DISCUSSION HELD AT THE FACULTY OF ACTUARIES

The Vice-President (Mr J. F. Hylands, F.F.A.): Our authors this evening are most welcome, Stephen Richards and Dr Iain Currie. Stephen Richards is a graduate of Heriot-Watt. He graduated in actuarial mathematics and statistics. He has worked since 1990 in life assurance and nowadays runs his own consultancy and software business specialising in the analysis of mortality and other demographics.

Dr Iain Currie is Reader in Statistics in the School of Mathematical and Computer Sciences at Heriot-Watt University. Iain's main research interest is in multi-dimensional smoothing methods with particular application to the modelling and forecasting of mortality rates. He is responsible for developing GLAM, which is a high-speed/low storage method of smoothing data that are arranged in multi-dimensional arrays. I am sure he would be delighted to tell people more about that afterwards.

I am delighted that the authors are here this evening to present the paper. The paper is the fruit of a research award from the Actuarial Profession where we are endeavouring to focus research activities on particular areas of interest or importance to the profession.

Mr S. J. Richards, F.F.A. (introducing the paper)**:** I am very glad to see so many of you in the audience tonight to talk about longevity risk, not least because we know that there are some other pressing macro-economic factors impinging on your time.

For example, when we first calculated the illustrative annuity factors in our paper, we used a discount rate of 5% per annum. This was the Bank of England base rate in early October 2008. Just four months later, the base rate has fallen to 1%, an historic low since the founding of the Bank of England. I know that there are other things on your minds besides longevity risk, but I am very glad that you could come along tonight.

Just like interest rates, mortality rates have also reached an historic low, not just in the UK but in most other developed nations. However, unlike interest rates, mortality rates still seem to have plenty of scope to fall further.

Figure 4 in the paper shows a clear downward trend for nearly half a century at various ages. The obvious assumption is that mortality rates will continue to fall. But how far? And how fast?

One view of mortality rates is so-called longevity risk, namely the risk that pensioners or annuitants will live longer than assumed in pricing or reserving calculations. A key question is what direction the trends in Figure 4 will take. This is not just a question of what the likeliest trend is, but of what future trends are possible and the likelihood of each. It is for this reason that the authors prefer statistical models for projections, rather than deterministic scenarios. A particular scenario might be interesting, for example if it involves extrapolating trends in causes of death, but a scenario on its own is not particularly informative without a statement of its likelihood.

Both life-office annuities and defined-benefit pensions must be reserved for prudently. There is no set definition of prudence for a mortality projection for pensions or annuities. A minimal approach — or one approach I might advocate — is that prudence could be defined such that mortality rates should be more than 50% likely to be higher than those assumed in the pricing or reserving basis. In other words, a prudent reserving basis would use rates below the central projection from a statistical model. The reserving basis can itself be a deterministic scenario, possibly even one derived from a cause-of-death-based extrapolation.

However, such scenarios would have to be benchmarked against the results of a statistical projection to prove they were prudent. There are two main problems with this approach, however. The first (which I think is the core of our paper) is model risk, namely the fact that you cannot know which model will be appropriate for projection. This problem exists for all forms of forecasting, but was succinctly expressed by Cairns (2004) when writing about interest-rate models: "probability statements derived from the use of a single model and parameter set should be treated with caution".

https://doi.org/10.1017/S1357321700005687 Published online by Cambridge University Press

This is illustrated in Figure 5, where two different parameterisations of the Lee–Carter model lead to two very different projections. This does not necessarily mean that either, or both, of the models are wrong. Rather, it shows that even slight changes to model structure can radically change the direction of projections and the apparent uncertainty over the projections. Model risk can be demonstrated, but not completely quantified — a "known unknown" to quote Donald Rumsfeld. For this reason the prudence of a reserving basis needs to be benchmarked against more than one statistical projection.

Alternatively, an actuary might include an explicit margin for model risk. But this is perhaps a bit more art than science.

The second problem with statistical projections is somewhat less philosophical, namely which dataset to use. Ideally, an actuary would have a forty-year dataset for exactly the portfolio being reserved for. In practice, this is almost unattainable: few portfolios have historic mortality this far back.

One solution is to use a separate dataset which does have the necessary history. However, this introduces basis risk, namely the fact that the data used for projections is not the same as the population for which the actuary is reserving.

The two most obvious candidate datasets in the UK are population data from the ONS and assured-lives data from the CMI. Neither is especially satisfactory. The ONS data is not ideal as private pensioners and annuitants have markedly lighter mortality than the wider population. The ONS data also have the problem that figures are available only up to age 89 whereas we want to reserve far beyond that age.

The alternative is the CMI data, but this is not ideal either. There is little data for female lives, for example, and Figure 7 suggests that the CMI assured-lives dataset has undergone radical shifts in volume and possibly changes in composition as a result. Neither dataset is ideal, but as actuaries we have to accept these limitations and work with what we have. However, as with model risk, an actuary should include an explicit margin for basis risk for pensioner and annuitant mortality projections.

A further problem with basis risk is that lives are not equal in financial terms. Figure D1 shows the degree of concentration of liabilities in a small proportion of lives in any given portfolio: the top decile of pensioners typically has round about half of all the pensions in payment. This goes for a large annuity portfolio (the top decile had 54% of all the pensions) and for a collection of defined-benefit pension schemes (46% of the total pension was being paid to 10% of the members). The possibility of this top decile experiencing different mortality trends from the rest is also a form of basis risk. You may parameterise a model using a particular dataset, but if 10% of lives drive more than half of the liabilities, you have a risk that what happens to that 10% is not the same as what happens to the other 90% driving the projections.

However, this particular aspect of basis risk can also be categorised as part of concentration risk, since there are other consequences of this lack of equality in terms of financial impact. For example, different social groups have different absolute levels of mortality, which can have a much bigger impact on the reserving basis than mortality projections.

Concentration risk is interesting because it also interacts with stochastic or binomial risk, namely the uncertainty over when someone dies even if you knew exactly what the underlying mortality rates were. For example, Tables 9-11 show the volatility in run-off by lives, but also that heterogeneity in pension size increases that volatility. For very small pensions schemes (of under 50 lives, say), the binomial risk or stochastic risk will often dominate all other sources of risk in the scheme, including even investment risk.

All of these risks — model risk, basis risk, concentration risk and binomial risk — can be quickly and easily explored with modern computers. Nowadays there is no need to pick model points to somehow represent a particular portfolio as entire portfolios can be simulated in run-off on a life-by-life basis in a fairly modest time frame. This requires de-duplication to be done, particularly for annuity portfolios. Figure 1 shows that the tendency to hold multiple policies is strongly correlated with the underlying wealth of the individual. Failure to de-duplicate would give a falsely comforting picture by both understating concentration risk and over-stating the

https://doi.org/10.1017/S1357321700005687 Published online by Cambridge University Press

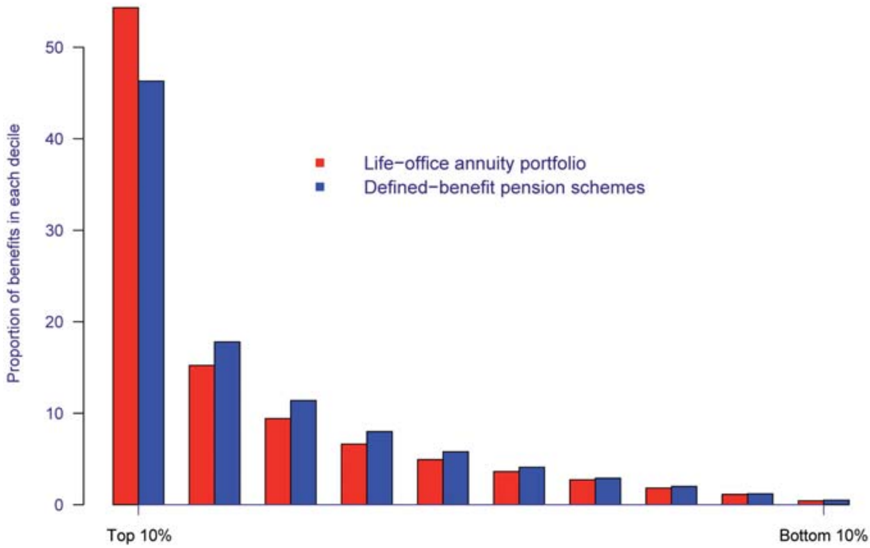

Figure D1. Liability concentration

benefit from the law of large numbers. By parallelising calculations across a standard eight-core system, even a portfolio as large as 200,000 lives can be simulated in run-off 10,000 times in an hour.

The resulting data can be presented graphically in a single chart, such as Figure 10. The horizontal axis shows the reserve for the present value of the benefits for this 200,000-life portfolio. The vertical axis shows the probability of a given reserve being adequate in run-off.

