

---

## Model Risk: Illuminating the Black Box

---

### Abstract of the Edinburgh Discussion

[Institute and Faculty of Actuaries, Sessional Research Event, Edinburgh, 29 June 2017]

This abstract relates to the following paper: Black, R., Tsanakas, A., Smith, A. D., Beck, M. B., Maclugash, I. D., Grewal, J., Witts, L., Morjaria, N., Green, R. J. and Lim, Z. (2018) “Model risk: illuminating the black box,” *British Actuarial Journal*. Cambridge University Press, 23, p. e2. doi: 10.1017/S1357321717000150.

**The Chairman (Mr A. H. Watson, F.F.A.):** We have three speakers from the working party with us. First of all, we have Rob Black, who has a BSc in physics from Durham University and a PhD in atmospheric science from the University of Edinburgh. He started working in Edinburgh in the web development industry for three years before changing career to actuarial work. He started as a trainee at Scottish Widows in 2004 and qualified in 2010. He is now at Standard Life where he has worked on model production, model development and results validation. This has included a couple of enjoyable years working on the significant modelling process development required for Standard Life to meet its Solvency II reporting timelines and is now seeing the benefits of that post-implementation.

The model risk working party is his first experience of a working party, and it has been good to learn more about the internal workings of the Institute and Faculty of Actuaries and the role that volunteers can play in the profession.

We also have Louise Witts, who is a senior manager in the pensions risk team at Lloyds Banking Group. She has a maths degree from Bristol University and spent 14 years as a pensions consultant at Barnett Waddingham, latterly as partner and scheme actuary. She then broadened her experience with interim positions at the Pensions Protection Fund and in wider fields.

In 2011, she took up a position at SL Investment Management in a firm specialising in traded endowments and life settlements. There she managed the actuarial and risk teams and developed a particular interest in model risk and considering the ways models are used to support claims made in the marketing literature. In 2015, she moved to Lloyds Banking Group to work in the second line of defence with responsibility for risk monitoring of the colleague pension schemes, such as Defined Benefit schemes, currently £45 billion in size.

Model risk is still of key interest with challenges in providing meaningful information for management reporting and decision-making by non-specialist colleagues.

The final member of the working party is Andreas Tsanakas. Dr Tsanakas is a reader in actuarial science at Cass Business School, which he joined in 2006. Prior to this he was at Lloyds, mainly employed in capital modelling projects. Andreas studied electrical and computer engineering at the

University of Patras in Greece, and carried out his doctoral research in risk management at Imperial College, London. He also holds an MSc in Control Systems from Imperial, and an MA in Modern German Studies from Birkbeck College.

His research interests are in quantitative risk management with particular focus on measuring portfolio risks, capital allocation and dealing with model uncertainty. He is interested both in the quantification of uncertainty and in its limits. A recent emphasis relates to the ways in which models are *not* used in the practice of decision-making other than how governance can address model risk. He has published widely in academic literature while also contributing to practitioner publications such as the prize winning working party report Model Risk – Daring to Open the Black Box.

**Dr R. Black, F.F.A. (introducing the paper):** I want, first of all, to acknowledge the wider contribution of the working party. It has been my pleasure to chair the working party for the past 18 months or so.

There were about ten contributors to the working party. Quite a broad church: Andrew Smith, from Deloitte, will be known to many of you. We have representation from HSBC, Prudential and Barnett Waddingham. Rob Green is a consultant who was at Deloitte but now runs his own company, and there is Bruce Beck from Imperial College.

It probably helps everybody to have a quick run through the paper so we know what we are all here to discuss later.

I thought I would start with a definition of model risk, which is shown in Figure 1. This comes from the Federal Reserve System in the United States. It underpins all the phase 1 work which led to the original paper: Model Risk – Daring to Open the Black Box. I thought it was useful just to repeat it for the phase 2 work. So, at least, we all have a common understanding about the term “model risk”.

Crucially, it is a system for calculation but involves assumptions and judgement, so we are keen from the start to repeat that there is judgement behind what the model is doing, and equally how we interpret results from the model.

In that sense, calculator type models that are rules based are not models, as such, by our definition of models because they can be fully tested and checked, so there is no debate about what the model is doing and, ultimately, what comes out.

The concept of model risk is twofold: first, models may have fundamental errors in them and therefore may produce inaccurate outputs. But a rather insidious problem is that a model may be

In the words of the Board of Governors of the Federal Reserve System (2011):

- “[T]he term *model* refers to a quantitative method, system, or approach that applies statistical, economic, financial, or mathematical theories, techniques, and assumptions to process input data into quantitative estimates. A model consists of three components: an information input component, which delivers assumptions and data to the model; a processing component, which transforms inputs into estimates; and a reporting component, which translates the estimates into useful business information”.

The concept of model risk is therefore twofold:

- Models may have fundamental errors and may produce inaccurate outputs when viewed against the design objectives and intended business uses; and,
- A model may be used incorrectly or inappropriately.

