the prior distribution of models.

You can also try to generalise the parameter concept to an area that Andrew Smith terms ambiguity sets. In my words, it is a family of models that is general enough to encompass many models. As this family of models is very general, it can have many models within its realm, and one could then try to come up with estimates of location and scale parameters within that large family and carry out some of the work that we did on prediction intervals. An important question, which I believe is still unanswered, is simply how big a set of models should you choose initially. It will be interesting to hear your thoughts on how different participants in industry come to a conclusion on that question.

In section 5 we also touch on a rather newer phenomenon in actuarial modelling, known as ‘lite’ or ‘proxy’ models. These are effectively a simplified version of the traditional ‘heavy’ model. There is some discussion on the different challenges faced, and also some discussion on Monte Carlo sampling error.

Finally, we also talk about Bayesian methods as potential tools with which to apply judgement.

**The Chairman:** Thank you Mr Jakhria. I would now like to invite Andrew Hitchcox to open our discussion. Andrew is the Chief Risk Officer at Kiln. He has been there for many years, and is one of the leading thinkers on Enterprise Risk Management in our profession.

**Mr A. N. Hitchcox, F.I.A.:** When I started in capital modelling, many years ago, we used to put the topics of parameter risk and model risk in the ‘too hard tray’. But they are now becoming part of the near future for actuaries for a variety of reasons:

- The world changes so fast these days, that we do not have the time to build up long data series: we have to rely more on modelling for financial projections;
- We are increasingly living in a modelled world in other parts of society outside the insurance industry:
  - In science, e.g. climate change forecasting;
  - In commerce, e.g. logistics and distribution companies use forecasting models, as do airline ticketing and reservation systems; and
  - In social media and internet services, with their profit-maximising algorithms.
- In our own insurance world, models create and consume capital:
  - If you improve a model’s accuracy which lowers the results of the calculation, you can release capital which otherwise would need to be held; and
  - If you make mistakes or change your model suddenly upwards, you can consume capital. That is of vital importance to investors.
- The Prudential Regulation Authority (PRA) in its recent launch document said:
  - “Internal models introduce additional risks that should be understood and managed appropriately by an insurer and its senior management”.

My first comment on the paper is on Section 2, which discusses the topic of judgement in depth and in a very helpful way:

- Expert Judgement is a big topic for Solvency II;
- But it is also a very important topic for capital providers and investors themselves;
- Investors are the ultimate risk bearers of management’s modelling choices; and so
- Those of you who are Chief Actuaries or Chief Risk Officers should study this chapter carefully, and think what lessons you can learn to help you in the governance of your model, and how you make transparent the expertise that you bring to bear. That is a very large part of the model, as the paper makes clear.

My next comment is on Sections 4.2 and 4.3 on the topic of model risk. I liked these sections very much, and would advise readers to pay close attention and make sure you get the structure of the questions clear in your mind.

My one request to the authors is as follows:

- Everywhere else in the paper, when you posed a question, you gave us very specific and useful examples and illustrations on how to tackle the issues.
- If you have the energy to do another paper on this topic, say in 2 years’ time, can you start with Section 4.3, and construct some more specific visual or numerical examples of the solutions and ideas that you discussed in words in the paper?

Finally, I want to emphasise to the audience the importance of Section 5, which deals with the topic of ‘Errors Introduced by Calculation Approximations’:

- It tackles two specific topics, namely ‘Proxy Models’ and ‘Monte Carlo Simulation Error’. These are both very important topics in practice.
- I’m sorry to quote some Solvency II text to you, but Level 2 Article 230 on Validation requires you to produce “an analysis of the stability of the outputs of the internal model for different calculations using the same input data”.
- So those of you involved in Model Validation work should study chapter 5 of the paper very closely, and make sure that you can address the issues raised there.

Looking ahead to the future:

- I mentioned earlier on the importance of understanding model risk as part of the governance issues;
- The impact of model error is a risk ultimately borne by investors;
- It is not too much of a stretch to envisage the eventual requirement for the public reporting of model risks or uncertainty in financial or supervisory statements;
- If you articulate the genuine model uncertainties that your firm faces, will it be seen as a sign of strength or of weakness?
- Remember that an important service to society given by the insurance industry is to take on those risks that individuals and commerce do not know how to, or do not want to, manage themselves. Yet the investment analysts are very demanding when we come up with modelling errors which are quite genuine and possibly not the fault of the modellers involved; so
- If the authors are looking for yet another area of future work, I would ask them put their collective minds to the subject of how to articulate model risk to outside investors in a way that is useful to all concerned.

In conclusion, I would like to say well done to the authors for opening up an important new chapter of actuarial endeavour, and to get us started on an important but very difficult topic.

**The Chairman:** I am going to hand over the floor to Louise Pryor.

**Dr L. M. Pryor, F.I.A.:** This is a great paper. The big message is, of course, not to believe your models. That is something that I have been saying for quite a long time. I do have a minor criticism. They do not use the best quotation of all from George Box: “All models are wrong, but some are useful”.