For a given projection model, the 10,000 run-off costs from our simulation can be arranged in ascending order to form an ogive, i.e. a cumulative distribution function of the probability that the given level of reserve will be adequate in run-off. This can be done for as many projection models as you like on the same graph, but Figure 10 compares just two different projection models for future improvements.

An actuary can use any basis he or she likes to calculate a reserve and place it on the x axis. He or she can then read off the probability that the reserve will be adequate with respect to longevity risk in run-off. Note that we are actually ignoring other risks like investment risk and various expenses, but these could be added as well, which would change the shape of the ogive.

As an example, Figure 10 shows vertical lines for a couple of reserving bases in common use. For example, the line for the long-cohort reserve in Figure 10 suggests it is moderately prudent under the DDE model with around 60% probability of adequacy. However, under the CR model (which is one of the core models in the paper, where we have decided to smooth the time trend) we can see the very same reserve would appear to be only adequate with a 10% probability in run-off. The same reserve; two subtle differences in the Lee–Carter model; yet two very different pictures as to whether or not that reserve is going to be adequate.

Conversely, a reserve which was minimally prudent under the CR model, which only just exceeded the 50% level, would have a probability of adequacy of 95% under the DDE model. This might be regarded by some as too prudent, especially taking into account other margins in the reserving basis.

To sum up, the prudence of mortality projections in both pension and annuity reserving can only really be demonstrated with reference to the results of a statistical projection model. The measures of uncertainty are the key here, and these cannot be obtained from a deterministic scenario alone. However, both model risk and basis risk are very substantial, and actuaries need to consider multiple alternative models rather than just one single projection model.

Professor A. J. G. Cairns, F.F.A. (opening the discussion)**:** I should like to thank the authors. It is a very interesting and relatively short paper, by Faculty sessional meetings standards. It is short in number of pages but contains plenty of good content. It is very thought-provoking and very focused.

One of the key points brought out in the paper is the focus on model risk. I was talking to both of the authors in the last week or two about this. The idea was very much to focus on one particular model, the Lee–Carter model, which is well-known to many people, but then to look at one or two sub-models even within that particular, relatively simple class of models.

The point of the paper, then, is to try to focus on some simple ideas concerning model risk without getting too cluttered up with some of the non-Lee–Carter models that are available in the published literature (see for example, Cairns *et al.*, 2009).

In this particular case the sub-models might be described as taking some different assumptions about smoothing of the various underlying age effects and the period effects in particular with this model. The general conclusions that come out are that there are significant differences between the sub-models. That is what has been talked about and it is illustrated amply in several of the figures that are in the paper as well as in the tables.

What I also take away from the paper is a point also highlighted in tonight's introduction by Mr Richards. Namely, that the margins for model risk are at least as large as the margins for things like parameter risk (in particular, the central trend), at least in the context of the particular numerical examples that came out in that paper. In general, there are some very nice illustrations there from where the different margins might emerge.

In general terms, the balance between the different risk margins does depend on the specific question that is being asked, but it is very often the case that the margins for model risk, however you want to quantify that, are at least as large as margins for process risk and parameter risk.

However, with parameter risk and process risk, you can, to some extent, build those into a very specific and objective way of calculating the margins. On the other hand, when you have two or three models it is rather more difficult actually to attach the probability to each of those models, making the assessment of a specific risk margin rather difficult.

Despite that, as Mr Richards recommended in his introduction, you really do need to look at these different models rather than just bury your head in the sand and say, "This is too difficult. Let's not think about it.''

In the sub-modelling, there is what I like to think of as being a philosophical issue. I describe it as that because it is something that is rather difficult to prove, I suspect, or argue one way or the other. That is how much smoothing do you do in each of the dimensions? You have age effects, cohort effects and period effects. Each one of those three factors in a model could be smoothed or not. A particular focus in this paper is how much smoothing you put into the period effect. You can certainly ask the same questions of the age and period effects.

If I were to ask all of you whether age effects should be smoothed, probably quite a lot of you would say that age effects ought to be smoothed. On the other hand, with period effects I suspect that there is more debate over whether period effects are random (that is, you have significant systematic deviations from one year to the next) or alternatively perhaps you would have smooth underlying signals plus some systematic noise, a substantial shock which affects all ages in the same way. These are the issues that I know Dr Currie has been working on in separate work over the past couple of years.

Model risk has been illustrated well in this paper. A bit of advertising perhaps for some other work that I have been doing, Cairns *et al.* (2008), where I have been looking at a range of different models, of which the Lee–Carter model is one out of a number.

Figure D2 is just another illustration of how model risk can be assessed through doing projections under different models.

The fan charts are similar to the ones you see in tonight's paper and can be interpreted in the same way as you would a Bank of England fan chart. There are six models and these are examples without any parameter risk. The point here is to focus on illustrating the differences between different models.

So you can see that there are relatively modest but significant differences between the different models in terms of the central trend, but you can also see that there are quite big differences in terms of the spread as you go further and further out into the future. All of these things are ones that need to be thought about, not ignored. They need to be thought about and used in a way which informs your calculations of best estimate reserves, and then also the margins that you add on to them.

Figure D3 looks at cohort life expectancy. It is one model but it is focusing on parameter risk. Again, what you take away from this is how the time horizon affects the relationship between process risk on the one hand and parameter risk on the other. Maybe over the first ten years the process risk, the underlying randomness in the underlying process, is the dominant effect, but the further out you go, you can see that parameter risk starts to take over as the dominant factor that contributes to the width of the fan. If you were to add model risk on top of

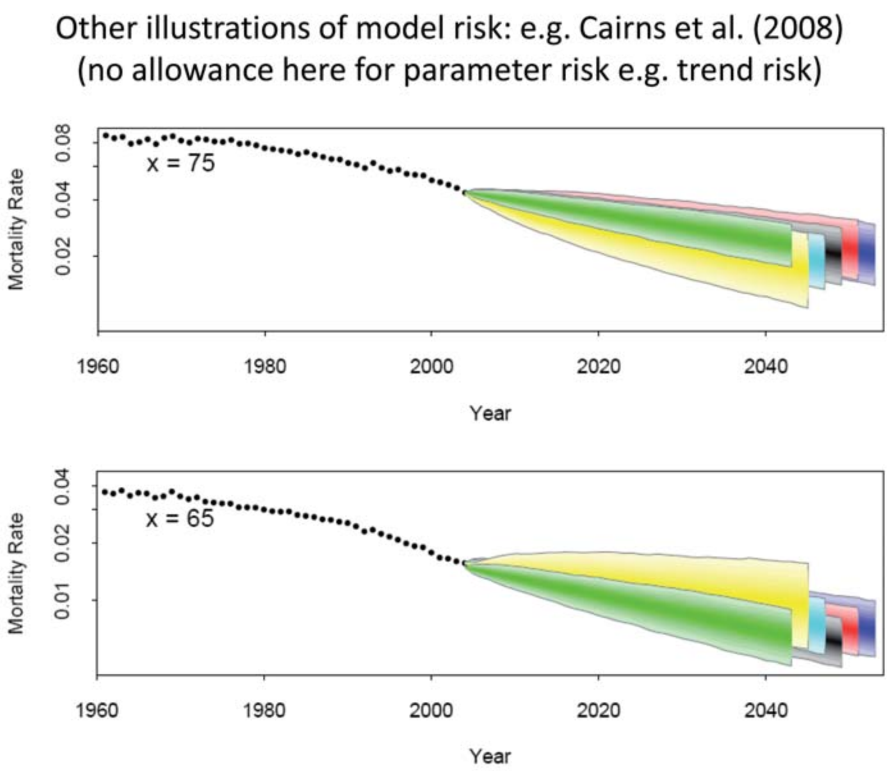

Figure D2. Model risk illustrations

https://doi.org/10.1017/S1357321700005687 Published online by Cambridge University Press

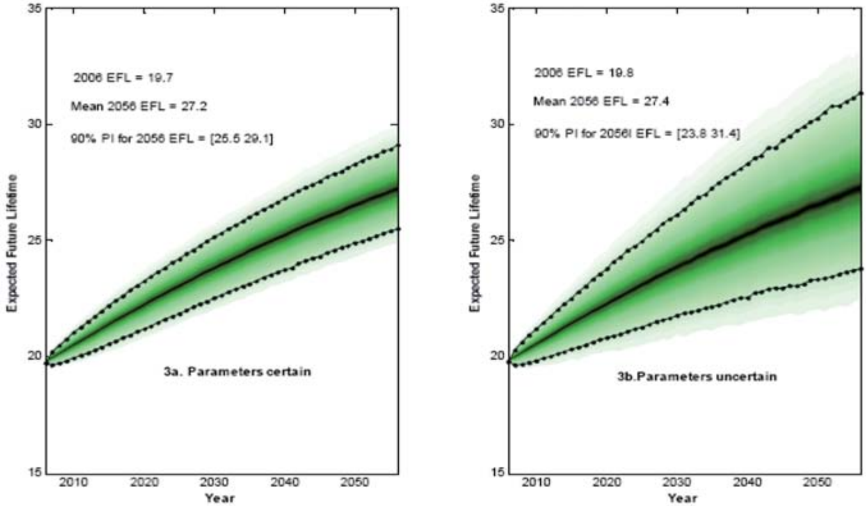

Figure D3. Longevity fan charts for 65 year old English & Welsh males:2006-2056

that of course it would be wider still. So reserving is one of the other issues, a very important issue in this paper.

