

Heavy models, light models and proxy models: a working paper

Abstract of the London Discussion

[Institute and Faculty of Actuaries, Sessional Research Event, London, 24 February 2014]

Contact

Christopher Hursey, F.I.A., E-mail: chris.hursey@charasys.co.uk

The Chairman (Mr E. M. Varnell, F.I.A.): We are joined by the following authors: Mr Christopher Hursey, Dr Matthew Cocke, Dr Matthew Modisett, Mr Parit Jakhria and Ms Cassandra Hannibal. Mr Hursey will introduce the paper.

Mr C. J. Hursey, F.I.A. (introducing the paper): This is a working paper and there is further work to do. You may have noticed a couple of appendices are missing, more will be added.

We were drafted in to produce this working paper over a short space of time and we hope to obtain further insights for possible inclusion in the next version of the paper, which will be submitted later in the year.

I will give a brief introduction to proxy models before discussing various aspects of model choice and design, a brief look at specific models, case studies and then some closing remarks.

So, part one: an introduction to proxy models. We start with a brief history of modelling methods. It is fair to say that this history is not comprehensive and is probably more applicable to complex liability models. If we go back two or three decades, we still had analytic functions and formulae in widespread use because computers had not been widely adopted. With advances in technology, we then had an increase in the complexity of models. The increase in complexity was as a result of the actuarial profession taking advantage of the increase in technology: the advances that were being made in terms of computing power and the tools that were available for us to use.

So, we moved from functions and formulae, consisting mainly of commutation functions, into cash flow models. Those cash flow models were in computerised spreadsheets. That made it very easy for us to start to take advantage of, and analysing, things like path dependency. It also allowed multiple time points to be evaluated rather than just a single point. We could project a series of cash flows, draw out various metrics and do more in-depth analysis.

Later on, regulatory changes, specifically a requirement for the recognition of options and guarantees, led to a need for stochastic models. A stochastic model need not necessarily be a stochastic cash flow model, but given that we were working with deterministic cash flow models, that seemed a natural progression at the time.

One point is often overlooked: in one regard, this was a step backwards because we were no longer able to evaluate multiple time points. We stepped back, again, to being a single time point evaluation. In order to do multiple time points, you would require stochastic on stochastic.

So, where next? Driven by increasing regulatory and risk management demands, there has been an increase in the number of scenario results that we need to produce. We have gone from a position where we need to produce scenario results numbering in the tens: our base liability value and a certain number of stress scenarios. We have not seen just a tenfold or 100-fold increase, but in some cases a 1,000-fold increase in the number of scenario results that we need to produce. This has led to a return to functions and formula.

We now have a situation where the demand for scenario results has overtaken the technology's ability to supply. So, we have returned to replicating formulae and other proxy models. The irony is that these formulae are often less sophisticated, computationally, than those formulae that were replaced 20 or 30 years ago.

So, what is a proxy model? In the widest sense of the word, all models are proxies in the respect that they model something. We distinguish between models that emulate reality and those that simply model a more complex model. So, for our purposes, and for the purposes of discussion here, we define our proxy model as a model that emulates a more complex model.

We can take it a step further. It is an important point here to distinguish that we are not only emulating a more complex model, but, more often than not, we are just emulating the output of a more complex model. We do not actually look under the bonnet and see what is going on within that more complex model: we just look at a series of outputs and then try to approximate those outputs. It is important to consider that point because, if that output or the data we are using does not cover some underlying behaviour, then it would not be in the proxy.

Moving on, we have choice of model and design. We find it useful here, because there is obviously a large number of choices to be made in picking the most appropriate model, to classify most proxy models as replicating formulae. We find that this covers a large range of models. Everyone is familiar with replicating polynomials. We also have replicating portfolios. These also fit into this classification of being a replicating formula, where we have formula elements multiplied by coefficients. In the case of a replicating portfolio, the formula elements are assets or market instruments. Effectively, you have a replicating formula made up of basis functions, where the basis functions are assets rather than polynomials.

Specifically, we have a linear system of equations. In one example we have n risk factors. Then we have k basis functions, all multiplied by coefficients β , and we have a linear combination of these, which we need to solve to find the best fit across a number of scenarios.

It is important to consider that if there are more scenarios than there are terms in the formula, then you have an over-determined system and you will be very unlikely to find an exact solution. You will be required to find a best fit, which is usually least squares.

We find it useful to use this classification because once you fit models into this paradigm, we find you can identify the common issues. When it comes to calibration, you find, whether it is a replicating portfolio or a replicating polynomial, you are still trying to determine the coefficients. If you are trying to determine formula structure, it is the same problem, just using different basis functions.

It is hoped this provides a common framework for comparison. It is important to note here that not all models fit within this classification. We accept that, since it does cover a lot of the common models in use.

In terms of the choice of model – this is a recurring theme throughout the paper – it is very important to consider the use to which the model be put. Models tend to expand beyond their original use. What do we mean by that? Every model is designed to fill a specific purpose. When it comes to our primary model, a heavy model – we use the terminology “heavy model” as opposed to the proxy model – the building blocks of that model will be things based in the real world, such as cash flows and such like. When it comes to future amendments, or improvements to that model, they can be adapted normally, providing it has been designed well.

With a proxy model, you sometimes find that the building blocks of the model are not based in the real world. That means it can be problematic to adapt it for future uses. So you have to be clear from the outset, when you are picking the model, as to the uses you are going to put it to, because, sometimes, you might not be able to adapt it for other uses. Reasons for this include the building blocks not being present in the proxy model, or there are features of the heavy model that have been omitted.

We draw particular attention to models that have been built for capital measurement. A lot of the work that has been going on in proxy models has been driven by the need to measure capital requirements, and in particular the “1 in 200” for ICA or Solvency II purposes. And yet, although we are building these models to measure capital, we tend to focus a lot of our efforts on particular scenario results. We have found from observations and experiments that scenario accuracy is very different from distribution accuracy. You can have a model that provides very good capital accuracy, but is very poor at the scenario level. So, it is important to consider the use for which you design the model, and not use it inappropriately because it may be simply not good enough to provide answers at a different level of detail.

Complexity versus accuracy is another fundamental choice that you are making when you pick a model. Very simply, you have a trade-off. In terms of our replicating formula classification, you can think of it as follows: as you increase the complexity of each formula element, or increase the number of elements, you increase the complexity of the formula. It is hoped you will increase the accuracy. By increasing the complexity, you increase the run time, which is why you have the association between increasing accuracy and increasing run time.