The paper is especially useful because of the examples. It is really useful to see how big an effect some of what one might think of as comparatively minor decisions can have on the final result.

The paper also points up something which I am sure you will not be surprised to hear me say: it is very important to remember the limitations of models when you are making decisions based on them, and it is extremely important to communicate those limitations to people who are making decisions.

I am going to take a minor detour although I think you will see the point when I get to the end. I am going to talk about a very important characteristic of models, which is robustness. If the model

results vary dramatically when the inputs do not vary by very much, the model is not robust and the results should be treated with caution. It is important that you have some idea of how robust is the model, and it is important that the people making the decisions have some idea of how robust is the model. One way of doing that is to provide full information. For very big and complex capital models you cannot provide detailed information about all the ins and outs of the models, the tests you have done, and all the data and assumptions. What you can do is make it possible for other people, should they have the time and inclination, to go through, check and run their own experiments.

There has been a really good public example recently about what can go wrong when this is not done. I am sure you have all heard of Reinhart and Rogoff. If you have not heard about them in the last week or so, you have not been reading the same news sources as I have been reading. They had a very famous paper which looked at countries' debt ratios and growth rates and claimed that if a country has a 90% or higher debt ratio, then its growth rate is going to be very low.

I have not actually read the original paper so I do not know if they went on to claim causation or just stopped at correlation. But many people have made the jump from correlation to causation and it has affected the macro economic policies of many large countries including the United Kingdom. The politicians have said "We must keep our debt ratio low, we have to take the austerity route, we cannot spend our way out of this crisis." It has had major effects on a lot of big economies.

Reinhart and Rogoff did not publish the data on which they based their conclusions and they did not publish full details about the model. It has emerged within the last week or so that the model was faulty. A student of two researchers at the University of Massachusetts was set an assignment to try and reproduce the results of any famous economics paper. He chose this one. He could not reproduce the results. He eventually arranged for them to send him their spreadsheet. He and his advisers found three major issues. First, there was an Excel error. This should not surprise people. Excel errors happen all over the place. This error meant that the data for, I think, three countries was omitted from the final averaging. Second, there was some selective omission of the data for some years for some countries. It is assumed that Reinhart and Rogoff had a reason for this. I do not know that there has been any particularly satisfactory explanation, and in any case it was a matter of judgement whether to omit that data or not. That brings us back to one of the themes of this paper. Third is another judgement issue: the weights they used for the countries in their averages were not the weights that many other researchers would have chosen. I think that that is a simple way of putting it. These three issues made a huge difference to the result. If you reinstate the missing data and change the averaging method 90% does not seem nearly so significant when looking at debt ratios to growth rates. I think this should give us pause for thought. I am not remotely surprised to find Excel errors. I am sure you will find them all over the place. But what it means is the results simply are not robust to omitting some data or putting some in.

If leaving out three countries makes so much difference when the results are based on only about 20 countries anyway, we need to consider the impact of the 160 or so in the world that are omitted. Admittedly, some of them, probably, are not useful because they do not have long-term data or they are very different types of economies. But, I bet there are five or ten countries that are comparable that could have been included. How much difference would they have made? What if the period covered by the data had been slightly different?

We can see that omitting some years for some countries made a huge difference. What if there had been another five years omitted off the beginning for all countries or another five years put on the beginning? Admittedly, they started at 1946 so starting five years earlier might not be useful. But there is a point that is made very well in this paper: you have to think about the period your data covers. Robustness is important, and it is really important that there is a way that people can check what you have been doing.

The moral of this story is that we should all be relieved that our models are not being used to influence the macro economic policy of large economies. I am sleeping a great deal sounder in my bed at night because of that. But capital requirements of large insurers are not insignificant and we should definitely be afraid of making mistakes.

We are all members of the Actuarial Profession, and we have technical and ethical standards that, I hope, mean that we act responsibly. But I also think that there are issues here that the Profession, as a public body, should take seriously. Actuaries understand the limitations of financial modelling as shown by this paper and the work done by actuaries in their jobs communicating with the people who make decisions. The Profession should be trying to ensure that others understand the limitations, too, and that they are realistic about how models can be useful even if they are wrong.

**The Chairman:** I will throw out a question. Given the lack of data, it is quite easy to arrive at the view that there is no point in modelling some risks at all and that something pragmatic would be just as useful. Have the authors or anyone on the floor any comments on that point? For example, does anyone here feel that there is no point in modelling persistency risk because there is so little data? Similarly, is there enough data to model credit spread splits?

**Mr Jakhria:** That is an interesting way of looking at things. I think what you are saying is that there is no obvious model to choose. I would rephrase that as “there is no obvious model we can reject”. What that means is that data could possibly come from a large number of models, which has rather different implications compared to using a constant model.