There is a focus on examples which are looking at more mature portfolios. If you are also thinking about younger active pension plan members, say 30 or 40 years of age, then the longevity risk does indeed start to manifest itself rather more than perhaps it does in the calculations that are presented in tonight's paper: the risk that has an impact on the cost of an annuity at the future age of 65. This annuity price depends on longevity risk as well as interest rate risk or real interest rate risk.

The balance that you get between those different risks very much depends on the different time horizons that you might be looking at. It also depends on the hedging strategy that you are adopting. If you are, say, following an equity-linked investment strategy, then the interest rate risk might be relatively significant. But even then, interest rates are mean-reverting in the long run, while, in contrast, longevity risk is not thought to be. So, the longer the time horizon, the balance tips away from interest-rate risk towards longevity risk. Even if you are not doing any hedging of interest rate risk, longevity risk begins to dominate in the longer run.

Mr Richards also touched on other populations in his introduction, since they are in the paper. The idea here of course is, even if you have the right model, when you are looking at historical data you have limited amounts of data and a limited number of years. You may also have small populations, so you have a lot of parameter risk associated with that. If you just fit a single model to the individual populations, then it is very likely that you will end up with divergent trends.

What you have to think about now is whether these divergent trends are reasonable or not. Most people would agree, I think, that you may get some deviations and you may have, say, one population which typically has mortality rates that are 50% of standard England and Wales mortality. In the long run you would expect both of them to move up and down in parallel, at least in terms of looking at the ratio. So even though both sets of mortality rates are going down, you would expect the ratio to remain relatively stable. These are ideas of biological

https://doi.org/10.1017/S1357321700005687 Published online by Cambridge University Press

reasonableness that you need to bring in to think about modelling two populations, so coupling populations for some consistency.

My view is that you need at least to attempt to model this coupling properly using some form of statistical model, rather than adopting some sort of ad hoc deterministic projection. Again, that is one of the key stories that is coming out in the conclusions in tonight's paper. So there is this basis risk. It is also about the trend risk, I would argue, in particular.

Summarising my comments, yes, deterministic scenarios are okay, but they should only be there as part of a bigger package of risk analysis using good statistical or stochastic models, and indeed again, as has been stressed in the introduction tonight, it needs to be based on not just one particular model and indeed one sub-model within the Lee–Carter, for example, you need to be looking at a number of different models just to get a good feel for what the range of results could be in the future without necessarily trying formally to attach probabilities to each of those models and particular reserving levels.

I certainly think that ignoring stochastic models just on the basis that there are too many and that it is too confusing is a bad idea: you will end up going backwards and using some deterministic scenarios. To me such a response is simply not acceptable: it represents poor risk management and therefore leads to significant operational risk. I thoroughly agree with the authors' conclusion that you need to you use more than one model.

References

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D. & Khalaf-Allah, M. (2008). *Mortality density forecasts: an analysis of six stochastic mortality models.* Working Paper.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A. & Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, **13**(1), 1-35.

Dowd, K., Blake, D. & Cairns, A.J.G. (2010). Facing up to uncertain life expectancy: the longevity fan charts. *Demography*, **47**, 67-78.

Mr A. G. Sharp, F.F.A. (in a written contribution which was read to the meeting)**:** I would like to thank the authors for another valuable contribution to actuarial understanding of mortality, and the implications of the uncertainties about mortality in future years for life insurers and pension schemes. I am sorry that I am not able to attend in person.

I wish to comment on Figure 7, on page 14 of the paper, which shows a rapid reduction in exposure data in the assured lives dataset between 1985, 1995 and 2005. The first comment is to clarify the dataset that is illustrated. This is described in the paper as "assured lives" — and I think this shorthand has also been used by the CMI itself at times — but it is more fully described as non-linked single-life permanent (Whole Life & Endowment) assurances issued after the provision of satisfactory evidence of health. This is a subset of the overall mortality data collected by the CMI from life offices, and is the subset that was used for much of the CMI's work on the P-spline and Lee–Carter methods that is documented in CMI Working Papers 20 and 25. Sample projections using these methods have been included in the CMI Library of Mortality Projections. In undertaking this and earlier work, the CMI did not look only at this dataset — in part because of the declining data volumes and the potential discontinuity as this dataset switches from predominantly endowment policies (at ages up to 65) to predominantly whole life policies (at older ages). However, it is only by looking at a dataset with worthwhile data below age 65 that one is likely to be able to detect cohort features at significantly younger ages. Other potentially suitable CMI datasets have not exhibited similar decreases in volumes, but they have not been collected for long enough to contribute much to projection work. Some other subsets of CMI data — such as those covering annuitants or life office pensioners — only contain significant volumes of data above age 65 and so are of limited use in this regard. This issue was noted in the CMI working papers at the time.

https://doi.org/10.1017/S1357321700005687 Published online by Cambridge University Press

The second aspect that I would like to comment on is the reduction in data volumes for the "assured lives" dataset that is starkly illustrated by Figure 7, which may surprise many readers of the paper. Some of the illustrated reduction in data volumes is due to a significant reduction in sales of endowment assurances since the mid- to late 1990s. Data volumes for some other product lines, such as term assurances, have not exhibited similar decreases over that period.

Overall, though, it is true that volumes of data submitted to the CMI by life offices have fallen in recent years. There are a number of reasons why this may have happened. One is industry consolidation; as life insurers have merged and when data is migrated from one administration system to another, consideration of submitting data to the CMI is probably at the bottom of the agenda in some cases, if it actually features on the agenda at all! So what are we doing about data collection? Some years ago we launched an extensive consultation process regarding the modernisation of the data collected by the CMI — the "Per Policy" initiative. This initiative — and the much higher profile now afforded to mortality — has persuaded many "new" data contributors to come forward. Unfortunately many of our existing data contributors have not yet been able to respond fully to this new initiative — and we are engaging further with these insurers to try to ensure that we reverse the decline in data volumes. I would also draw your attention to our recent announcement on mortality projections—

> "The CMI Executive Committee has established a Working Party to examine the latest data on mortality improvements and to develop a generic model, able to produce a range of different projections. This follows concerns that the Interim Cohort Projections, or variants of them, are still in near-universal use, despite these projections being based on experience data only up to 1999. The intention is to develop a relatively simple, spreadsheet-based model capable of widespread application. Good progress is being made, and the Working Party hopes to issue a paper for consultation, together with an illustrative version of the model, by March 2009."

Reference

CMI (Mortality Projection Working Party) (2007). Stochastic projection methodologies: further progress and P-Spline model features, example results and implications. Working Paper No. 20.

The Vice-President: I might just mention at this point a slight aside, but relevant to the matter under discussion, that the profession, in conjunction with the CMI, is about to embark on a consultation exercise with a wide range of providers and users of CMI data to assess how the profession can fulfil the most useful possible role in studies of longevity and morbidity. That has not formally been announced by the profession yet, but it will be coming along very soon. I am sure that some people in this room, in one capacity or another, will be receiving letters inviting them to contribute to the consultation, and if you do not receive such a communication but you would like to contribute, then I am sure your contribution will be most welcome.

Dr I. D. Currie (panel member)**:** I should like to pick up on a couple of points that Professor Cairns made. He pointed out that, given one of our two illustrative models, we were able to suggest margins for prudent reserving. We did this for two models. What you will look for in vain in our paper is a recommendation on which model to choose. That was not an oversight on our part. We had no way of discriminating between the two models. In the larger sense, if you had, say, a dozen candidate models, you could go through this exercise of deriving a prudent margin for each of these. But then you are left with the problem of deciding where to hang your hat.

From the mathematical point of view it is simply saying that we have no way of putting a probability distribution over these different models. They exist on their own. They are not related. All it does is give you a greater impression, more evidence, of the fact that there is a lot of risk connected with forecasting mortality.