Generally, as you increase the complexity of each formula element, if you want to keep the same level of accuracy, you should need fewer elements. If not, why would you do it?

This leads us to an important point. As you increase the element complexity you need fewer elements, which may have implications for the calibration because you lower the minimum number of calibration points that you require.

This illustrates that the trade-off between complexity and accuracy is not just in terms of the result. It also impacts on the calibration and the implementation cost.

There are various factors that we can use to validate proxy models, which are discussed in the paper. We have already talked about use of the model. We will discuss quality of fit shortly. Other issues are ease, cost and speed of implementation, which will all be influenced by the model’s stability.

Proxy models have a propensity to become out of sync with what it is that you are trying to model. This is a broad subject and needs more research, covering issues such as frequency of calibration and frequency of use of the model.

The final point is intuition. We talk at some length in the paper about intuition, and the model being intuitive. If you consider a replicating polynomial, it provides no intuitive understanding to the user. If the coefficient of equity squared multiplied by persistency has changed from 650 to 450, what does that mean? It does not have a real world meaning, whereas if you are using, say, for example, a replicating portfolio, the coefficient of a particular market instrument might have a real world meaning – for example, the moneyness of guarantees – and this can be drawn from a replicating portfolio in certain ways.

If we consider quality of fit, there are various statistical tests that can be used to assess the quality of fit of a model. Again, we come back to the issue of use. You must consider quality of fit in terms of the use to which the model is going to be put. We considered various uses, for example, daily reporting, where you might be interested in small movements. Then you have stress testing, where you would be more interested in specific risks being stressed individually, but by quite extreme amounts. So you have to think carefully about what we mean by accuracy. Are we talking about distribution accuracy, scenario accuracy or component accuracy?

With component accuracy we might be looking for specific risk components or we might break it into sub-products. There are various ways that you can break down the formulae.

This brings us to one of our key results, albeit with the caveat that further mathematical justification is required, on the distinction to be made between scenario accuracy and distribution accuracy. This is drawn directly from our case study on a replicating polynomial.

Figure 1 shows a scatter plot of 20,000 tests of the errors between the proxy and actual. The fit here is poor by any conventional measure. It ranges between $\pm 60\%$ and ± 60 million.

In Figure 2, we have ranked the actual results and then ranked the proxies and overlaid them on the same chart. You can see there that the distribution of results is very similar, despite the individual scenario results being very inaccurate.

Figure 3 will make this a little clearer. The orange line that you can see is the difference between the two lines in Figure 2, and is showing the difference between the ranked actual and the ranked proxy. The blue line shows the ranked errors. That shows you the difference between the scenario accuracy and the distribution accuracy.

How can a model be so inaccurate but the capital result so accurate? We talk about the curve of constant loss. This is not a new concept. Our interpretation of this result might be influenced by our experience of risks in one dimension. In one dimension, when we get to the tails, the errors tend to become larger and larger. It is tempting to draw the same conclusion in a multi-risk dimensional space. As you move closer to the tails, the errors automatically become larger. That is not actually true. You find that the errors at the tails are not larger because you are at the tail of distribution: it is because of error bias.

This is our assumption at the moment but it requires a rigorous proof. The assumption is that it is due to error bias. In one dimension, once you get to the tails, you are completely dominated by a

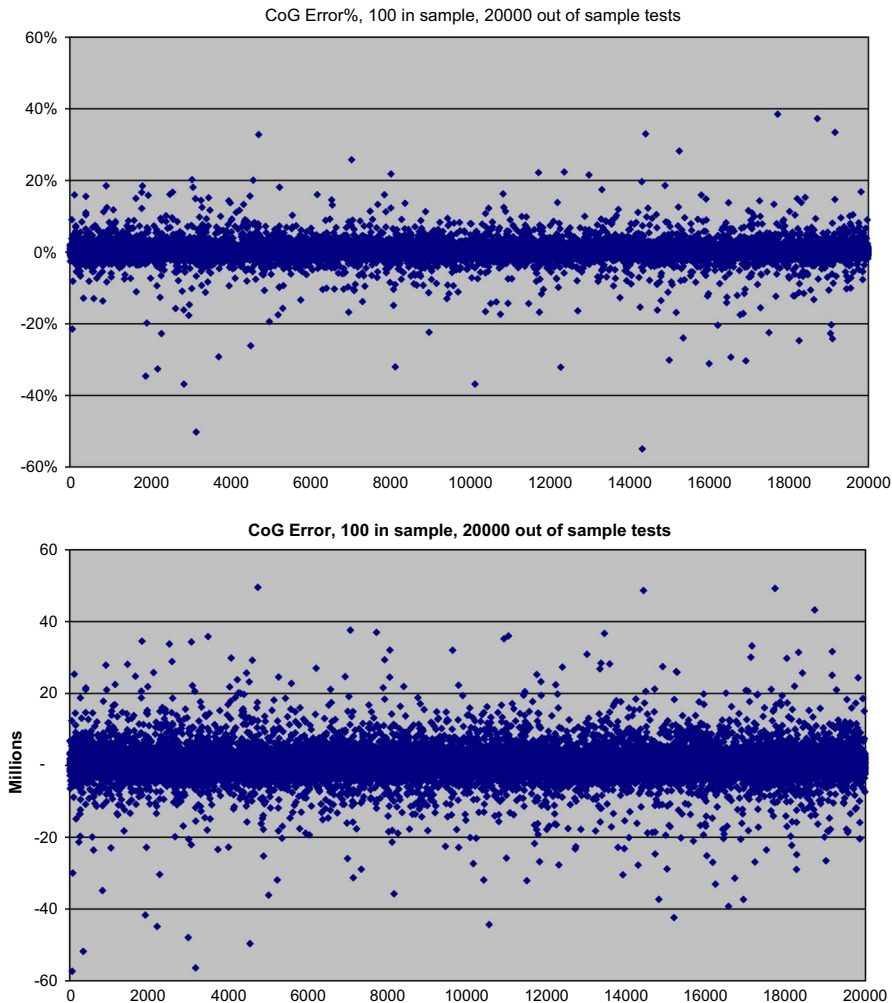


Figure 1. Scatter plot of 20,000 tests of the errors between the proxy and actual

single scenario. Once you move into more than one dimension, you have a constant loss. There are a number of scenarios producing the same loss. Along that curve, the proxy can have errors that can be positive and negative, in our case up to 60 million or 60%. The method of least squares, the formulae, the proxy, do a very good job averaging out the errors. We end up trying to minimise the error bias. That leads to our result: that you end up with a very good capital estimation even though the individual scenario estimation is very poor.