There is no easy way to deal with it apart from making sure, as we said in the paper, that the capital calculations for the risks that you have not modelled allow for some kind of grossing-up factor, so that you have to put aside some capital. As you obtain more information about the risk and better modelling techniques then you may be able to reduce that amount of capital, which gives you the correct incentive. You could always carry out some sensitivity analysis on models you think may be plausible, again bearing in mind modelling risk. Perhaps Andrew Smith may have more comments to add?

**Mr A. D. Smith (student):** One of the points that we picked up in the paper was the concept of an ambiguity set. That is a class of possible models under which the technique that I have applied demonstrably produces a one in 200 event. Outside your ambiguity set, there are other models which you have not considered and which you are saying nothing about.

Our primary difficulty is a failure to specify the problem, and not a difficulty in solving a well-defined problem. The way you formulate the problem, can end up being a social convention. If I asked everybody in this room what they thought six 8s were, there would be a large consensus around the number 48. If I asked people in this room whether it is appropriate to wear a tie to a sessional meeting, there probably would be a lot of people who would say yes, although some might disagree.



We need to recognise that those are different kinds of questions. We can prove mathematically that six 8s are 48 but we will never prove mathematically that you should wear a tie.

When we are building capital models, a lot of the time we are, I think, following some sort of social convention. We know that the result is going to be benchmarked against other insurers and we try to do something that is going to be socially acceptable and not raise the eyebrows of rating agencies or regulators. When we obtain a result by consensus benchmarking, there is a danger of overstating the scientific content. We might say “I have not much data and I have done something which is broadly sensible and socially acceptable within the community that I am in. But I need to recognise that a social convention like wearing a tie is not a result that I can prove with pages of theorems.” Sometimes, on the other hand, there are results which we can prove with pages of theorems. We have both of those kinds of results in this paper. It would help our communication I think if we are very clear about the distinction between technical proof and social convention.

**Mr J. G. Spain, F.I.A.:** I feel very diffident about speaking this evening because I know so little about capital modelling. I come from a final salary pension background. I am very much afraid that we are going to see some of this applied to final salary pensions in the not too distant future. It is scary because I had the privilege of being here this afternoon for the workshop. It was very informative. There was a lot of material I realised that I did not know which I thought I did. I wonder how many other people think the same way.

The answer to your question about lack of data is, if you do not have much data, a regulator may say: “You should not be writing this business. You should have a much better grasp of what it is you are modelling.”

I will go back to final salary pensions. Final salary pensions, as we used to know them, are not a short term environment in which to work: we do not emphasise one-year projected outcomes. For the profession, I am also scared that by concentrating on the one-year outcome, because that is what the regulators want us to do; we are potentially going to provide the wrong answer to the wrong question. Insurance companies, particularly life companies, and some general insurers with a longer tail, should be thinking much longer term. The paper is excellent, but if Solvency II is brought in it is going to divert people to the wrong solution.

**The Chairman:** I recognise that we have some members of the Prudential Regulatory Authority (PRA) and Financial Conduct Authority (FCA) staff here this evening. You might feel a bit embarrassed to speak, but Mr Spain threw out some interesting challenges to us about whether concentrating on a one-year VaR can in fact create instability and that, if you are looking at a run off situation, you might not want that. Does anybody have any strong views on this topic that they would like to share with us?

**Mrs K. A. Morgan, F.I.A.:** I am from the PRA. I want to discuss the one-year issue. Yes, the Solvency II capital requirement is calculated as 99.5% one-year value at risk of the basic own funds calculated according to the Solvency II balance sheet, but that is a capital calculation.

I have said this quite a few times, but not really in a pensions environment. The rest of Solvency II is about pillar two and three. Pillar two is about governance and internal controls. Pillar three is reporting around the whole of the balance sheet and lots of other information.

The key part of pillar two is the Own Risk and Solvency Assessment which specifically says it should be based on an insurer's own risk appetite and over the time period that they think is appropriate. So there is no requirement in Solvency II to say you should work out a one-year figure and only a one-year figure.

The point of the solvency capital requirement is to give supervisors a point where if the own funds go below it then we intervene, and if own funds are above it then we do not intervene, but we keep an eye on where you are relative to that point. That is one indication of the financial strength of a firm, but we also look at more qualitative aspects of the controls around the firm. I just wanted to clear that up.

**Mr B. Bergman, F.I.A.:** I want to talk a little about the communication of uncertainty. I have been involved in risk management for over eight years now. If I think back, I detect a definite reluctance on the part of risk management teams to make management aware that their all-singing all-dancing capital model on which they calculate returns on capital and do all their decision making is not actually quite as reliable as they may have thought.

I recall sitting at my own performance appraisal a good number of years ago and I got a bit of a ticking off for alerting people to the possibility that our model could quite easily be wrong. I was told: "People believe in this model. If we say it is wrong, it undermines us and the model."