Coming onto the second point that I want to mention, Professor Cairns happens to have an office along the corridor from me and I can absolutely assure you that I did not plant the following question. I do not think that I have discussed his comments about divergent trends. In an earlier grant that I had from the profession, I published a paper with a couple of colleagues in 2004 where I took the CMI data, pensioner data, on lives and amounts. I showed how it was possible to produce a consistent forecast fitting two surfaces so that they produced parallel forecasts. This is exactly as Professor Cairns suggested. I think that is very much the way forward so far as dealing with very small datasets where you cannot actually make a forecast with the limited data but could try to relate the forecast to some other larger, more secure dataset.

Dr L. M. Pryor, F.I.A.: I should like to thank the authors very much for an extremely timely and useful paper. To me its interest is that much of what they say is generally applicable to modelling in general rather than being specifically about mortality.

In ¶3.2 they describe four problems of forecasting which, I believe, apply to any sorts of models: model risk, parameter uncertainty, parameter stability and stochastic variability. In ¶¶8.2 and 8.3 they go on to point out that the risk which dominates depends on the size of the population that you are modelling. If you have a small population, it will be stochastic variability that is most important. That is especially true, of course, if you have a small pension scheme with a few lives that represent most of the liability. It only takes one long-lived ex-managing director for all your modelling to be absolutely pointless. But of course, for larger populations, model and parameter risk come into their own.

However, there are two other important problems in forecasting — the data and the judgement of the model builder. Everyone knows the data affects forecasting. The reliability of the forecast depends on both the size and the quality of the dataset. There are some passing references in the paper to the importance of removing duplicates, but there are many other problems that you want to be aware of and to try to do something about, if possible. Another important issue is of course how closely the dataset resembles the population that you are trying to model. We have also heard a bit about that.

Judgment is I think equally important. Judgment and, as I will come to, opinion. It permeates the whole of model fitting. What general form of model are you going to use? For mortality are you going to you use a causal model or an extrapolatory model? If the latter, are you going to you use Lee–Carter or something else? If Lee–Carter, which variant?

The fitting procedure, too, depends a lot on judgment. How much smoothing, where to place your knots, and all sorts of other matters. How much back-testing are you going to do? How much of your data are you going to use to fit the model, and how much are you going to use to do back-testing? All this is judgment or, as I think that Professor Cairns especially pointed out, opinion. Much of the time there just really is not any factual basis to discriminate between the possibilities. What you are doing is essentially using your opinion.

I think it is very important to be clear when you are making these sorts of decisions whether you are basing them on fact or opinion. And it is especially important when you are not responsible for the final decisions that are being made. Nowadays it is usually the boards of insurance companies, the trustees of pension funds, who decide on the assumptions to be used. The work you are doing is providing information on which they base their decisions, and it is very important that they know the status of the information they are getting. Are they making decisions based on fact or are they making decisions based on your opinion? I think that you have to make that very clear to them.

In ¶9.3 the authors point out that the effect of mortality models is financially significant for annuity business. That is a point that is well taken, especially in the context of how you can tell which is the right model.

Also, you have to remember that mortality is not the only significant factor in pricing annuities. There are going to be other models you are using for interest rates, and so on, that are equally, if not more, significant. I would be very interested to know what people think about the relative significance of the assumptions that go into these sorts of decisions.

The final sentence of ¶9.3 is particularly important for all modelling: "... the actuary's role should be a little less about calculating the value of the liability, and much more about demonstrating the depths of what we do not — and cannot — know." I think that is very important.

Finally, I just want to pick up on a comment in ¶7.3, which Mr Richards also alluded to in his opening remarks. The authors say: "... a claim of prudence has to be substantiated, and it is hard to do this convincingly without reference to a statistical projection model." I completely agree that a claim of prudence has to be substantiated. That is especially the case when it is not you who is making the final decision. You are providing information on which others base their decisions. If you are recommending an assumption as being prudent it is very important that the people who decide whether to accept your recommendation know why you think it is prudent and how prudent you think it is.

The authors say that it is hard to do this convincingly without reference to a statistical projection model. I take their point to a certain extent. I am afraid I am not giving any answers here. I am just raising more problems. Yes, statistics help, but given that we have seen that the actual model you choose may have a large effect, if you are saying something is prudent because it is 95% or 99% percentile, or whatever, that is only with respect to the model you have chosen. As Professor Cairns pointed out, there are a lot of different models you could choose, and you have to look at the whole range of models. An assumption that is very prudent if you are using just one model may in fact turn out to be positively reckless in the context of another model. You need to look at the whole picture and not just get tied up with a statistical analysis based on a single model.

This paper has given me a lot of food for thought and I hope it will give a lot of other actuaries food for thought, too. Thank you very much to the authors.

Dr Currie: Thank you very much for the very supportive comments. As a mathematician, of course, I am not immune to holding opinions. What I think is unusual about a mathematician is that his or her opinions are absolutely laid out clearly. The models that we have put out are there and you can shoot them down. It is very clear where we are coming from. If you do not like the model, you can produce the evidence which says that this model is not doing the job.

Of course, we do not just produce the models, we do try to see whether they describe the data, by doing back-testing, and these models work reasonably well, certainly well enough to do the kind of modelling work on mortality that we have presented tonight.

Mr O. J. Lockwood (a visitor)**:** I am carrying out a research project at Heriot-Watt on another subject related to stochastic mortality models. The authors' stated aim is to demonstrate that the choice of model has a material impact on the trend and uncertainty of future mortality projections and therefore on the range of likely values for longevity-related liabilities. The authors also discuss the relative merits in terms of applicability to a portfolio of annuitants or pensioners of fitting the models they consider to ONS data and to CMI assured lives data. They conclude that both datasets have advantages and disadvantages in terms of applicability, and that the choice between them may have a material impact on the modelled future trend and uncertainty.

I believe this paper will successfully achieve the aim of highlighting to its readers the importance of the choice of model and of the dataset it is fitted to, and my comments are not intended to undermine this importance in any way. I do, however, have some comments regarding the detailed calibration of the models considered in the paper.

There is no mention in the paper of the form of time series model fitted to the kappa parameters in the LC and DDE models. I should have liked to see this stated explicitly without the need to consult the references. My comments assume the model is a standard ARIMA-type model with no consideration having been given to the appropriateness, or otherwise, of assuming the residuals are IID normally distributed.

https://doi.org/10.1017/S1357321700005687 Published online by Cambridge University Press

In the CR model, the future projection of the kappa parameters follows a linear trend, with the slope being determined by the last two P-spline coefficients in the region of the data. The resulting projections for mortality at age 65 are shown in Figure 5 for ONS data and in Figure 6 for CMI data. We see that the CR model forecasts much higher mortality improvements than the LC and DDE models for ONS data, but all three models forecast similar rates of improvement for CMI data. This feature is commented on by the authors, but I feel it warrants further explanation.

I note that the main criticism the CMI levelled at the Lee–Carter model in Working Paper 25 was the lack of an explicit allowance for cohort effects. These effects, however, tend to appear implicitly in the kappa parameters in the form of an unusually large fall when lives experiencing a favourable cohort effect reach the most common ages of death. Longevity in the 21st Century (*BAJ* **10**, 685) identified a favourable cohort effect centred on the mid-1920s for CMI data and on the early 1930s for ONS data. This tends to suggest that by the time the data ends, in 2005-2006, the cohort effect is still causing the kappa parameters to decrease rapidly over time for ONS data, consistently with what can be seen in the bottom right graph in Figure 3, but no longer has much impact on the kappa parameters for CMI data.

My suggestion is, firstly, that this is what is causing the dashed line to slope more steeply than the grey solid line in Figure 5, whereas both lines have a similar slope in Figure 6. Secondly, I suggest that if this explanation is correct, then it makes the dashed line in Figure 5 inappropriate as a mortality projection because it is extrapolating a short-term rapid fall in the kappa parameters that cannot be expected to continue indefinitely. I am more positive about the grey solid line in Figure 5 and both lines in Figure 6. Nevertheless, without having analysed the residuals of the time series models used to construct the grey solid lines, I would question the validity of the IID normal assumption given that cohort effects have distorted the kappa parameters.

The question then arises as to how the work in the paper could be extended to incorporate an explicit cohort effect. The authors mention the work of Renshaw & Haberman (2006). Unfortunately, my understanding is that the attempts that have been made to fit the Renshaw & Haberman model to different datasets have revealed a lack of robustness. For example, adding a further year's data sometimes causes the parameter estimates to jump to completely different values.

More robust is the Age-Period-Cohort model, under which the logarithm of the force of mortality is an age term plus a period term plus a cohort term. This model does not, however, allow for the possibility that the sensitivity of mortality to period and to cohort effects might vary with age. It would, in my view, be of interest to consider a compromise between the Renshaw & Haberman model and the Age-Period-Cohort model under which the coefficients of the period and cohort terms are allowed to vary with age, as in the Renshaw & Haberman model, but with smoothness imposed on this age dependency by P-splines, in the same way as in the DDE model.