We did talk a little bit about calibration, again referring to the replicating formula classification. We recognise two stages of calibration. The first stage is determining the formula structure. This is the trickier of the two stages. It is determining which elements go into your formula.

For a polynomial it is: do you include X squared, X cubed Y , XYZ and so on? The number of possibilities is huge. As you increase the number of risk dimensions, I think the number of

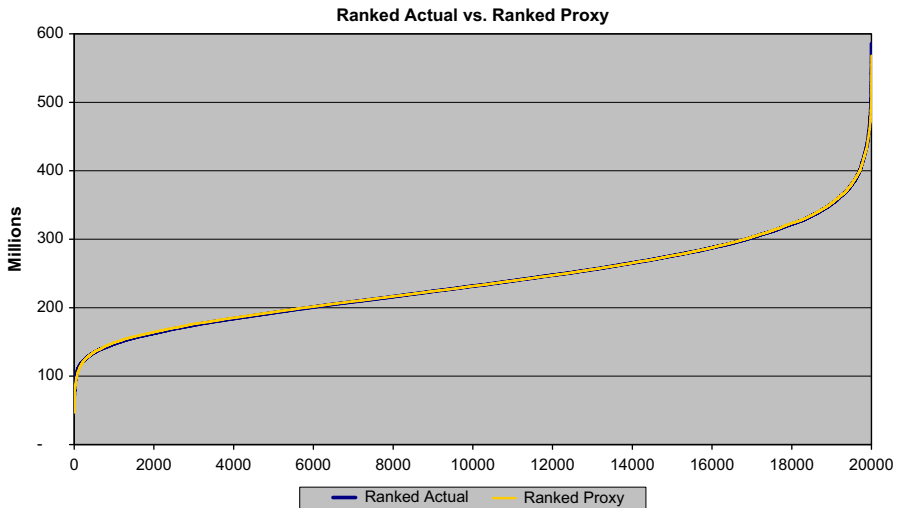


Figure 2. Ranked actual vs. ranked proxy

combinations increases exponentially. For a replicating portfolio, it would be deciding on which assets to include in the portfolio.

The second stage of calibration is determining the formula coefficients. There will be a target calibration, normally least squares.

It is useful here also to recognise the two environments in which the model will be run. We distinguish between the design environment where we are designing and building the model and determining this structure, and then the production environment. In the production environment you are often required to produce results within certain time constraints. We find that in the production environment often it is just a matter of calibrating to a given formula structure, although that is not always the case, since it depends on the methodology being used.

The design environment, even though you will be running both the first and second stages together, is done in an iterative refinement process, where you test the formula structure. You test it and then you refine it, and so on, until you obtain the formula structure that you want. The formula structure will often be linked to your method of calibrating the coefficients. We want the formula structure that minimises the square root of the sum of squared errors.

We talked a little bit about determining the formula structure. One further point to make is that we do find that there is a fundamental choice to be made between subjective expert judgement and objective automation. Again, there might be some simplification in this argument, and we should make it clear that when we talk about automation we are not necessarily talking about it being an automated computer. We are talking about the decision-making process, so we are talking about the use of automated decision-making processes, which, once you can codify a decision-making process, it then can be automated in a computer. But there is no reason why it could not be carried out manually. It is just that once you have codified it, you would be unlikely to do it manually.

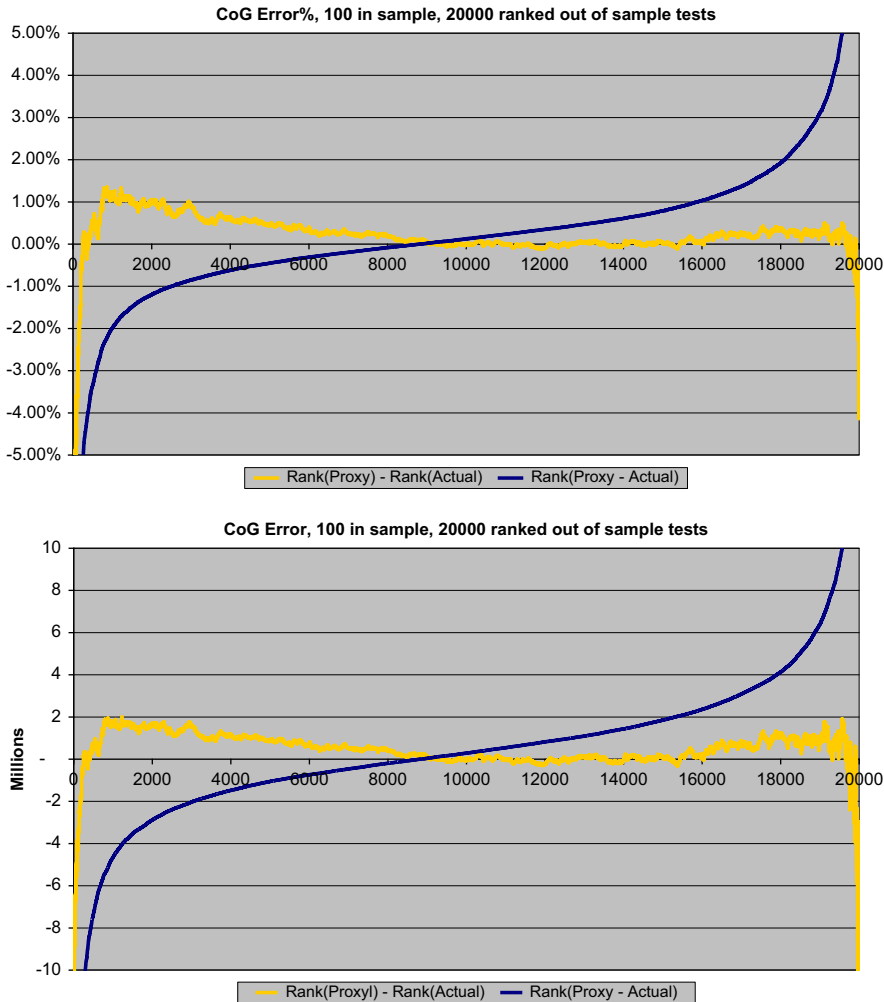


Figure 3. Scenario accuracy vs. distribution accuracy

So, effectively, you have subjective decisions that can be replaced by objective decision algorithms to allow automation. Automation allows us to trial a much larger number of formula structures, subject to having enough calibration scenarios to try all the structures that you want to try.