**Figure 1.** Model risk: a definition

used incorrectly or inappropriately. So, we might have the best model in the world but we do not use it for its purpose.

So, there are two definitions of model risk which form the foundation for all of our discussions in the last 18 months.

We will look very briefly at a few case studies which are described in more detail in the paper, as shown in Figure 2. There is no shortage of case studies of models behaving badly, or models going wrong, but there are some interesting examples in academic literature. We have pulled out a few that we thought illustrated interesting ideas.

We have had discussions within the working party about model risk communication. We by no means have definitive answers to what moderate/good model risk communication looks like. I hope there might be some discussion from the floor about that later.

The third point is practical implementation of a model risk management framework.

Bruce Beck was at a workshop in South Africa about financial regulations, and came back to the working party quite enthused that we need to put something out there which talks about what it means in the day-to-day office environment; what are people actually doing? It is one thing to read the phase 1 documents; but what does that mean in practice? What can we do as businesses, as companies, as actuaries?

In the working party, we focussed on practical aspects quite early in the process. Equally, we recognised from the outset that actuaries are not the only people or companies that use models. There must be things that we can learn from other industries.

Bruce Beck has quite a background in environmental science. He has some interesting things to say about weather forecasting, as has Rob Green on the aerospace side. Andreas (Tsanakas) will touch on those later.

Our paper is really quite a practical paper, with no equations anywhere and so is not a technical presentation at all.

The paper started out, in many ways, as a write-up of all the topics we had discussed, which was then shoehorned into what we hoped was a fairly logical running order. We benefited enormously from people's different perspectives.

Loss of NASA's Mars Climate Orbiter (1999)	Root cause of failure was human error. But this was not the main issue. After investigation it is believed the main issue was the fact that the error had gone unnoticed despite a number of quality control processes being in place.
Cumbria flooding (2015)	Demonstrates the challenge of forecasting based purely on historical data. Record 24-hr rainfall. 1-in-800 event? Just what is an "extreme event"?
Growth in a Time of Debt (2010)	Errors in original calculations were not picked up. However results supported authors' prior beliefs. Hence appears to demonstrate the wider point that model checking and validation can be heavily influenced by prior beliefs and biases.

Figure 2. Model risk: case studies

Some of the case studies are recorded in more detail in the paper, the Mars Climate Orbiter is quite a long time ago now, 18 years, but this was a fundamental human error around the units used in the software, in particular between American and, I think, UK engineers. There was no debate at the time. Once the event had happened and they did an investigation, there was no debate at all about what the error was. NASA went into very great analysis about how all their processes, checks and balances had not picked this up before the mission set off. They came down to the view that at the time they were under a lot of time and financial pressures to complete missions to budgets. That clearly was a contributing factor.

What was interesting for me was in many ways putting a satellite around Mars had become routine. They just expected that it would be another successful mission. They had done it so many times that there was a feeling within the investigation team that they had taken their eye off the ball. From that, one of the straplines in the paper is that we should all challenge our own notions and even if things have worked successfully many, many times, it pays to step back and check that we are happy with what has been done.

Three studies are described in more depth in the paper. The Cumbria flooding is an example of an extreme event which was very difficult to perceive. Within a 24-hour period, they had the equivalent of 2 months of rain (this was in December).

For some of us who are familiar with one in 200 type events, how would we even quantify this one in 800? We talk about a one in 200 event but what does that really mean?

Then growth in a time of debt. Some of you may be familiar with this example. It is an academic paper published by two Harvard academics. It was picked up and used extensively to champion policy-makers' ideas on quantitative easing. There is a very interesting follow-up story which we picked up in that academics tried to reproduce the results and found that they could not. They eventually obtained access to the model behind the results and found flaws in the model. The very interesting thing is that while the initial results of the paper were severely diluted, we think it is quite a good example where, perhaps, the authors beforehand expected and wanted to find certain results. Had they found completely counter results, they would have, maybe, looked much more closely at the model, and so on.

We spent some time in the working party talking about the value of good model risk communication. What does it mean? What are good examples? I hope that can influence the discussion later.

We are really looking at two audiences: internal and external. By communicating well on model risk, what do we hope to achieve? Well, we hope to achieve better decisions, and, conversely, avoid poor decisions. We hope that it allows senior management to communicate more openly and transparently around results to our external audience and we hope it gives greater visibility and profile to actuarial teams in organisations.

From our own experiences, what are better companies doing around internal communication? There is much more activity on the enterprise-wide risk side. We certainly see evidence of model inventories and model materiality frameworks.

We obtained a good steer in this area from HSBC, where a lot of effort is being put into championing model risk internally.

Equally, externally, what do we mean and want by communication? The universe of people that we might talk to is endless: regulators; shareholders; clients; analysts. In fact, anyone relying on the outcome of the model.