To end I would like to mention a point about the ICA stress scenarios in Section 7. This section has the potential to cause confusion regarding the significance of the 99.5%. In the ICA, rather than calculating, say, 1000 present values of the annuity payments and taking the fifth worst, 1000 possibilities are considered as to what the best estimate of the liabilities will be in a year's time, and the fifth worst of these is taken. I am not giving a view as to whether this defines an appropriate amount of capital to hold in respect of annuity liabilities; I am simply observing that this is how the regulations are currently framed.

Mr Richards: To reply to Mr Lockwood, I know that the 99.5% stress scenarios in the paper are not according to the definition of ICA regulations; ICA-style 99.5% stresses are one-year stresses, whereas these are in run-off. Personally, I find the run-off stresses more interesting and more appropriate for annuity business, but it is a point of clarification that, yes, these are not ICA stresses, which must be calculated on a one-year horizon.

I would like to pick up on the point about separating age from period effect from trend effect.

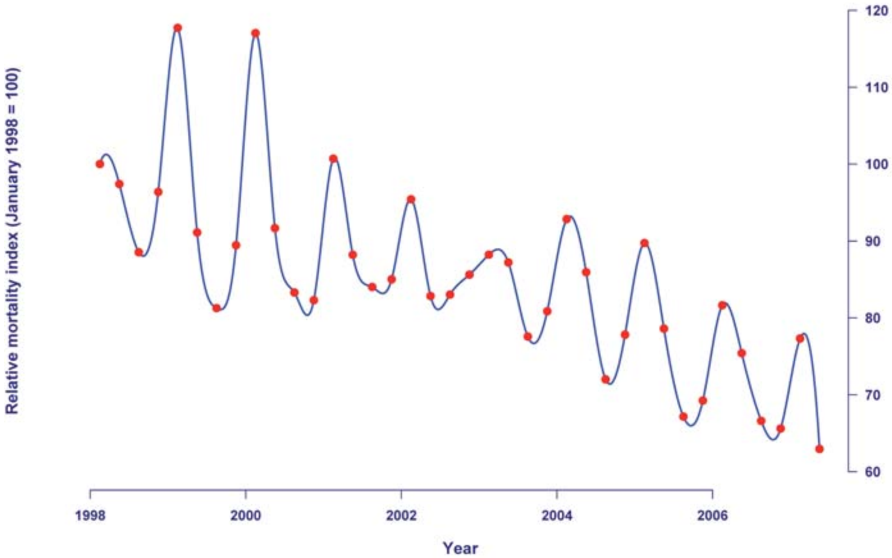

Figure D4. Seasonal Cox Model for a large life office annuity portfolio

Figure D4 is a Cox Model fitted for age, gender and obviously period. "Period'' here means season, aligned to the ONS's definition of season, so winter is November, December and January, with the other seasons following in three-month groups.

What we can see are two clear effects. One is the strong and obvious downward trend from left to right. These reductions in mortality are partly driven by medical improvements and partly driven by cohort improvements. We can also see the impact of period effects within this trend. We can see a clear periodicity within the calendar year.

All of the peaks, are winter months. There are also period effects between the years. We can see the winters of 1999 and 2000 are periods of much more extreme fluctuation because of the high mortality of those two winters. We can also see one of the points made by Professor Cairns about the independence assumption. There is a strong degree of negative auto correlation. The winters of high mortality tend to be followed by summers of light mortality.

Figure D4 is for life office annuitants. Figure D5 is for the defined-benefit pensioner portfolio which is referred to in the paper. It is also referred to in Richards (2008b). We see exactly the same things. We see strong periodicity. We have the same unusual pattern in 2003. We also have the same strong downward trend from left to right.

Often people think there is something qualitatively different about life-office annuitants and defined-benefit pensioners. These two graphs from two very large portfolios — there are about 250,000 lives at least in each — show that there is more that unites them than divides them. They are subject to exactly the same patterns, and almost exactly the same trends. The average aggregate annual mortality improvement in both charts is about 3.5% to 3.6% per annum.

Mr A. M. Eastwood, F.F.A.: I have a series of questions for the authors. Some of these points have been touched on by other speakers this evening. It always seems to me with any kind of extrapolation methodology that there seems to be a lot hinging on the time period covered by the dataset, and that by moving on a year or two you can completely change the trends. That

https://doi.org/10.1017/S1357321700005687 Published online by Cambridge University Press

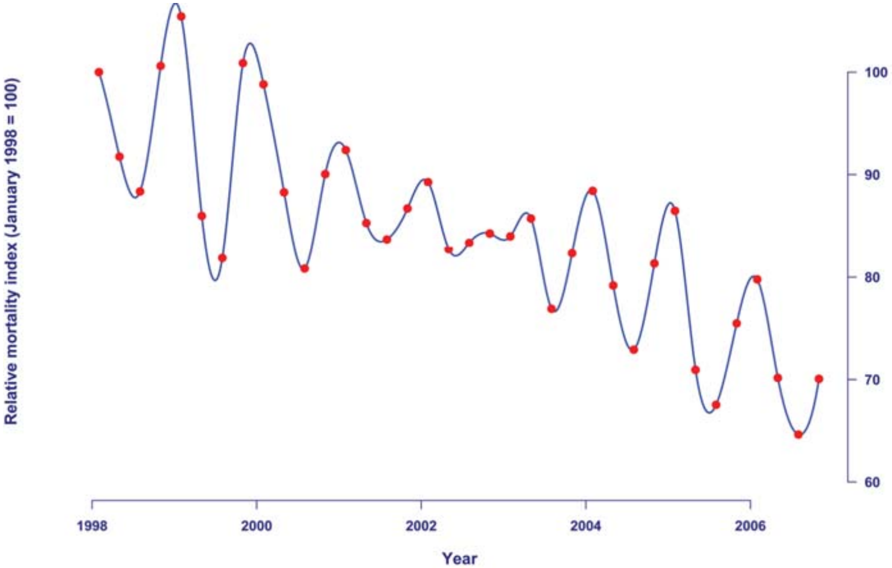

Figure D5

certainly seems to be the case with some of the P-spline work that I know Dr Currie was very heavily involved in developing. I am interested in any comment that the authors might have on the criticality of that kind of assumption.

My second question relates to the CMI library of model projections. Do the authors have a view on the extent to which that provides a sufficiently complete set of models or whether the profession should be looking to see that dramatically expanded so that we can get a better impression of model uncertainty for those of us that do not have time to delve into the myriad of statistical methods involved?

A final question, or a plea, for different annuity longevity or trend projections, somehow to relate the mathematical formulae in the model back to real life parameters. If I try to explain the choice of projections to a board, I would like to be able to explain the choice along the lines, "I have a choice of a number of models here and it is important you understand the choice we have to make. In essence here is what is going on. This one is ignoring x; this one is assuming that we can draw a straight line through that; but all the rest of the difference in models could be due to statistical fluctuation.''

I struggled a little bit in this regard. I could see that one of the two models that was covered in detail in the paper was smoothing to a far greater extent than the other. But I think just a brief summary with any of these kinds of models in words as to what the key differences are I would find tremendously helpful in trying to explain things.

Dr Currie: Thank you very much. You asked three questions. I will try to take them in turn. You mentioned that some of the P-spline work could be rather unpredictable if you added one year's data; completely different things seemed to happen. I think one of the problems with this was people were not using the model in the way that I had envisaged.

A big difference between the penalty method and the time series method is that the time series is fitted to the whole of the historic data, whereas the penalty method largely takes its projection from the much more recent past. It certainly fits the model to all the data. But the

https://doi.org/10.1017/S1357321700005687 Published online by Cambridge University Press

forecast largely follows what has happened in the recent past. This is where opinion and judgment come in, I suppose. I would not advocate using knots that were very close together because that automatically introduces volatility into the system.

I do know that the CMI software allowed you to specify knot spacings that were really much less than I had ever envisaged. I think this had something to do with some of the more unfortunate effects that were observed.

To come to your second question, we really are asking for a complete listing of models. I think you realised when you asked the question that that is not possible. We looked at two classes of models, the P-spline model and Lee–Carter models. Professor Cairns has taken a more encyclopaedic approach to these matters. He has a paper where he lists a very large number of these models. You can build others.

The Renshaw & Haberman models have been mentioned. I very much agree with the point that if they are going to be used for forecasting, then some amount of smoothing will introduce stability in the models and there is much more chance of them being successful if smoothing was added.

Finally, you asked how you might explain some of these rather abstruse mathematical models to the board. I thought we had really been quite helpful there. As well as giving the formula, what we did do in the paper was give a lot of pictures. I have a little mantra about mathematics which says: "If you cannot draw it, you do not understand it." When you are talking to your board, I would try to explain the models in terms of diagrams and explain what the parameters are through the pictures of the parameters and how the various forecasts work. I hope your board members would appreciate a little bit of graphics.