Moving on, we then have the second stage, which is determining the formula coefficients. It is a lot more straightforward as there is a lot less subjectivity. A lot of the time it is solving a least squares problem. There are other measures that you might wish to look at, such as minimax, for example. I think a lot of people have focused on the least squares problem, at least in the current software solutions that are being implemented.

As I said, this target calibration is linked to the decision-making process in terms of designing the formula. But then we will also find we introduce this concept that, when it comes to subjective

decisions, often what we think of as a subjective decision we are actually, consciously or unconsciously, applying a weight to various outcomes. If you look at it in that way you realise you can then just change from subjective to objective, and realise that it is a problem of solving a weighted least squares. There is still subjectivity involved in terms of translating your unconscious weight attribution and turning that into a function that can then be put into a formula and solved. But if we can solve a weighted least squares problem, then it provides, perhaps, a means of codifying subjective decisions in a format that can be solved and implemented.

It is important here to consider the distribution of calibration scenarios. You will only obtain an unweighted least squares fit if the distribution of the scenarios, the calibration, is drawn from a uniform distribution. If you drew all your calibration scenarios, for example, from a normal distribution, the greater density of scenarios in the centre of the distribution would effectively provide a weighted least squares fit, even if you did it in an unweighted least squares way.

There are further choices in calibration with regression fitting versus precise interpolation. As we saw with our replicating formula, if you reduce the number of scenarios to equal the number of formula terms, you end up with an exact solution subject to certain conditions. This will be very sensitive to the points that you pick and there are ways and means of picking the best points.

You also have to choose between optimal components versus the optimised whole.

We find that, if you optimise the components, the greater the impact of non-linearity, then the further from optimal your whole solution becomes. And, vice versa, if you optimise the whole formula, the greater the impact of non-linearity, the further from optimal the components become.

This is important because if you have optimised the formula as a whole, then you use that formula to test a single risk stress of the 99.5th percentile, for example, you may obtain larger errors than you would if you had optimised individual components, and that can be significant.

We talked briefly about least squares Monte Carlo. The main point here is that we do not consider this a different type of model. It is worth mentioning because it is growing in stature and use. Least squares Monte Carlo is normally implemented to calibrate a replicating polynomial. It is a method that is characterised by its method of calibration rather than being a different formula type or a different model type.

We did some work on case studies, which is as yet incomplete. We have included in the case studies two of the model types we looked at: replicating polynomials and radial basis functions.

In order to do this work in the real world we are unable to test every scenario. It would be useful if we could test all 100,000 scenarios, but normally we have a limited number, maybe 300–500, and then from that we have to extrapolate and infer conclusions about the whole risk distribution or the whole scenario distribution and the distribution of errors. What we were trying to do here is think what would happen if we could test every single scenario, and then, it is to be hoped, rather than infer results, we can actually draw some firm conclusions.

So we built what we termed a semi-light model, where we replaced a stochastic cash flow approach with a closed form solution. It is purpose-built for our research. It included nine risk factors. It covered a single product, and we modelled asset share and the cost of guarantees.

Briefly, the replicating polynomials and all metrics tested, as you would expect, with a regression fit. The quality of regression fit improves as the number of calibration scenarios increases.

We found that the fit-to-asset share was very, very good, which you would expect because of the linear nature of the asset shares. We tried not to focus too much on the asset share. It was so large compared to the cost of guarantees. If you looked at percentage errors, and if you looked at the total liability, there was no valuable conclusion to be drawn. What was interesting, though, was that we did find that the interpolation fit achieved near-optimum results. We had a 63-term formula in this example: 63 terms and 63 precise calibration scenarios. We were able to achieve results on a par with a regression fit using between 400 and 500 points. This is useful if efficiency is an issue, particularly for with-profits. Interpolation, if done in the right way, can obtain good results.

We also found that we had good capital accuracy across the board whether we were using 100 regression scenarios or 1,000, even where the scenario accuracy was poor.

We found radial basis functions did obtain better accuracy than polynomials for the cost of guarantees, but not the asset share, which is interesting. I think the linear nature of the asset share meant polynomials provided a better fit.

Briefly, a word about commutation functions. The hope here, I suppose, is that this is a generalisation of the replicating formula idea. You can have a more bespoke function that includes assets from replicating portfolios, but then include basis functions based on actuarial functions, such as persistency assumptions, mortality assumptions, interest rates and yield curves; 25 years ago, we used commutation functions with a level interest rate. We can use technology to calculate commutation functions on the fly using variable yield curves and various techniques to generate a range of yield curves.

Some closing remarks. As I have said, this is a working paper. We have not drawn any firm conclusions. What we have found, though, is that there is a recurring theme throughout the paper. It is that any design, choice or implementation for a calibration of a proxy model involves making a series of compromises.

If the question is: what is the best proxy model? I am afraid that is something we cannot answer. The use of the model is a recurring theme. It might be more appropriate to ask: which is the best model for a particular use? Then again, the tendency to try to extract maximum value from a model means we then want to adapt it and use it for other purposes. So the question might be: what would be the most flexible or adaptable model? Unfortunately, these questions are not easier to answer.

It is hoped that what we have tried to do is provide a flavour of the issues, which need to be discussed and addressed, and we have tried to pull it all together for further discussion and research.

Going forward, the trend has been that as technology has advanced we have increasingly wanted to take advantage and build ever-more complex models. We do not see that stopping. As knowledge of the problem increases, and we become more adept at using proxy models and understanding them, our hope is that we can take advantage of the technology in increasingly efficient ways.

Finally, we wish to develop more sophisticated proxies. A lot of the work is being done in polynomials. There are limitations with polynomials. We hope we can develop basis functions which

will draw more on our knowledge of the complex model which we are trying to emulate, so that we build basis functions that better match the problem, and we hope to see some gains in that area.

Mr D. Georgescu, F.I.A.: There was little in your presentation on final validation fit on out-of-sample testing as opposed to regressional in-sample fitting. Are you comfortable that errors average out, and would you consider giving more emphasis in the paper to out-of-sample testing?

Mr Hursey: No, we are not comfortable. Your question really does cut to the heart of the problem. It is a good question. The advantage that we have had with our case study is that we have been able to evaluate 5,000 scenarios, which, in reality, you cannot do. The trick now is to work out how we can take that observation across to the situation where you can only evaluate 500 scenarios and then draw the same conclusion. We are not quite there yet, and there is further work to be done to build a firmer mathematical foundation.