We feel it is difficult to communicate externally on model risk if you focus only on the models you have and the uncertainty in their results. No doubt that would lead to further conversations, especially with the Regulator. So, you have to be quite nuanced in your approach to communicating externally.

The main thing for me is the practical implementation of the model risk management framework, bringing to life what companies are doing that I have seen, and colleagues on the working party have seen.

First and foremost, an essential is a central inventory of core models. HSBC, anecdotally, have a global database of every model they use. That sounds a very impressive position to be in, no doubt with a lot of time, effort and money spent in reaching that position. We start from the outset assuming, for any reporting period, internally we need to know exactly which models we are running and to have confidence that they have been through all aspects of testing and sign off.

Crucially, from a personal perspective, we have to have evidence that the right models are being used. This is where it becomes very interesting between the theory on the page and the practicalities of people doing their day jobs. We need frameworks in place to ensure that you cannot use an out of date model and if auditors want evidence that you have used the right, latest, signed off model, that that can be provided at the right time.

So, the central inventory of core models is a must, we felt. Equally, we felt it was very helpful to start a discussion around model roles. Named individuals will have responsibility for certain models. But within our working party we probably all have a different idea of who the model user is, who the model owner is, and who the model approver is. We think it is helpful for companies and consultancies to think through who wears these hats. Clearly, people can wear several hats at the same time. In terms of assigning roles and then responsibilities, it is helpful for organisations to consider who does what within this framework.

In terms of third-party model software, I am aware that many life insurers use a well-known brand of economic scenario generator (ESG). Is that a good thing? Do we, inevitably, then all end up with the same view of the future, whether that is a good or bad view?

The paper talks about one of Moody's marketing points. If you use a Moody's ESG, the regulator is very familiar with that. Lots of other companies do. If you use an unknown, unbranded ESG, what will happen? Of course, the regulator is going to take a lot more interest.

The third-party software can be seen as: "Another team, in a separate room". The fact that we are buying in a service from a third party, gives us a core piece of kit without which we cannot do many functions across the business. That does not absolve us in any way from fully understanding how that third-party software works. Really, we would just want to treat it as another in-house model with all the routine testing and sign off. This is quite a challenge: if you really go to the *n*th degree and you want to be as comfortable with all the third-party software we use, as well as our own software, that is quite an ask.

The final comment for me is around the culture challenge. The first phase 1 paper talked about four different types of model users.

Many of us think that building bigger and better models is the way forward. Equally, though, there are some people who, for all very valid viewpoints, like to make decisions from the gut, if you like, from their experience of 30–40 years in the financial services industry, the way that they see markets go, regulations change and so on. We are very conscious that when we think of a model user and the recipient of a user, it is a wide spectrum.

I am sure that you will recognise that there are some individuals who are completely key to the success of model development, whether that is because they know the methodology better than anyone or they just have the expertise to code it. Is that a good or a bad thing? Personally, I think the more transparency that we have around how models work, and so on, the better.

In terms of social pressures perhaps the higher we go up in the company, the less senior management want to hear about model issues. They might be discussed fiercely within the development teams, but does that bubble up?

In terms of balancing model change and innovation, within the model risk framework, we all accept that stage gates, and following a well-established process, is the thing to do; but equally we are aware that human nature, a lot of well-meaning people working in companies want to fix bugs in models as and when they see them, so we put that out as a challenge.

I will pass on to Andreas to mention a bit more about the weather forecasting and aerospace.

**Dr A. Tsanakas:** As part of the work that we did, and this is also something that followed from feedback from the phase 1 work of the working party, we were asked by the audience at the last sessional event to look also at other industries that use complex quantitative models, and how they dealt with model risk. Actuaries are not the only ones who have to deal with such issues.

We looked at a few different fields. I will talk only about weather forecasting and the aerospace industry.

Of course, there are many parallels to actuarial work and there are lessons that can be learnt; but also not everything transfers between fields, for very good reasons. It is important to understand the parallels between fields but it is also important to understand the differences. But it also points to, let us say, limitations of our approach.

Starting with weather forecasting, one thing that is very noticeable with weather forecasting is that big errors in weather forecasting are public. You find out about them. There was the well-known “Michael Fish” moment. When Andy Haldane gave a lecture some time earlier this year, I believe, he mentioned the inability of economists to predict the financial crisis as a “Michael Fish” moment for the profession.

Meteorologists probably have good reason to be annoyed by the comparison, given the accuracy of weather forecasting has increased spectacularly over the past 30 years or so. Economists could not say the same.

There are several reasons why prediction accuracy has improved that much in meteorology. One big part of it is that we understand the system in a very fundamental way. The weather is Newtonian mechanics. This combined with rigorous data collection and computational power leads to better and better predictions up to a point. Weather is a chaotic system. The development of weather patterns depends strongly on initial conditions which are never perfectly known. Still, there has been an enormous amount of progress.