A few years later I asked an expert in natural catastrophe risk modelling how confident he was with the probable maximum loss (PML) estimates coming out of the model. "A factor of Pi" I was told, which gives one an idea of the uncertainty surrounding the model. As we have seen from the paper, there is so much uncertainty that we have to deal with in our capital models: parameter uncertainty; model uncertainty; risk factor uncertainty; not to mention all the other decisions one has to make. The numbers coming out of the model could plausibly lie within a huge range. But how can we tell management that this number that we are calculating is plucked from a very wide range? This would really undermine what we are trying to do! For these reasons I think there is a tremendous amount of reluctance on the part of risk management teams to tell management exactly how uncertain are the numbers coming out of their model.

**The Chairman:** That is quite a challenge. I wonder how that will play out in Solvency II if a Solvency Capital Requirement (SCR) is calculated by an internal model. I know you are caricaturing it a little, but, if these numbers are completely worthless, then it is quite difficult to pass the use test, one would have thought. Does anybody want to pick up Mr Bergman's challenge on communication?

**Mr H. T. Medlam, F.I.A.:** I want to blend together the first point that was made and then the point that models are all wrong but some can be useful.

When you calculate the capital number you can be wrong by a factor of two or a factor of three. Once you decide on that number, you have to make it useful. You make it useful by linking it back to the exposure inputs. If your input exposures change, then the capital model number changes at the end. It is useful as long as the capital that is calculated is sensitive to the input numbers, and as long as that is the case then you are driving the correct management decisions and it is being used in the correct way by management.

**Dr C. D. Pickup, F.I.A.:** I am glad that this point about there being not enough data to have any idea about what a one in 200 event is in most cases has been raised. It is really important. We touched on it and as a communication challenge it is undoubtedly very difficult.

I think Mr Smith's suggestion about trying to distinguish between things which we could be more or less confident about and things which are just a kind of convention is a useful point. I welcome that. However, even when we do have lots of data and we think that we are confident about what might be, say, a 1 in 200 event, there is still an underlying assumption that the future is going to be like the past.

Finally, I would like to disagree with Dr Pryor's point about testing the robustness of models, one of the tests being that if you make small changes to the input to a model, you should be more comfortable if you get small changes to the output.

This is not true. I commute, and if, when I came in today, I had left a minute later, I might well have been 10 minutes later into the office or I might have been 20 minutes later. Depending on the time of the day, if I were driving, for example, that could be magnified. One minute could mean 20 minutes or half an hour, depending on where the sudden blockage was that I hit. Whilst this may seem to be a trivial example it is not irrelevant, because if you think about lots of models, they have floors and ceilings to different values or elements are carried forward if they go negative. So, in fact, in many cases you would expect to see discontinuities if you cross certain thresholds.

My point is that this robustness test would be a false comfort unless there are other very good reasons to believe that we should be obtaining small changes to the output from small changes to inputs.

My view on models is that if you have not found any mistakes in them you are just not looking hard enough.

**The Chairman:** I understand the point that Dr Prior was making. I suspect if you were modelling where you would end up after you do a jump, and you are standing right beside a cliff, then you might get a very different answer depending from where you started.

I think Dr Pickup also makes a very good point. It underlines the importance of thinking through what it is you are modelling, and understanding the dynamics of what is happening.

**Dr Pryor:** I think that Dr Pickup makes a good point. Just to be clear, I was saying that if the model is not robust and you obtain very big changes in the results from small changes in the inputs, then you should be very wary about trusting the results. I do not think I said, and I certainly did not mean to say, that if you get only small changes in the results from small changes in the inputs, you should then trust the results. You probably should not trust them then, either, but for different reasons.

**Mr R. J. Houlston, F.I.A.:** I am going to make comments on what Dr Pryor has said. I agree that data is very important. I had not noticed the paper making significant comment on what data to include. Where the analysis of equity performance is included in the paper, because we like data, it goes back over very long periods.

Another thought is on the idea of actuaries not influencing the world. I would like to put forward the idea that we have been suggesting recently that bonds are better to match pension liabilities

than equities. I suggest this has probably influenced pension schemes' investment strategies. The pension schemes' increased need for bonds may have added to the boom in credit, leading up to the recent crisis. The basic ideas may not have originated with actuaries, but I believe there are times where the work we do can affect the wider economy.

**Mr Jakhria:** I have a couple of responses. One comment was around data and another was that in our models we may inadvertently, or otherwise, assume that the future is going to bear some relation to the past.

My apologies. Firstly, I should have highlighted the data issue in the summary talk. On the first page of the paper we discuss data issues as follows. There are two problems with data which are almost conflicting. One is there is far too much breadth of data. You can obtain data on almost anything. You can obtain data on the width of coffee beans for the last 20 years. What that means is that it is very difficult to distinguish the important data from the less important data. If you are doing a regression analysis, and throw a lot of variables at your problem, you will inevitably secure much better regression. Does that mean you are better off? Perhaps not, depending on the data. So, there is a huge amount of judgement involved in trying to understand how components are linked together.