Mr P. C. Jakhria, F.I.A., C.F.A.: A lot of the logic that we have been applying applies to out-of-sample tests as well. The work that we are doing asks what type of basis functions best fit the nature of the problem we are solving. In that way we would minimise the total computer time, including both in-sample and out-of-sample testing.

Our aim is to make the problem as easy to solve as possible. In practice, you will need to carry out out-of-sample tests for every in-sample exercise for validation purposes. Choosing the best basis functions in the first place, which is our current focus, allows you to (i) reduce the number of out-of-sample tests and (ii) reduce run times.

Dr M. Cocke, F.I.A.: The approach to out-of-sample testing is an interesting question. There are a number of possible methods of generating the out-of-sample points. Our research so far has focused on generating these points from the underlying risk distribution. Other methods are possible, but it is not an area we have started to look at in any detail at all.

Dr M. C. Modisett, F.I.A.: Just to be clear, I do not think that we are trying to be prescriptive to say that these particular functions will work in all situations. We have done a case study.

We are trying to describe a journey of how somebody might look at their own risk profile and how they could start analysing it to see if this methodology would allow them to come up with some functions that would be good for them. We are not trying to say some functions work and others do not: we are trying to display a methodology for checking whether or not your function is going to fit.

Mr A. D. Smith, F.I.A.: I should like to thank the authors for the useful summary they have put together of current practice. As they say in paragraph 7.1.2, they illustrate a lot of methods and none of them can be demonstrated to be superior to the others.

I was a little bit uncomfortable with the extent to which professional judgement seems to be necessary on what strikes me as a mechanical task. We have a lot of crossing fingers and hoping for convergence when we ought to be trying to prove theorems.

I should like to see a lot more consideration of the mathematical conditions under which we could claim that these methods converge to within a given tolerance and with a certain number of simulations and a certain polynomial order. There is a list of diagnostics in paragraph 4.2.1. I did not

see any link between those diagnostics and how confident you can be in the accuracy of what your model says is a 99.5th percentile.

I had hoped for a synthesis of some of the results. What I actually found was Appendix 1, which is entirely blank. That is a stark reminder of what has actually been proven and is somewhat more candid than the software vendor promotions, which I have encountered.

There is a real need for some solid theoretical work here, especially in the area of proxy model error, which has also been called spanning error. The example in section 6, you have mentioned already, had nine risk drivers. The maximum order of two of any of the risk drivers. It is quite similar to an example in the paper by Frankland *et al.*, that you do reference, which was discussed here about a year ago. I encourage you to release the model points of the formulae you have used for the two models, which Frankland *et al.* have done, so others can comment on that and try to replicate the results which you obtained, some of which are quite interesting.

Mr Georgescu has already pointed out that as none of the proxy methods, so far as I am aware, can actually be proved to work, the brute force of sample validation is really important.

In paragraph 4.3.9, I think you agree that although we have become good at making the initial fitting more efficient, all that does is to put validation onto the critical path. In your example, looking at two-way interactions, there are potentially 36 handkerchief plots of which you have shown a subset, and for each of those you estimate four non-linearity coefficients. Suppose you have interactions of up to fourth order, then you have 126 handkerchiefs plots, each of which requires a five-dimensional printer to print, and nearly 3,000 non-linearity coefficients to estimate.

I wonder whether that is approaching the stage where it is so onerous to do the validation that it actually defeats the saving that you had from building a proxy model in the first place. I welcome the views of the authors on whether even that would be sufficient, that is, can you ever get to the stage that you are really validating the model without having done as much work as just running the full heavy model.

Mr Hursey: I cannot disagree with any of your points. They are all valid, specifically the last couple of points that you mentioned about the 3,000 non-linear coefficients. Again, that really is the thrust of the problem. I have alluded to the limitations of polynomials. As you increase the number of risk dimensions and then increase the order and then you look at the interactions, very quickly the light model does not seem so light any more. You need to produce 3,000 heavy model runs to calibrate it. It does defeat the whole purpose.

The general theme in the paper is to allude to these things without being critical of any one model. I think we could be a little bit more explicit in the limitations of replicating polynomials. We will review that part of the paper.

Dr Cocke: There are a number of convergence theorems. For example, if the risks are in a compact space and the unknown function you are trying to estimate is sufficiently smooth, then as you take more terms in the approximation, the approximation will converge to the unknown function under certain circumstances.

Those types of convergence theorems specifically apply to polynomial approximation with fitting points generated from the roots of Legendre polynomials. However, if you use uniform fitting points,

those results fall over. So if you are interpolating using a polynomial, using Legendre polynomial roots as fitting points is quite an important part of the toolkit.

There are similar convergence results for radial basis functions – the unknown function needs to be sufficiently smooth and the risks come from a compact space.

Where the academic research seems to be slightly light is in answering if you are doing one of these approximations, and then doing a stochastic simulation to estimate Value-at-Risk, what is the combined error from the error in the proxy model combined with the stochastic error from having potentially not enough simulations? It is to be hoped that Appendix 1 of the report will fill one of those gaps.

Mr G. J. Mehta, F.I.A.: Should we have a greater number of stresses to calibrate the polynomial? We should try, once we have calibrated a polynomial, more out-of-sample testing to validate the effect. What is the authors' view of where is the balance?

Mr Jakhria: That is a good question. As I explained earlier, we are trying to minimise the total number of calibration scenarios plus validation scenarios by coming up with the best proxy model.

A further question is that if you take for granted whatever proxy model you have, what is the best split between the two? Ultimately, you need to optimise trade-offs in two dimensions. One is the total run time, and the other is the least squares error in the validation run.

What we found is that there are no easy solutions other than experimenting on a case-by-case basis. One of the other areas that we are looking into is, depending on how you define about what you are concerned, looking at your validation scenarios either uniformly distributed or weighted to other scenarios in which you are more interested. Carefully choosing the basis of your validation scenarios should help reduce the number.

Mr Hursey: I will just make one more point on that topic. On validation scenarios, there is an issue of calibration in terms of when you use regression or interpolation. You obtain greatest efficiency by interpolating, picking points in a specific way. There are techniques that allow you to use the same methods to pick the best out-of-sample points. The problem is you are very reliant on inferring from, however, many out-of-sample points you choose and that that is the maximum error. But all you can infer is that that is the maximum sample error. However, there are ways that you can estimate the points at where the maximum error occurs.

The problem is this. It is very simple in one dimension and two dimensions. But as soon as you reach three, four or nine, ten or however many, it becomes increasingly difficult to pick those points either for interpolation or for the testing of out-of-samples.