If you compare this to actuarial work, the first thing to notice is we have no laws like Newton's laws. You could say that we know something about the law of large numbers and maybe can speculate that "no arbitrage" assumptions mostly hold, but never when it is most interesting for them to hold. So, we are not quite in the same place.

What also complicates things more is that, when you are talking about practices like weather forecasting, the data drives the forecasts but the forecasts do not influence the weather in the same way as writing insurance policies does not cause earthquakes.

But, much of financial risk is endogenous. So, in much financial risk the output of models, and the actions of agents as a result of those outputs, also coordinates changes in the system you are trying to describe.

In this fairly fundamental sense, there are substantial limitations to the extent that you can ever predict financial risk.

Another aspect of interest to us is the very high emphasis that the world meteorological organisation places on user satisfaction, and on the user perspective, in relation to weather forecasts. They place about as much emphasis on accuracy and technical validity as on satisfying the users and addressing questions.

This is again something that we can think about in terms of actuarial models, about to what extent should, or can, the user perspective be integral to model assessments. Should we consider validation as a completely separate exercise to satisfying user requirements, or is it something that should be seen under one broad umbrella in the sense that models are abstract simplifications. Which way you start, which details you focus on, of course depends on the users. There is a quite an intricate mix that links technical verification and the user perspective.

Moving on to aerospace, which is another field where failures are often very public and very catastrophic. It is a high stakes industry from a safety perspective. Again, there is a high emphasis on complex quantitative models. For example, computational fluid dynamics to test and develop design of components and automatic flight control systems.

One thing that is quite a characteristic of this industry is the extent to which failures and near misses are thoroughly and independently investigated. These investigations lead to changes in regulation.

This is again something that we can think about in terms of actuarial models. To what extent are we learning from failures and near misses? Before I continue with this, there is a question of when do we know that we have a failure or near miss? When a plane crashes we know that there was a failure.

But when do you know when the model has failed? Sometimes you know, but often you do not. If you get some model parameters wrong, these are unobservable. If you are writing long tail business, it may be a very long time until you realise that something is not quite right. So, there is a difficulty fundamentally in financial models even defining what the failure is, in many circumstances.

Other concepts of relevance that we picked up from the aerospace field was the idea of operational envelopes. When components are developed, there is quite a lot of work which goes into specifying the conditions under which they can be used. Again, there is a question regarding actuarial models to what extent are we equally careful and rigorous in circumscribing the uses that models can be put to without leading to a potential for adverse outcomes?

Finally, in that field, something that is also going on in aerospace engineering, especially in the development of components, is that there is some parallel independent testing. On the one hand, if you have computational fluid dynamics, so there is a lot of numerical work going on in the computer, but at the same time you have independent experiments in wind tunnels, so you have verification using two very different independent processes: numerical computation and experimental ones. Can we have something like this in our models?

Maybe, in the past, you had the famous back of an envelope calculations that could be used to that effect. But there is a question about whether the still plausible as insurance risk models become more and more complex and try to reflect the dynamics of complex portfolios.

These are some open questions for consideration for actuarial models. I will stop there and handover to Louise.

**Miss L. F. Witts, F.I.A.:** We are looking at other industries, and also what we do in our day jobs. Within the paper we wanted to pull out some of those areas that we more normally work in and how this approach can be applied in those areas.

One of the key things that we were looking at is around expanding some of that discipline into other areas rather than just capital modelling. Are there other uses?

Rob (Black) talked about the difficulties of putting together the inventory. When you have a framework, and you have change controls, somebody who spots that something is wrong with the model then need to fill in a form and send it in. It may be some time until the model governance review takes place and the model cannot be changed until that happens, which could cause an issue. How do you address that while also stopping somebody changing the model because they are using it for one purpose, and the change makes it inapplicable to another purpose?

Model risk has really started to be focused on in life offices, and as the techniques are seen to be helpful, perhaps in the Solvency II area initially, they may be expanded further out.

In banking there is a lot of rigour in the use of models. Some of that comes from the Prudential Regulation Authority. Andrew Bailey gave a talk in November 2015. He outlined the fact that he expects non-executives to have a sufficient understanding of the model to be able to provide a challenge. They are not expected to have detailed technical knowledge but they are expected to be able to challenge and hold people accountable for the use of the model. They need to have support.



With a bank, you have a wide range of complex models in different areas. How do you make sure that the board is able to put in that kind of challenge? If that is what the regulator is saying, we need to put in place those processes.

The majority of banks now have independent model governance teams or model approval teams totally separate from those who are building the model or using the model. The models are set for a purpose and they are allowed to be used for that purpose only. If the purpose is expanded, it needs to go back to the independent team.

The models that we use have to be reviewed annually by model governance to make sure that they are still in line with best practice and thinking at the time. That is a bit more frequent than Rob suggested in the life industry.

One of the things is helping to make sure that the board understand where the model is reliable, where it works well, what moves the dial and what assumptions are reasonable. We are also developing performance indicators for models so that you can report on performance of models to management.