Mr C. R. Barnard, F.I.A. (in a written contribution which was read to the meeting)**:** The Lee–Carter method for projecting mortality rates, first proposed in the early 1990s, consists of a stochastic loglinear model belonging to the family of Generalised Linear Models. As such, it allows for extensive statistical inference, and can be used to make stochastic mortality projections by modelling its time-varying parameter as a random walk. The model is fitted to historical data.

The Lee–Carter method is easy to understand and control, and is widely used by government agencies, mortality offices and supranational bodies to project mortality rates. For these reasons it has become a safe option to choose for mortality projections. However, as with all uncertain, long-term projections, we must be careful not to put too much confidence in the results, or to present results with too much confidence.

For example, when using the projected mortality rates to construct tables for annuity pricing, the following risks need to be considered.

Model risk or mis-specification: The annuitant mortality might not have the structure and development as described in the model. For example the mortality development might change to a non-loglinear development, or it might display heavy cohort effects.

Similarly, the population used to generate the table might not coincide with the future group of annuitants. For example, different selection effects might prevail, or the average group of annuitants, for which the table was determined, may not represent the target group of a particular insurance entity.

Parameter or estimation risk: Even if a representative model is chosen, the parameters are only deduced from the available data and might thus be inaccurate.

The above risks are types of systematic risk.

Risk of random fluctuations: Tables only predict mean values over a large population of annuitants. As the typical annuity portfolio of many insurance companies is quite small, random deviations from the mean values could be considerable. This is a type of unsystematic risk.

The Lee–Carter method gives useful base mortality forecasts, assuming that present mortality trends continue in the future. I have found that carrying out the mortality projection by sex, geographic region, and even cause of death is even more informative, where enough credible data exists to do this. One of my company's Eastern European subsidiaries had 16m exposed to risk

https://doi.org/10.1017/S1357321700005687 Published online by Cambridge University Press

data over 6 years (it used to be the State insurer). We used this, with additional national data, to disaggregate our Lee–Carter projection as above, with very interesting results. The base projection is not used unadjusted for pricing voluntary annuities, as the 16m dataset relates mainly to life insurance lives. However, it would be more useful for pricing annuities under the country's mandatory second pillar pension scheme.

In summary, we must be careful in putting too much confidence in model forecasts of mortality rates without considering the risk factors mentioned above, the data used, and the purpose of the mortality projection. We should also educate non-experts in the problems and limitations of using such models for long-term mortality forecasting.

Dr Currie: I fully support what Mr Barnard said. He has a very large dataset and that allows him to do this disaggregation which will clearly give forecasts that are relevant for the various sub-groups involved. I very much support these comments.

Mr A. C. Martin, F.F.A.: I am speaking as an actuary and also as an independent pension scheme trustee. I can endorse the huge significance of concentration risk as outlined in ¶8.3.

In my practice and in my experience, I would suggest even schemes up to 1000 members will have huge concentration risk. Typically there are directors, executives, owners or the long-serving members who will have the vast majority of the liability. I really cannot underestimate how significant that is. Indeed, on risk registers, which trustees should have for internal control purposes, I would suggest that any longevity risk should actually include the concentration risk which is often of at least equal significance. It is a huge element of trustee knowledge and understanding.

The following conclusion is that buyout should be considered as a funding target or the default investment option. If there was any consideration by scheme sponsors that this might seem quite expensive, I would merely suggest that they look at the concentration risk and the possible impact of that. If that was not enough, I suggest they capitalise the cost of administering the scheme over the run-off period with some realistic compliance and administration costs — with an appropriate escalator for actuarial fees!

Finally, may I pass on a perception. I think that perceptions can be as important as fact. That is again relevant to the concentration risk. I recall when the Pensions Regulators put their head above the parapet in February last year with a consultation paper on mortality, there was a huge fuss at the mere suggestion that something more prudent than long cohort was necessary. Several of my fellow trustees have actually looked at this in detail and interestingly came out with the same conclusion. That consideration involved looking at a graph with the short, medium and cohort effects running down to 0. The end result was a reflection, "I see, medical advances are going to stop then, then and then.'' So how relevant and how prudent is that? I thought it was quite an important perception and one I think we should bear in mind when we are doing projections in the future.

Mr D. C. E. Wilson, F.I.A.: I was very much struck with parallels between this paper and the last sessional meeting paper that we had on stress tests for investment risk. The reason being the main conclusion that I take from the paper is simply the amount of judgment or prior views on the form of future mortality improvements that is required in order to come up with a suitable stress test that covers a one-in-200 year event, for example.

In the same way as for investment, there seems to be no way that we can say with any confidence what an appropriate capital allowance should be. If that is the case, at the risk of repeating what I said last month, I feel that the regulation under Solvency II, which has a distinction between a standard model with the QIS 4 type stress test which is discussed in this paper, and an internal model which requires somebody's own judgment, has a severe danger that the use of judgment is prevented through the use of a standard stress instead.

Given the impossibility of coming up with a consensus on judgment, then having regulators state a particular stress, like the 25% stress discussed on the QIS 4, is not by itself a problem.

Then it becomes simply a standard and we should stop calling it a one-in-200 year event because we do not know that is what it is.

But it would appear from the work that the authors have done that the form of the QIS 4 stress, which is applied simply as a fixed stress to a mortality rate at all different ages, is inappropriate, and using different stresses at different ages is more likely to capture the level of risk in fact in annuity business.

Mr Richards: I would agree with Mr Wilson that the QIS 4 structure with a flat 25% reduction is perhaps over-simplistic, and indeed penalises portfolios of a high average age. The benefit of it, of course, is that it is simple. I would agree that that percentage should ideally be tiered by age of the portfolio.

Mr S. D. Baxter, F.I.A.: I should like to echo a couple of the comments that have been mentioned earlier tonight. I profoundly agree with earlier speakers that it is necessary to look at a variety of different models in order to understand model risk. But I also agree with Dr Currie that it is almost impossible to put any kind of probability distribution over different models.

It is also very important to choose the right dataset to fit the model to — by 'right' I mean the most relevant dataset for the population to which the model will be applied. This mitigates the basis risk we have been hearing so much about tonight. As an example, my colleagues and I have done a lot of analysis of large occupational pension scheme datasets, and we have noticed very material differences in terms of mortality rates and life expectancies between different individuals based on characteristics such as socio-economic background, affluence, lifestyle and retirement patterns. Different groups of individuals with like characteristics also demonstrate very different age and time patterns to improvements in mortality rates. These observations should influence which projections we consider to be appropriate for different annuitant portfolios and pension schemes.

This leads me on to my second point which is to do with the dangers of pure extrapolation without considering our projections in a wider context. Like many actuaries, I am from a mathematical background and appreciate the beauty of a mathematical model; but I am also acutely aware of how models are perceived by the recipients of our advice. For certain audiences mathematical models can work well. However, in my experience of advising pension schemes, mathematical models do not tend to work well when speaking to a lay trustee. Invariably, the question they return to is: "Putting it simply, are you telling me we will find a cure for cancer?"

I doubt any of us would feel comfortable answering whether we will find a cure for cancer. Nor, are any of the models we have seen tonight as specific as predicting what will happen to mortality rates by cause. But I do think that we need to be extremely cognisant of the perception of the profession. To my mind this means being able to convert the results of mathematical models into examples of real-world scenarios that the results would be consistent with.

Mr S. J. Makin, F.F.A. (closing the discussion)**:** The paper certainly adds to our profession's already-rich contribution in the area of mortality forecasting. The authors are to be congratulated for that.

The authors point out that their paper presents a methodology rather than an answer. I would agree. The paper raises many questions; but far from being a criticism of the paper, I think this is to be welcomed.

Ms Pryor asked earlier for people's views on the relative significance of different assumptions. I will not give my views but suffice it to say that for the first year end in quite some time the focus is not on annuitant mortality assumptions, which for me is a very refreshing change. It is important to remember that mortality is often just one assumption amongst many others and that, depending upon the context, there may well be other assumptions of equal uncertainty and even greater importance.

Picking up on a comment made by Mr Lockwood, many firms calculate their one-year VaR ICA for longevity using a run-off approach but with a lower probability of certainty. The reasons relate to practical convenience.

https://doi.org/10.1017/S1357321700005687 Published online by Cambridge University Press

In terms of the paper itself, I have a few thoughts and observations which I would like to share. I did think that Section 2 — essentially 4 pages on data de-duplication — was a little excessive. With my tongue firmly in cheek, it is just as well that de-duplication was not applied to the References section of the paper, else Mr Richards' long and impressive list of papers would have appeared somewhat shorter than is the case, albeit he would have been the author of one much larger paper, provided that the de-duplication routines worked of course!