There are ways. Rather than just blindly picking 500 at random, you can pick specific points where you expect the maximum error to occur. Again, it is based on Legendre polynomials.

My observation is that, in terms of validation, there are efficiencies that could be made.

Dr Modisett: This is a point that Mr Hursey brought up in the presentation. The issue of having a good fit in one dimension is an issue. Then if you take the loss function and you try to fit it in higher

dimensions, there is a second issue. Which of those you are trying to fit is the best? That requires the expert judgement that a previous speaker mentioned.

We have this proxy function. It may be so good at one dimension. It may be so good at the second dimension. But the question that is asked for something like a Solvency II implementation is: how good is it at obtaining the distribution of losses?

You can look at one dimension and say that with this amount of error in our proxy function for the loss that might lead to this amount of error in our capital. When we do it with two dimensions, it might look worse because you are saying in two dimensions we have fewer points. In a larger dimension we could really be quite off in how, on a point-by-point basis, we match the proxy model versus the actual loss function.

Mr Hursey was trying to show that there is a third level of judging how good or bad is a model: not on a point-by-point basis in one dimension; not on a point-by-point basis in many dimensions; but considering number scenarios and setting out the loss distribution.

The loss distribution seems to be what a lot of people are concerned about these days. The surprising result was that, even if on a point-by-point basis the losses were very poor, somehow it all evened out. And this is the point for which we would like to derive some theorems.

Intuitively, it seems right. If you could say that the errors were somehow evenly distributed, the loss distribution could be quite accurate, even though the loss function point by point is inaccurate. However, it is quite elusive to come up with a mathematical proof because we do not have a base, standardised formula for a generic loss function for all different products.

Mr Mehta: We are modelling loss functions. Should we try modelling assets and liabilities separately and then net assets on an NAV approach? Should we directly approach fitting the losses to NAV, for example, in the with-profits case, or directly to asset share and so on?

One set of experts says if you model assets and liabilities separately, you might double up errors. You do not know whether the errors cancel out or not. At least your validation should be on NAV or asset share, or something like that. But in terms of Solvency II, you would always like to see assets and liabilities separately rather than NAV, because you would lose information if you want to create hedges, for example. Have you formed any view about this?

Ms C. J. Hannibal, F.I.A.: I think it ties in with use of the model. What do you want to look at? If you want to look at your assets and liabilities separately, then you need to consider fitting at the component level.

What you are actually interested in is how you are going to analyse the results and manage your business and use the results. That has to be your primary consideration.

Our research showed that there is not one great model that will work in all circumstances. You have to prioritise what it is you want out of your model. If you make that decision up-front then you will obtain your optimal result.

Mr Jakhria: I absolutely agree with Ms Hannibal. If your bottom line was to estimate the loss distribution of your cost of guarantees, I think that is what you should model.

A very subtle point is that it also depends on what functions you are using. I will try to give a simple example, using cost of guarantees for a with-profits product. If you go back to basics and consider the ideal shape of the function, and what behaviour you would expect that function to represent, you would define two behaviours:

- (i) As the asset share goes to infinity, the cost of guarantee goes down to zero.
- (ii) As the asset share goes down, at some point the cost of guarantees would be a linear function to asset share such that, as the asset share goes below the guarantees, it needs to be fully matched.

If you think of that function, it is asymptotic to zero on the one side, and it goes up to infinity on the other side.

If you think of all the possible polynomials, there is no polynomial that has the kind of asymptotic behaviour that will go to zero at one end and to infinity at the other end. All polynomials go up to positive infinity or down to minus infinity.

That is why we have found that, if you are trying to do cost of guarantees, polynomials are not a natural fit, which is why some people may argue that it is a better fit for the assets and a better fit for the liabilities, and hence you do that. However, that is sidestepping the question because you should be saying: First, what is the most important item for you? Second, given that, what is the most natural basis function to use?

One of the points Mr Hursey made, and which we cannot emphasise enough, is that, as the knowledge of your problem/liability increases, we hope that you choose more tailored basis functions to solve the same problem. They do not have to be incredibly complicated, but as long as they have the same generic properties you would go a long way towards good convergence.

Mr Hursey: I would like to re-emphasise another point on this distinction on optimising components.

We use an example in the paper where we split a formula into the marginal risk functions in optimising individual components. You can scale that up to any level whether you are looking at product level or even the whole balance sheet.

The same principle applies throughout: if you optimise your components, the sum will not be optimal.

It depends what you want. If you want granularity then you need to optimise the components. If you want the optimal total, then you need to optimise together, which causes a problem because you do two calibrations, one for assets and the other for liabilities, and then try to make them consistent. There is no easy answer.

We refer in the paper to the principle of optimising components and optimising the whole in 5.3 to 5.19 onwards. Then we have a very specific example of a two-dimensional polynomial in Appendix 2.

Dr Cocks: One other point on modelling of assets is that some companies are, in practice, modelling at the individual asset level. Closed form formulae are used to revalue the assets in the economic stresses which should be, it is hoped, reasonably close to the true answer. So, with assets, there is the possibility of not using proxy modelling but using some sort of semi-heavy model.

Prof A. D. Wilkie, F.F.A., F.I.A.: I found this paper very interesting because I have been doing stochastic simulations for a long time. But this is taking it a long step further than anything I have done in the past. I thought it was very good.

You mention minimax methods of fitting. If you are wanting to obtain all the points not too far out, then the minimax solution might be quite a good one. That is pulling in the furthest out error: minimising your maximum errors. Although the average error may go up, there seem to me to be advantages in that approach. If I were wanting to fit a polynomial, for example, or something like that, to an existing published curve, say, a table of q_x in mortality rates, I would look for a minimax error because I want the error of my worst case to be minimised.

Minimax might have multiple local minima, I suppose. That could be a problem. It depends on your function.

Another observation: I imagine that your behaviour at the extremes depends mostly on the extreme values of the underlying random variables that you use to generate things. Often, one may use a normal distribution. It may be more realistic to use something more complicated and slightly harder to fit. But a difficulty seems to me to be that whatever you do, it is very difficult to estimate the extreme tails of any distribution.

If you wanted to consider a 1 in 200 year event, it seems to me that you really need 400 years of data, and none of us have that. Even the meteorologists do not go back quite that far. I think it is extremely difficult, estimating the 1 in 200 tail, even though the supervisors might like us to do it. I am not sure whether it might be worth arguing for something different.

My final comment is on the slope of an option going asymptotically to zero at one end and asymptotically to Y equals X at the other end.