Another problem with data which is quite important is that although there is a huge amount of breadth, there is probably not enough depth, even with some of our highly rated datasets, to narrow down the choice of models.

The other point was whether the past is similar or different to the future. I would draw attention to the book on 'black swans' where there is extensive discussion on this very topic. One of its themes is that something may happen in the future that is so completely unrelated to past events that all the previous models (built on past events) are useless. When interpreting that statement, one needs to be very careful. It may well be true that there are scenarios where that could happen. However, you would need to go back into your past to when you chose the model. It may actually be a model that you could have picked from the data i.e. a model that you could not statistically reject, even though it was not the most credible model at the time. You need to be very careful before you can attribute everything to being a black swan event.

**Mrs Morgan:** Listening to the discussion, I think we have this the wrong way round. Models are useful, and there are lots of problems with models. But models are everywhere. They are not just in insurance and banking. They are all over government.

One model that has been close to my heart over the last year is the assessment of the bids for the West Coast main line, and the problems that there were with the model used. There is a recent HM Treasury paper on the use of models in government which covers all the issues about governance, control and understanding. As far as I know, the armed forces use war games and other simulations when they are planning their manoeuvres. They are using models.

I do not think we should become too depressed about difficulties with models. We are the Actuarial Profession. We have the intellect to tackle these tricky problems. And somebody does have to tackle them. We, the actuaries, are in a good place vis-à-vis these models. We know the short-comings. We can communicate. We like thinking about problems, and we like improving things. We have developed a public interest type of thinking. I think that we should seize this opportunity and start

influencing government. Not to make particular decisions in a political way but to make better decisions so that they understand the pros and the cons of what they are doing, the shortcomings of the data that has been used and the shortcomings of the models that they are using. I think that we can do this. This paper is a major start. I say let us go out and change the world.

**The Chairman:** That was very helpful and has grounded our discussion.

**Mr Bergman:** I should like to make a few comments, this time on expert judgement in the context of calibrations. Some of the risks we have to calibrate, ‘mass lapses’ for example, are really difficult risks to calibrate. We have no data, yet we have to calculate a number. The regulator wants numbers! Hence we are required to produce a number, and justify it! The calibration is full of expert judgement. Would one approach be to say there is a standard formula out there which already has calibrations for many of these difficult risk types? Let us look at the expert judgements involved in the derivation of the standard formula calibrations and assess how our portfolio differs from the underlying portfolio assumed in the corresponding calibration of the standard formula. Based on these portfolio differences we can adapt the expert judgements made in the standard formula to derive a calibration suitable for our portfolio. Adopting this approach we would at least ensure that we come up with a capital number consistent with the standard formula, taking any portfolio differences into account. I wonder whether people have tried this and whether they have had any luck trying to go beneath some of the expert judgements in the standard formula.

**Dr M. C. Modisett:** To respond, the way I took the comment was that you could start from the standard formula and ask the question: can we reject this answer? Most of the uncertainty that has been mentioned in the table would probably lead to the answer “No, that seems to be as good a model as anything else.”

**The Chairman:** Maybe there is a distinction to be drawn between studying the risk for your own benefit and arriving at a capital requirement for regulatory purposes. I remember, with the internal model, one of the plans was there would be an incentive to do the modelling because you would end up with a lower capital number. I suspect that, in practice, that may not be achieved in all cases.

**Prof D. Leech (guest):** I wanted to make one or two methodological observations. The meeting is about how to model, or take account, of extreme events. It seems to me that it is not really useful to take more data. By gathering more data, we are observing more normal events, in a sense. We are not necessarily gaining evidence about the likelihood of extreme events. There is an assumption, a law of large numbers or central limit theorem there, which does not really apply when we are talking about extreme events.

The other point that I wanted to mention was a lot of the variables we are talking about are actually human variables. What we are actually considering is the modelling of the behaviour of human beings using probability models. I think we need to talk a bit more about how extreme events occur in that sort of situation.

If we think of the events of the past few years, the financial crisis, the credit crunch, and so on, which involved many extreme events, the origin of them was the fact that people suddenly changed their behaviour. So there was a discontinuity, a ‘black swan’, if you like. This is something that needs to be thought about. I do not know the answer. I think that it is not simplistic. It is, perhaps, too easy to think you can model human institutions like prices, asset prices, and so on, in purely

statistical terms without taking into account the essential, uniquely human, behaviour that is underlying. We are not looking here at natural events. We are looking at people. That is just an observation, really.

**The Chairman:** We have had quite a few suggestions for the next paper from the Extreme Events Working Party. Mr Frankland, would you like to say something as the The Chairman of the Working Party?