Sometimes de-duplication is unnecessary. It really depends what is being done. For example, de-duplication is not particularly needed if the exercise is to place a value on a portfolio of liabilities. It is, however, clearly necessary for the sorts of portfolio simulations discussed in Section 8 of the paper, and, for that matter, in fitting models of individual mortality to detailed portfolio data.

Section 3 is where the paper really began for me. In ¶3.2, as Ms Pryor has mentioned, the authors point out the problems with forecasting methods, being the well-known ones of model risk, parameter risk and stochastic variation. I have to say that I particularly liked the authors' approach to splitting parameter risk into two sub-components: parameter uncertainty and parameter stability. These are often not presented separately, and I found this presentation to be both useful and insightful.

At the end of ¶3.2 the authors note that only if model risk and parameter risk are "successfully negotiated" will the model's confidence intervals give a true reflection of uncertainty. In ¶6.7, however, the authors suggest that "it is not possible to know if the model is (or will be) correct". An extension of these points might well be that the authors consider mortality forecasting to be impossible, and I look forward to their comments on this later. Whatever their views, though, I am drawn to the authors' later sense that mortality forecasting requires both caution and humility.

In ¶3.4, the authors discussed an extension of the Lee–Carter model, proposed by Delwarde *et al*. (2007), which smoothes the age responses in the traditional Lee–Carter model. This model is appealing, particularly for application to smaller datasets, but this does create something of a dilemma in my mind: where we would wish the model to be of most use — on small datasets — is exactly where the issue of parameter uncertainty rears its ugly head posing all sorts of questions. That said, at least one can use the model's own estimates of parameter uncertainty to derive confidence intervals around the resulting projections.

Both the original Lee–Carter model and the Delwarde extension of it achieve their forecasts by time series projection of the period responses. Notwithstanding that the authors were clear that they had used the same approach to time-series forecasting as the original Lee & Carter paper, I, like Mr Kirkwood, would have liked to have seen some discussion and some detail on this approach in tonight's paper.

In ¶3.5, the authors present their model. In addition to looking to smooth the age responses, the authors look to smooth the period responses also, and the method of choice, as discussed, is penalised spline regression. This then gives the authors a means of forecasting, the rationale being to use the regulating properties of the penalty function to smooth the B-spline coefficients and, as already discussed this evening, to linearly extrapolate from the final two coefficients. The underlying approach to forecasting here is very similar indeed to that which underlies the CMI's P-Spline model, which is very familiar to us.

At this point, though, I would pause to note that Figure 3 raises questions about the validity of this approach to forecasting. The period response parameter has not been linear, and so the assumption of future linearity driven by the final two parameters does seem to be something of a strong one.

That said, I would point out that this approach to forecasting does pay better regard to the broken trend in the traditional Lee–Carter period responses. This is a phenomenon which is observed in the study of many countries' data; around the mid-1970s there is a step change in the slope of the period parameters. When I read tonight's paper, I was reminded of a paper by Chan *et* al. (2008), which was presented to the Society of Actuaries in America. That paper considered broken-trend stationary models which dealt with this feature by having an explicit break point to permit changes in both the slope and intercept of the time series.

https://doi.org/10.1017/S1357321700005687 Published online by Cambridge University Press

By virtue of its parametric nature, the authors' model is somewhat less suited to stochastic modelling than is the case with the Lee–Carter model or the Delwarde variant. That said, the quantification of confidence intervals is somewhat easier.

On the subject of confidence intervals, the authors note in ¶5.5 that, when applied to the ONS dataset, their model has much wider confidence intervals than the Delwarde variant. I was rather left wondering why that was the case, and in particular how much of this was an artefact of the time-series forecasting method — an ARIMA(0, 1, 0) process is known to produce rather narrow and in fact ultimately parallel funnels of doubt on the log scale. We see later, though, that this is not a universal relationship, and in particular applying the authors' model to smaller datasets leads to the by-now-familiar (albeit paradoxical) result of tighter confidence intervals. Intuitively this result is still unappealing, albeit that we understand the reasons for it.

In Section 6 the authors present a financial quantification of the effects of applying different models, and this is a useful way of getting some kind of handle on model risk. I did find this Section of the paper to be particularly interesting, although I felt that the comparisons to the 25% longevity shock, as envisaged by QIS4, were capable a little mis-interpretation. The QIS4 requirement of a 25% longevity shock is intended to cover all aspects of longevity risk — current mortality risk and future trend risk — yet the authors have made comparison only to trend risk.

That said, and echoing some comments from Mr Wilson, the QIS4 requirement itself is arguably ill-conceived. When I tell you that it was set relatively arbitrarily in the QIS3 study, not the current QIS4 study but the previous QIS3 study, with reference to not too much more than the available anecdotal evidence, why would you believe it be correct? And even if it were about right overall, why would it be correct for each entity carrying annuity liabilities?

The authors point out that the QIS4 requirement has different effects for different portfolios. More fundamentally, though, the risks associated with different portfolios vary considerably, and specific consideration of each portfolio's own circumstances must be preferable to the one-size-fits-all herding that this aspect of the proposed Solvency II framework may well create. Internal models would seem to be vital for firms with what might reasonably be regarded as a material exposure to longevity risk.

Now, a thought for all those life offices writing standard non-profit annuity business. Based on your own models, just how much of the 25% would you allocate to the risk of mis-estimating today's mortality? And, again based on your models, just how much would you allocate to the risk of mis-estimating the trend? Taking all of that into account, how much do you have left when you stop to reflect on the risks posed by the anti-selective effects of the enhanced annuity market? The enhanced market selects many of the unhealthy lives, leaving behind what might be considered to be the super-healthy ones. What models do you have for this effect? How up-to-date are those models? And, most importantly, does your Board understand the models? Is it happy with the capital being held against those risks, if indeed any capital is being held against those risks?

So, where next for the model presented in this paper? It would seem to me that the model needs something of a road-test. I would have liked to have seen some analyses on backtesting. Like Mr Eastwood, I would have liked to have seen some analyses on the sensitivity to the final year of the data, or for that matter to knot spacing. I would also have liked to have seen the forecast (median) mortality curves in future years. The application of judgement as to whether these are sensible is an important part of the model validation process, and to finish, I also would like to say a few words about judgement.

The authors point out in ¶7.3 that "prudence is ... a matter of judgement" and that "a claim of prudence has to be substantiated". I agree with these points entirely. The authors then go on to say that "[substantiating prudence] ... is hard to do ... convincingly without reference to a statistical projection model".

Well, this is where it all becomes rather circular, essentially because one must first have some belief in one's statistical model before validating prudence relative to the model. An illusion of prudence can easily be created using a statistical model, and we do not need to look very far these days for examples of the consequences of blindly relying on models: we will all have heard, for example, of the assertions made by banks having experienced 25-sigma events several days in a row.

To put that in context, under a certain modelling assumption of course, a single 25-sigma event is about as likely as winning the National Lottery 20 weeks in a row having bought just one ticket each time. The Bank of England had a simpler explanation: the models were wrong! (Haldane, 2009).

Just how much belief did the banks have in their models, and what real-world scenarios did they test them against?

And to the models here tonight, what do the projections mean in biological and medical terms? This approach may be helpful to Mr Eastwood and his plea for help to explain the models to his Board.

By my rough and ready reckoning, the median projection of the Lee–Carter and Delwarde models are broadly consistent with eliminating all circulatory mortality over the next 10 years, albeit seeing no other or further improvements. Circulatory mortality is the biggest killer of annuitants in the UK, so this is no trivial assumption. The median projection of the authors' model is consistent with the additional eradication of the two largest cancers — lung and colorectal — over that same 10-year period. My question is: is that believable? I make no comments in either direction, but the use of such language and currency gives a means of discussion in the wider fields. This can be no bad thing for model validation.

References

Chan, W.S., Li, S-H. & Cheung, S-H. (2008). Testing deterministic versus stochastic trends in the Lee–Carter mortality indexes and its implications for projecting mortality improvements at advanced ages. http://www.soa.org/library/monographs/retirement-systems/living-to-100-and-beyond/2008/january/mono-li08-6a-chan-abstract.pdf

CEIOPS (2007). QIS3 Calibration of the underwriting risk, market risk and MCR (CEIOPS-FS-14/07).
http://www.ceiops.eu/media/files/consultations/QIS/QIS3/QIS3CalibrationPapers.pdf

Haldane, A.G. (2009). Why banks failed the stress test.
http://www.bankofengland.co.uk/publications/speeches/2009/speech374.pdf

Mr Richards (responding)**:** I just have one point for Mr Makin. The extensive material on data and data preparation was originally just one page, but one of the referees requested that we back up our claim that people of higher socio-economic status tended to have a larger number of policies. The referee did not doubt it, but wanted to see evidence that was the case. Apologies if we provided too much evidence.