Moving from banking to the pensions industry, I think that there is a lot less thought about model risk in the pensions industry at the moment. In insurance, we have this insurance manager's regime; in banking, we have the senior manager's regime. Those encourage you, if you are responsible for an area under those regimes, to want to know for what models you are responsible and for what purposes are they being used.

There is a possibility that a similar sort of regime may be expanded into all financial services markets in a couple of years' time. There is also talk of the regulatory environment requiring more model risk management.

The pensions industry, for example in longevity, is an area in which model risk is examined and people will run different models.

If you are doing pensions settlements and buying out liabilities, which are once and for all transactions on both sides, the longevity prediction is one of the key parts of the de-risking process for the pension scheme, and one of the key issues for the insurer taking on that risk.

Something which is looked at very closely is which longevity model you are using, which model your competitors are using, and what difference that makes to the results.

So, throughout, we are starting to think a bit more about model risk. It is something that is moving into the pensions industry and we will probably see more of that.

Finally, actuarial standards. As a working party, we fed into the development of the standard. We think that the principles in this paper are in line with what the profession is asking from us, which is fundamentally that: models are fit for purpose; they are well documented; controls and tests are documented; models are kept up-to-date; and communication highlights the key issues around the methodology, the rationale for using it, and the limitations, which are most important.

**The Chairman:** Thank you, Louise. I am looking forward to an interesting discussion. Who would like to open the discussion?

**Mr P. O. J. Kelliher, F.I.A.:** Thank you, Chairman, and thank you to the working party on yet another excellent paper. I have just a few observations on model risk and then one question.

You had the definition of models from the Federal Reserve, but one definition that I prefer is the one in the Oxford English Dictionary. It talks in terms of models being a simplified representation of reality. When you start talking about that it is clear that no model is perfect; there is no unique model; they are all just simplifications.

There is no unique one in 200 year event, either. That is one thing we need to communicate as actuaries. If we take, for instance, our equity stress, there is no unique one in 200 equity stress. I think the Extreme Events Working Party pointed out that using certain data you can have any number between a 20% fall and 90% fall for a 99.5% event. We need to be better at communicating this point that there is no unique one in 200 event.

The other thing is the importance of models in terms of businesses and business decisions. An historical example would be collateralized debt obligation and the models used for pricing of those which were obviously flawed. Another example is models used for house prices for the use of calculating no negative equity guarantees in lifetime mortgages.

In my experience, you can come up with very significant differences in the costs of no negative equity guarantees, depending on the model used. I think that is feeding into the pricing of that feature. Certain kinds of firms could be under-pricing the risk, depending on the model that they are using. On the other hand, certain firms might be over-pricing the risk.

The final observation I would make is when we talk about the fact that there is no unique answer, no one super model, I think this is something we need to get the regulators on board with as well. What we are seeing from Basel is a push back on models because one of the things they are saying is we get lots of different answers in terms of, let us say, credit ratings for the internal ratings based approach. I think the regulator seems to be quite perplexed by the fact that different businesses are assigning completely different probabilities of default, for instance, for similar books.

We need to educate them that really that is quite legitimate. People have different views based on the models. That is producing a different perspective as to what the parameter should be.

The question I would really like to ask is: in terms of model risk management, you mentioned model risk appetite. I am wondering, how would you express model risk appetite?

**Miss Witts:** It is a difficult issue, partly because of the number of models. You need to assess whether each is in line with its key performance indicators (KPIs), and whether it is within its boundaries, so models are given a Red, Amber, Green status as a result of those assessments. This is then qualified by the level of materiality of each model.

You end up with a set of indicators which in some way can express the spectrum of how you feel about the models, which can be an input into an appetite-type indicator.

**Dr Tsanakas:** I have no further comments on the risk appetite. I can come back to some of the earlier comments, though.

I think that this question of being able to formulate many different models on the same data and getting very different answers from them is a key actuarial concern. It is one way of describing uncertainty.

I think that the difficulty of communicating these points to the regulator, and indeed to boards, is quite structural in the sense that we understand models as modellers. To us, we use models to try somehow to find the proper representation of the world. We get puzzled and frustrated by this multiplicity of possible representations.

Boards, on the other hand, do not seem to care so much. With a colleague, I am doing a quantitative project on how models are used in the London market. What seems to be the case is that boards see models as instruments to give them the capital constraint if they are going to embark on a new venture, or something of the like.

In that perspective, a model is not so much a representation of the world but an instrument. This changes the perception of uncertainty completely. If this is the way you think about the model, you know it has limitations but you are not particularly interested in the multiplicity of possible models. You are interested in the multiplicity of different perspectives, and the way the decision is to be taken.