First, it looks like a hyperbola, which does precisely that, so there might be an advantage in looking at a hyperbola. It is also exactly what the Black–Scholes function is, with its normal distribution coming in. That is another way of calculating that particular one exactly, which, in turn, might be a way of approximating a more complicated position.

Mr Jakhria: I did spend a lot of time thinking what type of function is asymptotic to Y equals X at one end, and to Y equals zero at the other end. It took a while experimenting with various functions of the exponential and their inverses.

It did strike me, after a lot of work, that somebody had done all this work about 40 years ago, which led me to the Black–Scholes formula!

Dr Modisett: Your last point is one of the major things that we would like to push. There are polynomials that you can use as basis functions. But there are certainly other formulae such as hyperbolae and Black–Scholes. Black–Scholes itself is the difference of two cumulative functions.

Mr Hursey: In terms of the minimax, I think we have observed that, in a similar way that for least squares you use Legendre polynomials, with minimax you would use Chebyshev polynomials. You can pick particular scenarios that will give you the minimax solutions in one dimension at least. I have not looked at scaling it up to multiple dimensions.

As far as extremes are concerned, estimating 1 in 200, we do have a couple of people from the Extreme Events Working Party here. I do not know whether they have any thoughts.

Mr Jakhria: Some of the thinking is agnostic to probability, and there are methods which can help you estimate the calibration in a wide number of scenarios. If you can imagine a cross between an economic scenario generator and a wide-angle lens, so for the 1st year your scenarios go out very, very wide, and then they go out as they normally would for least squares Monte Carlo. What this allows you to do, in terms of your fitting points, is to be able to fit across a large number of scenarios, while at the same time using the law of averages or the central limit theorem to obtain an approximate result for each individual point in those scenarios. That will allow you to fit against a large number of extreme events without needing to put a probability on each of the events. All you are saying is that whatever distribution you have, you are simply stretching it further than you would have otherwise.

That is not in this version of the paper, but we are hoping to have more results on this aspect in the final paper.

Dr Modisett: I would emphasise one limitation of the paper. We are not trying to solve the 1 in 200 problem. It is definitely a known problem. The proxy model paper is not trying to invent 400 years worth of data. It has a very limited scope in a sense. We are trying to model that other model, so in some sense we are a little bit agnostic on the question. We are certainly cognisant of it. In some sense we are just trying to say if that a model was run to obtain its “1 in 200” and we are able to replicate that in our proxy model, then our job is done. It is up to somebody else before or after the paper to decide whether that original model was a good one or not.

So we are cognisant of the problem but it is not the issue that we are trying to tackle.

Mr Hursey: In terms of the case studies, I think you make the point about the asymptotic behaviour. We are looking at a case study for replicating portfolios and commutation functions.

One of the problems that we should draw out here is that we are almost guaranteed a good result. That is perhaps why we have not quite got there yet. In order to be able to run 5,000 scenarios, we have used a simplified complex model. We are still within our definition of a proxy, because our proxies are approximations for a more complex model. A more complex model uses a closed form solution, in particular, it uses Black–Scholes for the value of cost of guarantees. So once you are using commutation functions or assets or market instruments as a proxy, you are guaranteed a good fit because you are going to be using Black–Scholes to proxy Black–Scholes. You need to get over that before you can proceed.

Also, you mentioned the point of a simpler version. We have to accept that it is difficult to know how to pitch the level of the paper. As much as some people might want simpler examples, there are going to be people who want more complex and more mathematical examples. We will take advice before publication.

Mr D. M. Pike, F.F.A.: I am going to end up with a plea rather than end with a question. I agree that the acid test of the proxy model is the validation: how well it fits. If we are going to improve that, then I suggest that we need to pay much attention to how we transfer information from the heavy model when we are calibrating the proxy model. I think that that is just as important as what kind of model we use.

Heavy models have often been developed for a different purpose, for example, for shareholder value rather than for risk measurement. The paper also points out the tendency to take a single point in time, and there is also a tendency to look at the mean rather than the distribution of the stochastic results of the heavy model.

It seems to me that the route that offers most hope is that of least squares Monte Carlo, in terms of trying to obtain more information, and to capture more information efficiently from the heavy model.

So I would like to put in a plea for more work to be done on least squares Monte Carlo in the final version of the paper.

Dr Cocke: We are intending to include least squares Monte Carlo as one of the case studies.

Mr P. J. H. Smith, F.I.A.: I have no experience of proxy models at all. Indeed, before reading this paper, I had not even thought of what the issues were that are involved. There is a very considerable level of complexity needed in order to make major savings in running times.

I have a two-part question for the working party which follows up Andrew Smith's points on validation.

Are you aware of anybody who has developed a proxy model able to demonstrate Solvency II-level compliance with validation on statistical quality standards? If you are not aware that that has been done, would the working party consider addressing the practical issues that would be involved in so doing?

Mr Hursey: No, we are not aware of anyone who has achieved that end. And yes, we will follow up on that point, given a recurring theme here in terms of validation.

Dr Modisett: I do not want to say who has passed statistical quality validation or not. But I think when you are doing 100,000 scenarios on a whole company, most companies I have worked with are using some sort of proxy. This is the issue that if you use one of the big, heavy models, even a large organisation might be hard pressed to get 50 scenarios out of something like that. If you are going to sit down and test 100,000, they would have to have made some assumptions.

Mr M. H. D. Kemp, F.I.A.: I thought this was a good paper. It occurs to me that some of these problems are ones that other financial experts have also tried to tackle in different contexts. I would therefore encourage you to look further at the literature on some of the more complicated tools now being used for derivative pricing. We have already talked about Black–Scholes.

Two further approaches that might be applicable here that I see fairly regularly appearing in such literature are:

- (a) The use of transforms such as Fourier and Laplace transforms. In your context they could be thought of as corresponding to special sorts of basis functions (that have unusual characteristics relative to other ones that you have been looking at).
- (b) The use of weighted Monte Carlo. In this approach we have a set of simulations but instead of assuming that they are drawn at random equally across the probability density function, we deem them to have been drawn with non-uniform weights.

I think that you may find it particularly helpful to analyse whether a weighted Monte Carlo approach could be used alongside some of the other weighted techniques that you have described. For example, perhaps in the limit, a weighted Monte Carlo approach can be viewed as equivalent to a basis function that is concentrated onto a single observation, which would then require no polynomial fitting or the like. Weighted Monte Carlo might then offer a relatively simple way of shifting from a “heavy” model to a “light” model. Weighted Monte Carlo does potentially seem to be quite a general tool although some of the underlying complexity is merely hidden by the way in which you then have to choose weights to ascribe to each individual simulation.