Dr Currie (responding)**:** First of all, I should like to thank all the contributors very much for what has been largely a supportive set of comments. I should also like to thank Mr Makin, in particular, for his remarks.

I did scribble this down quickly as he said it, but he did make the rather startling suggestion whether we really think that mortality forecasting was in fact possible at all. I guess my answer to that is of course it is all too easy to do it where you just produce the model and the forecast follows. What of course is much harder is to produce a forecast that is actually going to turn out to be right. As we stand here in the present day, all forecasts are good. But who knows 20 years down the line?

That is why I think it is so important that we look at a wide range of models. It is only that way you can begin to get a prudent view of things by looking at a large range of models rather than saying, "I really believe in this model. This is where I get my prudence from." I do not think Stephen and I are suggesting that for one moment.

Mr Makin was very critical in Figure 3 of the linear forecasting that took place on the period effect. This is using the smooth method which extrapolated the final two coefficients in the B-spline fit.

My answer to that is this. I think that the tone of your remark was that the linear forecast did not sit with what appears to be a curve in the data. I think you were suggesting that we should have had more dramatic improvements. Paradoxically, if you look at Figure 5, in fact it is the P-spline method that gives the greater improvements. I have not prepared a graph for tonight's presentation, but it is possible to fit some sort of quadratic function and the improvements that you get there would give half the people in this room a heart attack!

Mr Makin was also a little bit worried about the fact that the confidence intervals for the smoothing method were considerably wider than the confidence intervals given by the time series method as presented in Figure 6 on page 13. I do not worry about that. I just see this as an aspect of the model. If you look at Figure 5, the two models give different forecasts. In Figure 6, using a different dataset, using the CMI dataset, the two forecasts are largely consistent, but the confidence intervals this time are different. So I think it is just an aspect of using a different model.

Finally, Mr Makin made a point about wondering whether the improvements that we were suggesting could actually be realised, given that they were equivalent to wiping out all premature deaths from circulatory diseases in 10 years, or possibly even adding to that deaths from lung cancer and colorectal cancer.

Well, it is a startling claim, but I would draw your attention to a statistic that Stephen gave earlier this evening to say that we have been seeing recently, over the last ten years, a 3.3% (I think that is the figure) annual rate of improvement in mortality. At the present day mortality is improving very, very rapidly. So I would just bat it back at you and say: why should this improvement suddenly come to a juddering halt?

Thank you all again for your remarks. Given the discussion, it is quite clear that forecasting mortality is a daunting business indeed, almost as bad as forecasting the change in GDP, which has been commented on tonight already, of course. Pity poor Mervyn King at the Bank of England, who, just a few months ago, was not forecasting negative growth; even the most pessimistic forecasts in his fan charts did not stray into negative territory. And Mr King's forecasts look only three years or so ahead. Pity the actuary who has to forecast at least 50 years ahead.

"What is the most difficult thing in forecasting?" Harold Macmillan would have had something to say about this. It would be something along the lines of, "Events, dear boy, events." However, this is all very gloomy.

Can we throw some light on the matter? In ¶3.2 we identify four problems — as a number of people have commented already — that the forecaster must face: (a) model risk; (b) parameter uncertainty; (c) parameter stability; and (d) stochastic variation. One of the great advantages of a statistical model is that we can do something about (b) and (d), parameter uncertainty and stochastic variation.

Figure D6 uses the CMI dataset on male-assured lives, together with two-dimensional P-splines, to forecast the mortality of 70 year-old males. Years 1950 to 1975 are used as the training set to forecast the mortality from 1976 to 2005. Certainly, our central forecast did not anticipate the accelerated fall that occurred from around 1975, but the dramatic decline was certainly well within the funnel of plausible futures.

One further problem we identified in our paper is that of basis risk. There has been quite a lot of mention of that this evening. Ideally we should base our forecast on the dataset with the same mortality as those lives we wish to insure. Such datasets are hard to find, sadly; we certainly had to compromise in the work for tonight. The situation is as follows: we have a large dataset which can be used for forecasting and a small dataset of interest which is not suitable for forecasting. Can we somehow use the forecast from the "wrong" dataset to help to forecast with the "correct" dataset? I call, rather whimsically, I think, such a model a piggy-back model. Stephen and I hope to report on this work at some future date, though I will not attempt to forecast when that might be!

Our paper tonight is largely concerned with the first of the forecaster's problems: model risk and, in particular, the financial consequences of model mis-specification. Again, just for simplicity, Stephen and I have stayed within a very small family of models, the Lee–Carter

https://doi.org/10.1017/S1357321700005687 Published online by Cambridge University Press

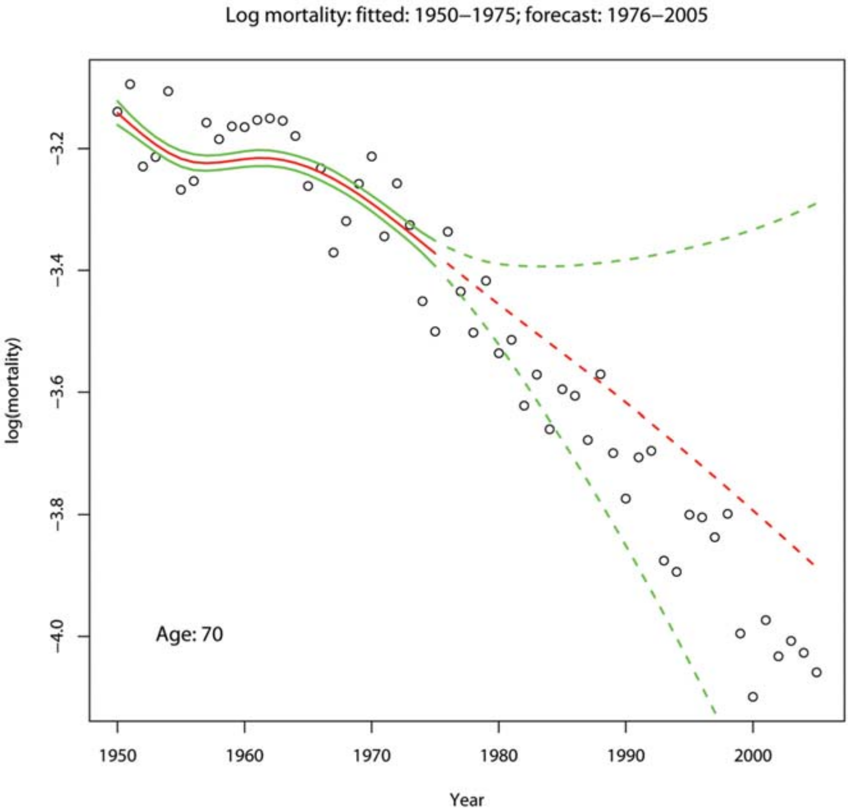

Figure D6

family. We find that even within this small family, forecasts vary substantially. More importantly, the financial consequences of these differences matter for both annuity and pension providers, and equally for their policyholders.

So, to sum up. The golfers among you will know that the further you hit the ball, the more any deviation from the straight and narrow will be magnified; so in mortality forecasting. Our paper has concentrated on model risk and has shown that uncertainties in mortality forecasts are financially material for annuity and pension providers, and for the policyholders in their care.

The Vice-President: When I was reading this paper I was reminded of an anecdote that I heard many years ago about a very distinguished actuary, still, happily, with us, although not present this evening. He was at the time the Appointed Actuary of a major Scottish life company and was asked by the company's auditor if he could say something about how confident he was that the actuarial reserves that he had just calculated were correct and how he could be thus confident.

https://doi.org/10.1017/S1357321700005687 Published online by Cambridge University Press

He replied in two words, which were, "Pure guesswork!'' This, you can imagine, distressed the auditor considerably, but, then, distressing auditors was one of the few innocent pleasures available to appointed actuaries at the time!

I thought that many, many years later with much more experience under our belts and techniques significantly developed, the authors are still able to write in this evening's paper that perhaps the actuary's role should be much more about demonstrating the depths of what we do not and cannot know.

I think we have had this evening a stimulating, thought-provoking and accessible paper, one which could be read and absorbed in quite a short period of time even by non-specialists; one which has taught us significant lessons about the use of models and, as one speaker said, not simply mortality models, but models more generally.

On behalf of the Faculty let me express my thanks to Stephen and Iain for their research, for the work they put in in preparing the paper and for presenting it this evening, and may I thank also everyone who has contributed, and particularly Stephen Makin for closing the discussion on what has been a very useful and very stimulating evening.