If one turns to the regulator, there is also the situation where a regulator views models instrumentally. For example, in Solvency II models, the models are there to give a capital amount. In that case, from the regulator's perspective, if you have two companies with very similar characteristics, it is puzzling why they should hold different capital levels, which again leads to practices of benchmarking about which people are reputedly very unhappy.

This is a very roundabout comment to your point. I guess the difficulty in communicating and convincing people of the importance of model uncertainty and the multiplicity of possible specifications is quite structural. It is a problem that we will never get rid of as long as models are used. By virtue of using models, you end up with this conflict.

**Miss Witts:** Another point that we discussed within the working party was if there is a point where you should push back on being able to model something? Generally, models are becoming more and more complicated. But sometimes that extra complication is not adding any value.

Your board, or your directors, may have asked you to model something and then totally believe in the answer that comes out of that model. But you may feel there have been such assumptions made along the way that it would have been more sensible at the beginning to have said: "You have to make this decision without any further modelling. We cannot do anything further to help that decision".

**Mr L. van Vuuren:** I work for Tesco Bank. I have just been helping to develop a new model policy. A lot of the debate was on categorising the models, what sort of oversight they would have. I just wanted to elicit some comments or thoughts from you on how you categorise models?

**Dr Black:** The first thing we would say is we need to know all the models that we are using, which for many companies, I am sure, is not information that they have to hand. From my perspective, it has been a journey that has taken a good few months to arrive at a position where we know at every quarter that this is our universe of models in which we are deeply interested.

Then we want to draw out the distinction between those models that are really fundamental to reporting, for example. There are models used across companies that are not for reporting, such as policy quotation systems. These models will all be models in our sense of the word, and we can try to rank them by importance to reporting numbers or to policyholder outcomes, and so on.

We thought that the universe of models in which we have great interest need not be very big. HSBC may have every model they use in a database. I have no idea how big that is. The models that work through to their report and accounts may be a very small subset of the total. It could be that we end up with a dozen – perhaps even fewer – models that are absolutely critical to the business that are going to be high impact category.

You may then have a tier around that which are helping run the end to end process. But you may have comfort in all the controls during production, for example, that those models are used correctly.

Really, it is all driven from understanding what are the models that we use, and, of those, which are the really crucial ones of which we cannot lose sight.

**Mr E. Barnard:** Thank you for both of your papers which have really helped inspire us to make key changes to the software that my firm makes.

It is almost inevitable that there is going to be a knowledge gap between the teams or the individuals who have developed a model and the users and other interested parties in that model. Have the panel any practical advice how to reduce that gap where they feel it is necessary?

**Dr Black:** I think we do insert many unintended barriers across the business to people's understanding of models: how they work; how they are run; and the benefits that they bring. There is therefore an impetus to make things more transparent.

We have all heard the term: "operating in silos". Probably the modeller is as guilty, if not more, as anyone.

Something very interesting to me on the working party is the cycle of independent review and challenge which the banking industry has had to address and, crucially, be seen to have addressed since the financial crisis.

In the banking world, you have teams of reviewers who have not developed the models, or who do not run the models, and they have a remit to go and challenge how models are built and work. That is a very strong place to reach. I do not pretend it is an easy journey to get there. If it could be demonstrated that models had been reviewed on a timely basis, that would really help.

The inventory can be very formulaic and simply list all models and any characteristics of those models. We tend to be quite reactive, and change methodology, perhaps, when the regulator tells us to do so. Crucially, if you have a separate review of models and methodologies, that provides a good

chance to be objective regularly, and say that we are still happy with the way we are doing things. The crucial element is that we have a separate team to make that call.

**Miss Witts:** On communications I have tried to bring out what moves the dial, and also does it move the dial a little bit or a long way? If I had made some slightly different assumptions, or used a slightly different model, would I have come up with a slightly different answer or would I have come up with a fantastically different answer? That is one of the points that is always worth putting across without giving a whole range of sensitivity testing.

As you say, the person who runs the model has a very good understanding of what it does and what it does not do. When I was working, modelling life settlements, the people running that model really had a good understanding of where there were some heroic assumptions. The sales people just wanted to clinch the deal. The task was trying to make sure that you had conveyed to them whether the model was sensitive or not to small changes.

**Mr J. E. Gill, F.F.A.:** Just to follow-up on the last question; in paragraph 4.5.14 you talk about social pressures and what can be done, and the hand off between model creators and the subsequent model users.

There is a very well-documented case study of another NASA mishap. It was not mishap, it was a disaster. It was the second space shuttle disaster, the one that burnt up on re-entry and nine lives were lost as a result of the heat shield not operating appropriately. Why did that happen? It was because more of the heat shield tiles fell off nine days before when the space shuttle took off than had ever happened before.

NASA had a team of people whose entire job was to model the number of tiles that had fallen off. They did not know because they did not have a mechanism to count and they did not have a sensor that told them. They had a team that just took hundreds of photographs when the shuttle took off, and from that they created a model that said how many of the tiles had fallen off. They presented their conclusions within an hour of take-off which was that the result was out of model range. That was a phrase which went all the way up the organisation in NASA because no one in NASA at the higher levels understood “out of model range” meant that the heat shield will not work.