**Mr R. Frankland, F.I.A.:** I would like to say thank you to all the speakers this evening. I feel that we have had a very warm reception. I must admit that I was rather nervous that we might be in for a fair amount of hostility given that we seem to be put in the place of saying yet again that you need to be cautious with modelling, and capital estimates tend, because of the uncertainty of those estimates, to require capital add-ons to cover that uncertainty.

One of the things I have wondered about for a very long time pulls together what Mr Bergman was speaking about earlier in terms of relying on the standard formula but also relying on regulatory judgement. Are we, in trying to set a 1 in 200 year estimate actually asking the right question, in the sense that it is something that can be meaningfully modelled? Is a half percentile really the capital measure at which we should be looking? There are a number of alternative possibilities. They may be no less, or more, onerous in capital terms, but at the same time they remove a lot of the difficulties that we experience in this type of modelling. For example, setting parameter movements of, say, three times a one in 10 year capital adjustment might be a lot more robust, having fewer issues with the use of limited data to estimate those stresses. Also, such a definition is no less arbitrary than picking 1 in 200. There is nothing special about 1 in 200. One could use 1 in 300, or 1 in 100. They clearly have different implications. The smaller you make that number, obviously the lower the capital. But, at the same time, the more reliable is the estimate and the more meaningful are the numbers.

A separate issue relates to comparison with the standard formula. If an office goes down the standard formula route then all this judgement is taken away from them. My understanding is that they just apply whatever the standard formula is, right or wrong, with no explicit justification on their part. If we were looking for a model based on a standard formula, then I think that there is one approach which could be applied, and that would be to derive the standard formula on a published basis with published justification.

If the regulators believed the standard formula should be based on 1 in 200 years, there perhaps should be an add on for standard formula users to account for the fact that standard formula entities are not looking at the peculiarity of their own risks and overlooking risks not captured by the standard formula. To the extent that internal model companies identify and model their own specific risk variations against standard formula assumptions, they would adjust the standard formula, but only be required to hold the add-on components to the extent that they do not explicitly model their own specific risks. This might result in a situation where an internal model company can hope to see its improved modelling leading to reduced capital requirements, giving an incentive to improve risk identification and management.

**Mr Spain:** Statistics is about uncertainty and the answer is not a scalar, it is a vector. We should be saying to managers of capital organisations that we do not actually know the precise answer, but our best guess is, say, 2.58 plus or minus, and then leaving it to the managers of the organisation to make up their own minds on which number to use.

**Mr J. Waters (student):** Dr Hare, you mentioned the use test and that companies have to use their internal model for a range of uses, such as investment strategy and capital allocation.

As an industry, many organisations have taken the following view regarding the internal model: “The standard formula does not apply to us. There is not one size fits all. We need something specific to our business.”

A question for the authors of the paper is: do you think we also need different models for different purposes? If expert judgement is so important in influencing what decision will eventually be made, is it right that we should decide on our expert judgement at a company level and then use this in all our decisions? Or would you say we should make one judgement for investment strategy and one for risk management, and so forth?

**The Chairman:** I am afraid I am going to have to close the meeting now. Martin White is going to summarise the discussion for us. Mr White is a non-life actuary who was one of the early actuaries in Lloyd’s when non-life was in its infancy. He now works for Berkshire Hathaway.

**Mr M. G. White, F.I.A. (closing the discussion):** I should like to echo the remarks of those who have congratulated the authors on their achievements with the paper. I think the subject is a good example of the more you know, the more you know you do not know. That has been very evident in the discussion this evening.

As the authors explain in their conclusion, one of the triggers for the paper was the perceived failure of risk models prior to the events around 2008.

The paper is well structured with discussion of the overall modelling structural choices and judgements, followed by choices of which risks to model, then plenty on model and parameter error where the difficulty of the task is made plain. The section on errors introduced by calculation approximations is an interesting bonus and clearly reflects the depth and practical experience of the authors.

Right at the end of the conclusion, the authors set out what reads to me like an objective for the paper, which is that these techniques will allow actuaries to close the gap between the risks we capture in our models and those revealed in the wake of financial losses. So perhaps this should be the test of the paper. Will this paper help to close the gap? All the gap? Some of it? Most of it? I listened to the discussion with this question in mind.

I came to the meeting today with some of my own thoughts on unacknowledged, and therefore unmodelled, elephants which may be in our room, and I was interested to see whether they would be touched on by our contributors this evening. The answer was not a lot, though Mr Leech talked about the human drivers. I will touch on them shortly.

On judgement and the overall framework, which was the bulk of the discussion, people not familiar with the modelling, and that will normally include the clients for whom the exercise is being done, will have no idea how much the choice of framework influences the outcome. The paper does a really great job of illustrating this point. I think everyone’s remarks point in that direction.