Mr Jakhria: One interesting point that you highlight, where we have done some research, is on the subject of transformations. A lot of our thinking initially boiled down to choosing what was a nice basis function. However, in the event that we did not come up with a nice basis function, we thought about whether it was possible to transform it into a nicer function.

The reason for transforming it is that you can still use the linear set of equations and optimisation afterwards. There is work that we have done on this which, perhaps, leads you down the path of going towards a replicating formula and/or commutation function, because it turned out that the transformation we were trying to do increasingly looked like something we already had either in terms of option prices or even commutation functions.

We agree that this is an area for further research.

Dr Cocks: One point worth making about the range of techniques we look at is that least squares Monte Carlo does have its origins in derivative pricing. Specifically, it is the paper of Longstaff and Schwartz, which looked at pricing American options, which has been adapted to come up with a way of estimating Value-at-Risk.

Dr Modisett: On the weighted Monte Carlo point, I am not sure exactly the usage you envisage, but the targeted Monte Carlo example is usually trying to focus on one aspect of the distribution.

At the moment the focus of the paper is trying to see how well we can estimate the entire distribution of losses, but you could use the targeted approach to say that we would like to judge our model on the basis of how well it comes up with a particular portion of the distribution, such as the 1 in 200 part. We will consider that aspect further.

Mr Hursey: Just a comment on weight. I do not know whether it is the same point. We drew attention in the paper to the equivalence between drawing points from a uniform distribution and then applying a weight function. Is it the same as drawing the points from a particular distribution and then just doing an unweighted least squares. You get the same result.

For example, if you draw points from a normal distribution and then perform an unweighted least squares, you will get the same result as if you draw the points from a uniform distribution and then apply a normal weight function to it. You end up with Hermite polynomial roots as opposed to Legendre, for example. I do not know whether that is relevant.

Mrs K. A. Morgan, F.I.A.: Thank you very much for this paper. I very much enjoyed reading it. I have a couple of questions around the use of proxy models. You mentioned that choosing the proxy model is a trade-off between use, complexity and accuracy, which I completely understand. But does

that mean that, if and when the use changes, maybe because you have some lovely intuitive model that people understand and then they want to use it in different ways, that you then have to go back and change the proxy model and maybe lose some of the intuition that you have gained?

The other linked question is if you are very happy with your proxy model, does that make it harder to change your heavy model even though that might be a better reflection of what it is you are trying to do?

Dr Modisett: The answer is “Yes”. If you have different uses to which you are going to put the proxy model, it is an expert judgement. If we have calibrated your model to be good for one particular use, and then we use it for another purpose, you have to sit down and ask whether the proxy model is still appropriate. It may lead to a different model.

The other one is to think about when you are changing the original heavy model. In our paper, if you change the original model you have go back and revisit the proxy. However, with that said, it depends on the change. Maybe the old one works. Maybe you can still salvage it.

Mr Jakhria: A related point you might want to ask yourself is: how do I convince myself this same model will work next year? When you have rolled forward your liabilities, what guarantee do you have that it will work? I think that that is a question that a lot of insurers are finding is very practical and common, but difficult, question.

Will the same basis functions work? For example, if you had $Y^2 X$ as one of the functions, will you still have $Y^2 X$ if your coefficient for the $Y^2 X$ changes from -400 to $+17,000$?

This raises a very pertinent question as to what you want from the model. Do you want it to match liabilities at any one point in time. Do you want to re-do the validation within certain time intervals, or do you want something that is a bit more versatile? That is where our section in the paper on intuition comes in. You should really try, to the best of your knowledge, to choose the factors, etc., that intuitively make sense, and then try to create a basis with them. In fact, you may even want to give up a little bit on fitting error just to make sure you have an intuitive fit rather than a spurious fit.

If you throw enough variables at the problem, you will, by definition, get closer and closer to the answer. However, you may want to take intuition as a serious consideration when building your model.

I wonder, if we knew that we were going to do this, we would have built heavy models in the same way. The answer is probably not. If we knew that we would come up against the constraints of computing powers, we probably would have used cleverer maths to create nice functions and put less “if” statements and loops in our Monte Carlo code.

The problem is that insurers do not have a clean sheet of paper at the moment. They have a heavy model that has already been created. Within a short space of time they would need to come up with a light model that is within a certain tolerance of the heavy model. If you had the fullness of time and a cleaner sheet of paper, I think what you would do is just make a nicer heavy model that does what you want it to do and do away with the (supposedly) light models altogether.

Ms Hannibal: Just on the point of the uses of the model changing, it might be possible to recycle some of the calibration work that you did to get your proxy model in the first place. So it is not

necessarily a case that you would actually lose all of the knowledge that you have gained and all of the work that you have done. You would differently need to go back and re-evaluate whether the proxy model is still optimal for your new use.

Mr J.-L. Chauca: I should like to know if you considered using redacted techniques like common factor analysis based on common variance in trying to reduce the complexity of the models that you are using as a proxy?

Dr Cocke: I think that is an interesting idea. The dimensions of quite a lot of these models can be fairly huge. You can have over 100 risk drivers. So, reducing the dimensionality would be a clear win if you could manage it. It is not something that we have looked at yet.

Mr Hursey: I would just add, again referring to the content of the paper, that we very much view this as a jumping off point. We have tried to provide a broad enough paper to provide interest and to give a starting point, so it is hoped in the future to be more focused on specific ideas.

In that regard it may be that some of the ideas are a little under-developed and they do need further work. It is clear that we have barely scratched the surface. It is a huge, broad subject. That is clear from the questions today.

Mr Jakhria: On the dimension reduction, there are two key questions:

- (i) Do you reduce the dimensions before you start solving the problem, which is one way to look at the issue. This is not in the scope of this paper, but I do know that there was another sessional paper, I believe last year, on “difficult risks and capital models”, which had a chapter devoted to dimension reduction.
- (ii) Secondly, and perhaps this is a question that applies more generally to all actuarial models, can we do what we do but be a bit cleverer about modelling so that we do not have so many dimensions that we need then to carry into a 1 in 200 stress test?

Dr Modisett: On a practical issue, it might be difficult to do that. In a typical situation you might have a lot of variables and very few surplus scenarios in which to do some reduction.

The Chairman: It just leaves me to express my own thanks to all of the authors for their work and to the speakers for asking many interesting questions.