A tragic example, unfortunately, of where the modellers did their job. They came up with a model that gave the answer but they failed to communicate the answer to those who could do anything with it because they used modelling language rather than: “And here is the impact”.

I liked the phrase in the paper in 4.5.14 when you talked about fostering a culture to encourage substantiated dissent. How do you do that? How do you make sure that those difficult messages are passed up in language that the users can understand?

From the many discussions to which Rob (Black) referred, I hope you can give us some light on how to make that happen.

**Dr Tsanakas:** In my organisation, as an academic, it is very easy to dissent. I am not an expert on organisational behaviour, a subject which I expect most of you have studied in the university of life.

One thing that can be said about this topic is the distinction between roles can be quite helpful in the sense that, for example, different stakeholders in the model can take very different perspectives, so as long as you have people who can legitimately, let us say, draw red lines, and they have an acknowledged role in that process, then you can do something.

But the question is very broad. How do you include a substantiated dissent? It is not only to do with model risk; it is to do with managing any organisation. The question really is too big for this working party.

**Dr Black:** From my background, I recognise that model development is a collaborative activity, which is a good thing. It probably reflects working for quite a large organisation. No doubt it would be different in a smaller organisation. There is no one person who dreams up a model development, implements it, runs it, signs it off. It goes through lots of hands. If you have a mix of personalities, you will get criticism and comments and approval or dissent. That is a natural process.

Equally, a “lessons learnt” session is an extremely valuable activity. We talked about some very high-profile model failures, often with a human cost. We get a lot of benefit through reviewing how models have performed, how processes have performed, to try to make things better and quicker next time, not to explicitly acknowledge that something went wrong, merely just to say, in the Solvency II world, say, we are doing this every quarter, how can it be better next time?

So, lessons learnt we found to be extremely constructive. People all come with an open agenda. There is no finger-pointing. It is really an honest appraisal: things went all right this time; it would be nice if they went even better next time. To have that in the quarterly cycle of developments is very helpful.

**The Chairman:** Do you want to say something about third party software? I am just wondering the extent to which we have the issue that if many people are using the same companies' ESGs, how are these decided upon? Is it some committee of seven people in a room that is deciding half the economic scenarios of the financial industry in Britain, or something like that? I am sure that there is a risk there but is that something that you have thought about?

**Dr Black:** This question came up in a working party discussion. Many people seem to have implicitly assumed that third-party software is correct, infallible, and will work. But it cannot be any different to the software developed in-house or models developed in-house. That being the case, we need to give it as much attention.

We probably all accept that third party software plays a crucial role in a lot of organisations but I do not think that it receives anything like as much scrutiny. Often that is because when the software is brought in, it is enough of a challenge just to understand it at the level at which we want to use it. To take an extra leap and really become expert about what is under the bonnet is asking a lot.

We certainly thought that there was merit in trying to up our game, if you like, in terms of understanding how third-party software works.

**Miss Witts:** Often in the process of purchasing external software you might go through that process with a lot of rigour. But when you have been using the software for a number of years, perhaps that rigour has been lost with the people who knew about it initially. For example, the external software

may have been changed along the way and new improvements may have been introduced. Whether those improvements impact the purpose for which you are using the software is another matter.

One of the possible solutions is to try to embed the discipline of looking again at the model on an annual basis, to see what has changed in the way you use it, and what has changed in the software.

Such an approach would keep the rigour that was initially undertaken alive throughout the use of the model rather than just at that one moment in time.

**The Chairman:** It is not just ESGs. I am assuming that we are all sitting here assuming that the Continuous Mortality Investigation (CMI) have their sums right and that their model of improvements is the best available. But do we know that we are using it properly? Or do we just put blind faith in what we are using from the CMI and assume that it is correct?

Is there any more we should know about communicating model risks?

**Dr Black:** It would be good to hear stories around communication more externally. That would be a challenge for the working party.

**Mr D. G. Robinson, F.F.A.:** I am thinking about communication of models to boards because the regulator is encouraging boards to have diversity in all senses: cognitive diversity; and diversity of experience. You have actuaries on boards but you also have people with other skill sets, and other backgrounds. Yet the regulator also requires these people to verify, to confirm and to take responsibility for the model that has been used in the business.

I am wondering what sort of experience the panel have in communicating models to boards, but also whether anyone in the audience has experience of what regulated companies are doing to help non-actuaries, non-experts, to understand models.

So those are two questions: experience around communicating to boards, and what boards might be doing to help their non-executive directors to understand the models being presented to them and for which they have responsibility.

**The Chairman:** Have you done any training for boards about this is what your model does?

**Dr Black:** Personally, no. The Solvency II program, prior to launch, extensively engaged the board with our latest models, not just executive directors but non-executives, too. I do not know whether it is on a renewal type basis, whether it happens every year, or whatever frequency.