I should like to start with the objective of the modelling. Mr Hitchcox explained that the ultimate clients are frequently the shareholders. It is their capital that we are using or releasing and we should

ensure that they are not misled. So what is the real reason that the modelling is being undertaken? I do not think that we can take it as read. Is it simply to get the model through the regulators? As an aside, Mr Frankland made the point that a one year, 1 in 200, target may not be the most intelligent regulatory objective. I have some sympathy with that but it clearly does not qualify as a criticism of the paper. The paper was about how do you make these assumptions? How do you make these judgements? How do you interpret what is coming out? Is the objective just to justify a number which is seen as the most acceptable to those at the top of a company? We had a number of comments which indicated constraint in that direction. How objective is the whole approach really trying to be? Given the huge power of judgement in driving the result, these are not minor questions and we should be very aware of the professional responsibilities that are involved in carrying out this work. That brings me to what I regard as a very large elephant that we may have in the room which is the context to the modelling work and the initial view of the state of the world with which it starts.

Section 2, at the beginning of the paper, refers to accounts and a true and fair view in accordance with accounting principles.

What if the accounting principles themselves lead to a serious mis-description of the world? A great illustration is the recently published report on HBOS by the Parliamentary Banking Standards Commission, which I do recommend as really worth a read. The bad debt provision signed off in the accounts for HBOS was less than £1 billion. On subsequent examination, new estimates came out with bad debt materially in excess of £10 billion. Any modelling which had an opening assumption that the accounts of HBOS represented a true and fair view in the sense required by the Companies Act, in other words that distributable profit be prudently determined in order to protect the company and its creditors, was bound to be wrong and so any capital modelling starting from that was going to give you a nonsense answer. So the insurance analogy, before doing the capital modelling, is to ask the question: “do we believe the start estimates for the assets and liabilities really are as prudent as mean estimates?”

A closely related question, and highly relevant to the use of judgement in modelling, is: “do we believe we have good enough understanding of how this business, and the things that affect it, really work?”

Without that understanding, judgement, as the paper and discussion have illustrated, can become a race towards locked-in accepted wisdom and ultimately potentially dangerous. It is not easy to deal with this.

Dr Pryor talked about the expectations on us as a profession. As a profession, I feel we should encourage challenge, both within and from outside, uncomfortable as it may be. Of course, meetings such as this achieve precisely that end.

On the choice of risk factors to model, there was very little comment, and I do not have much to say.

The section on model and parameter error was a very revealing part of the paper. Admittedly quoting from another paper, Currie, Richards and Ritchie, we see how different models of mortality improvement give rise to widely varying results. So wide that the mean output from two of the models of, in this case, an annuity value, was outside the 99th percentile from two other models. This is one example, but this theme runs throughout the paper, of how a deeper understanding of



modelling difficult risks gives not greater confidence but a depth of humility. The more we understand what we do not know about the world, the less we are going to be over-confident in our conclusions.

This takes us to the huge challenge of communicating our conclusions. There were, as there always are, a number of comments on this point. In the Solvency II context, I think it is unfortunate that there is so much process, and that terminology such as validation is used. We can appreciate the wish of the legislators, all too aware that people might cheat, wanting to put in requirements to protect against that possibility. But, I am a little doubtful as to whether those same legislators were really aware just how uncertain are the things that are the subject matter of the modelling, such as the financial condition of a complex financial institution. “Can’t you just use the best techniques to reduce the uncertainty?” they may ask. I think the answer to that has to be, “No, that is not how it works. The best techniques help us say, with some confidence, that the uncertainty is at least this big.”

One speaker mentioned how management may not want to be told how uncertain the estimates are. But, surely, it is our job to tell it like it is.

On the errors introduced by the calculation approximations, Mr Hitchcox, for one, emphasised the importance and referred to model validation requirements. But it did not attract a great deal of comment.

In conclusion, the paper reflects a great deal of experience and thought. As I read it through, I thought this is a superb paper of a really high standard. It contains a good mixture of explanation and example. I am in no doubt that the ideas in it will help to close the gap between adverse model outputs and reality. I do not think that the gap will ever be closed completely. All we can do is our best, which must include, as the paper implies, a recognition that there must be some judgemental grossing up for both the known unknowns (or the known simplifications), and also the unknown unknowns.

The subject matter of the paper is a very difficult area. Where the acts are not just acts of God but also acts of man, as Mr Leech points out, such as the behaviour of financial values, there is little theoretical basis for particular distributions. We are not observing the numbers of petals on buttercups, as I remember once doing at school. We are venturing into the unknown, where financial relationships between parties are always evolving, and where economic theories in popular use may actually shed more darkness than light.

Better understanding of the modelling process, which I believe the paper helps reinforce, has to inform not just those charged with doing the modelling, but also those charged with running companies, and those whose job is protecting the policyholders from failed promises. In my personal view, it argues powerfully for a principle of prudence in the running of companies, and in consequence that all the incentives, both behavioural and financial, that the system gives to the participants, are tilted towards prudence.

With that, I should like to thank the authors again for an excellent piece of work that I believe will be very useful to the profession in future and which may also attract, and I think deserve, a wider audience.

**The Chairman:** Thank you, Mr White. I would now like to ask Andrew Smith to respond on behalf of the authors.

**Mr Smith:** Thank you, Dr Hare. I should like to thank all of the contributors for their remarks about the paper. We are pleased with the warm reception that it has been given. I am going to pick up just a few points that have been made in the discussion.

First of all, I liked Mr Bergman's suggestion that one could set capital for mass lapses relative to the standard formula. A stress test for a mass lapse event might be an example of a social convention. Maybe we are guilty of misrepresenting this as the output of analysis that we all really know does not exist. Would we be in a better position, as an insurance community, by facing up to the fact that the stress test we apply is a convention and we do not really know what the right technical answer is. The world may put us in the box of brilliant statisticians, but that does not mean we have to pretend everything we produce is based on pure statistical theory.

There was an interesting comment from Mr Leech, about collecting more data just resulting in more normal events and not necessarily shedding light on extreme events. One of the best examples I have seen is the analysis of liability from nuclear insurance business. In this particular account, we found there had never actually been a nuclear accident resulting in a claim. We had liability data from slips, falls and employment disputes, but no nuclear accidents. We can all see how silly it would be to extrapolate nuclear accidents from data on harassment allegations and people falling over. For some of the risks we are looking at we do have a history of extreme events as well as moderate ones. The boundary between moderate and extreme events is blurred. In insurance models, we usually want to model a whole distribution. We do not know in advance what combination of equity moves, interest rate moves, lapse changes, longevity changes, liquidity premium changes etc. comprises the most plausible severe stress for an insurer. We do not have the luxury of focusing only on the extremes of a distribution because we do not know in advance what the critical combination is going to be.

Mr Waters asked whether we need different models for different purposes. Our paper concludes that focusing on just one model is a dangerous thing to do. Our regulatory regime forces a logical leap from *an* internal model to *the only* internal model. You have to come up with a model then have it validated and approved and signed off. You are discouraged from signing it off in a half-hearted way that says, on the one hand, we could use this model and, on the other hand, another model might also be good. We say it is unscientific to single out one model and to say that is the single internal model that on which we lavish all our attention. There is a danger that a single beautiful model compromises our awareness that there are many, many other models which could be driving our risks.

I am going to comment on Mr White's elephant in the room. I am not going to say anything specific about HBOS. But I have had some experiences of various organisations that have found themselves in messes and including numbers subsequently found to have been mis-reported. When you sift through the wreckage, it is pretty rare that you discover the losses came completely out of the blue and could not in any way have been foreseen if you looked in the right place.

Looking at the history of banks with bad loans, it is often the case that somebody within the organisation has a good idea that provisions are inadequate before this is externally acknowledged. The fact that that awareness was not reflected in the accounts is a problem of governance. We might, for example, try to address governance problems by ensuring challenging reviews take place, or by encouraging whistle-blowers. We need to be clear about what problem we are solving here. Although to the outside world the bad loans appear suddenly this is not necessarily a black

swan problem. It could be a problem that whoever is drawing up the accounts and signing them off is not aware of all relevant information, or they are not seeking it out.

I am going to finish by picking up another of Mr Leech's points about modelling human behaviour. I think the Actuarial Profession really needs to develop its own brand of statistics. Many of our established statistical procedures have been developed in the context of a discipline, but not an actuarial one.

The frequentist statistical techniques in our paper are based on the work of Ronald Fisher, and his particular interest was the genetics of crops. We teach the principles of statistics as abstract mathematics. It seems hardly relevant that Fisher was interested in crops. But in actual fact there were lots of things about crops such as being able to repeat experiments and construct randomised trials which do not apply to the risks that concern financial firms. For example, nobody suggests that we should raise interest rates to 30% to see what it does to unemployment. We cannot run those sorts of experiments. We cannot do randomised trials in the same way at all. In our workshop we also looked at the Bayesian approach which has been particularly popular with geophysicists and works well for them. They may not be able to repeat earthquakes to order but there are at least some underlying physical laws that are well understood. Actuaries feel as though they have to choose between them: Bayesians versus Frequentists; geneticists versus geophysicists. I do not think we need to make this choice. With an understanding of the social context of both statistical theories, we should seek a third way that is useful for our environment, rather than feeling that everything has to be copied from alien disciplines. I hope our paper sets a useful direction for that journey.

**The Chairman:** Thank you very much, Mr Smith. In declaring the meeting closed, may I ask you to thank the authors for what I thought was an excellent paper.

## Reference

Currie, I. D, Richards, S. J. & Ritchie G. P. (2012) A value-at-risk framework for longevity trend risk. BAJ, forthcoming.