I have seen no evidence that this has been done.

**Miss Witts:** As a bank, we do not have a lot of actuaries on the board. A lot of people are from a banking background so they may be more familiar with credit risk models or banking models or various mortgage models. One of the things we have to do is train people on longevity and that is always an interesting discussion. People like the statistics around how long they are going to live.

We introduced model risk KPIs into our risk appetite. And we have an independent model risk approval team of the type Rob (Black) mentioned. These are more common in the banks now. They

went through a process of training the board on the type of models at which they were looking; the number of models at which they were looking; the type of things that they were putting in the KPIs; and how they arrived at the process of suggesting a risk appetite that the board should monitor.

Part of the reason that we are asking questions around this area is because we do not know whether there are any ideal solutions. It is challenging different people; and, as you say, we have lots of different cultures.

You will be training your non-Executive directors using their pack, and some will ask different questions from others. If they have been trustees of pension schemes in the past, they might ask one set of questions. If they have come from a purely banking background, they might ask a different set of questions.

Trying to communicate the uncertainty, and going through some of the features and the types of assumptions, is all that we are trying to do at the moment. But we would love to learn from other people who have had good experiences in this area.

**The Chairman:** There are a few non-executive directors here. I wonder whether any of them have any experiences that they would like to share. Do you feel that you have been trained well by your advisers?

**Mrs E. E. A. Jeffrey, F.F.A.:** You are talking about model governance teams and model governance committees. I am thinking about the reliance on external providers. What happens when critical external models break in extreme circumstances? Is there anything you have seen where the committees and the team have asked for independent verification, or how otherwise to do it in emergencies?

**The Chairman:** What do you do when you break it? Have you thought about that?

**Miss Witts:** It has been a challenge for some of our software for sure. “Work around” or “fudge” is often the answer. You are back to the back of the envelope that we started from.

That is where I guess my comment earlier about complexity comes in. You have to draw a line on complexity if you have broken the model.

**Dr Black:** I do not have experience of third-party software blowing up in that sense. We do comment in the paper about the need to subject third-party software to user acceptance testing.

One option would be to include it in your end to end process with a typical extreme stress or scenario, for example. I have not seen examples of external software breaking during critical reporting periods. There is a big risk, especially bringing in a new piece of software. We need to accommodate it in the same way that we do with all change.

**Mr K. Luscombe, F.F.A.:** Picking up David (Robinson)’s question about boards and diversity, I think that it can be difficult for those on boards who, perhaps, come from a less technical background to understand the intricacies of financial models and how the world might fall out in future years.



But, I think that it can be more helpful the other way round. If you take a specific scenario and say the way that the world could go in this scenario, how would business look? Then you can have a more rounded discussion including those from a technical background and those from a wider one. You can then say: “will that make sense” and you can say: “you could do this or that”, particularly looking around potential management actions or mitigations that could be done to help the future. It could help the other way. But it helps if you translate what is a technical model output into what does this look like in the real world.

I was interested in the earlier example you had of the weather systems where you are saying we cannot influence the weather for tomorrow no matter how well or otherwise we are able to project it or forecast it. Of course, that may be true of tomorrow’s weather. We do not have much influence over that. What will the weather be on 29 June next year? It could be anything.

In the very long term all of successive forecasts could eventually come into future climate and climate change. There is no absence of models about how that might develop in the future, given some of the actions that are being taken. Of course, there are equally diverse views on how climate change will go, using that same information.

My question, bringing it back to financial matters, which might be more relevant to this audience, is how much of a connection do you think there is between the model writers and those using the models?

Is there an understanding of what is in one box and how it is being used? There are a lot of assumptions built into models, such as in this scenario we will take this action and do these things. Those are then built into the model to refine and perhaps temper or change the outcome.

How much in terms of the scheme of overall risk do you think that set of assumptions or otherwise is having in terms of model risk outcomes? Is it quite well modelled or do you think that the variation in assumptions is as big and significant a risk as any other?

**Dr Black:** I think from the perspective of a practitioner, there is a great risk, potentially, that models are not used correctly; that your starting assumption set is wrong in some way and in a way that the model developer did not intend.

Equally, from experience, that is where a lot of effort is put in to ensuring that models are set up in the correct way, time after time. There is an increasingly automated approach, so we do not have people keying in numbers to Excel workbooks to start the process. That part is becoming automated as far as it can be, and we are increasingly seeing the benefits of automation.

As we move to worlds where we have to model hundreds of scenarios, ideally we want to model them more and more. There seems a pressure to do more modelling. I do not think that is going to go away. Automation of set ups is where we are heading, which significantly reduces the risk.

**The Chairman:** It remains for me to express my own thanks to all of you and the thanks of all of you to the authors. I declare the meeting closed and ask you to show your appreciation to our authors and contributors.