
A Value-at-Risk framework for longevity trend risk

Abstract of the Edinburgh discussion

[Institute and Faculty of Actuaries 19 November 2012]

Contact

Stephen Richards, F.F.A., E-mail: stephen@longevitas.co.uk

This abstract relates to the following paper: Richards S. J., Currie, I. D. and Ritchie, G. P. A Value-at-Risk framework for longevity trend risk. *British Actuarial Journal*, doi:10.1017/S1357321712000451

Mr G. C. Wood, F.F.A. (Chair): Dr Stephen Richards will introduce the paper on behalf of the authors, after which Mr Douglas Philp will open the formal discussion.

Dr S. J. Richards, F.F.A. (introducing the paper): Longevity trend risk is most commonly found in annuity portfolios and defined benefit pension schemes. The risk is that mortality rates improve faster than expected. A loss is incurred because the insurer or pension scheme pay pensions for longer than anticipated. The paper addresses two related questions posed by the holders of this risk:

- How might expectations of future mortality trends change over a single year? and
- What financial impact could these changes have?

These two questions might be the definition of a one-year, value-at-risk assessment of longevity trend risk. However, this risk unfolds over a long period, specifically the lifetimes of the annuitants, so a one-year, value-at-risk approach is not a natural way to view longevity trend risk. Despite this, regulatory requirements in the shape of the individual capital assessment (ICA) in the United Kingdom and Solvency II in the European Union push insurers to consider all their risks through a one-year prism, not least because many other risks are sensibly viewed in this way. The paper therefore grew out of the regulatory need to put longevity trend risk into a one-year, value-at-risk (or VaR) framework and, more specifically, to estimate the capital required to cover at least 99.5% of scenarios arising in one year.

We believe we should use stochastic projection models for this task, in line with other experts. Writing about the solvency capital requirement (SCR), Börger (2010) stated: “The computation of the SCR for longevity risk via the VaR approach obviously requires stochastic modelling of mortality”. On the same subject, Plat (2011) wrote: “Naturally this requires stochastic mortality rates”.

It is worth illustrating the merits of stochastic projection methods. We can fit a Lee-Carter model to the mortality experience data from 1961 to 1992. The Lee-Carter model would have been state-of-the-art for mortality projections in 1992. We can then plot the central Lee-Carter projection, together with the equivalent contemporaneous CMI¹ projection.

¹ CMIB (Continuous Mortality Investigation Bureau) (1999) Report Number 17, Institute and Faculty of Actuaries

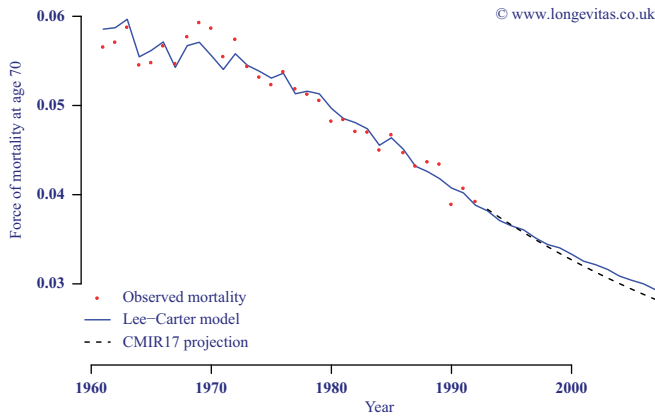


Figure A.

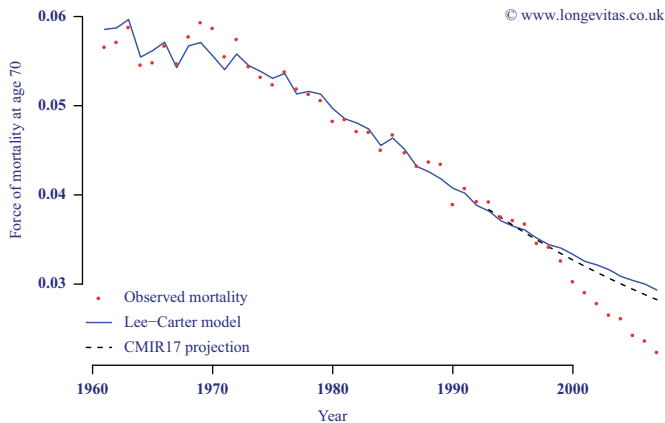


Figure B.

The resultant graph in Figure A displays substantial agreement between the two approaches, at least at age 70. The CMI projection is more prudent, as expected of an actuarial projection. A comparison of the two projections with actual mortality rates experienced since 1992 is shown in Figure B.

The figure shows that both projections were wrong, because of the emerging “cohort effect”, as introduced to the actuarial profession by Willets (1999)². However, although both models were wrong, the statistical model still has value in the confidence intervals, which are added to the graph in Figure C:

While improvements were faster than either method predicted, the confidence intervals of the Lee-Carter model would have correctly alerted actuaries to the possibility of improvements being faster.

² Willets, R. C. (1999) *Mortality in the next Millennium*, Staple Inn Actuarial Society, London

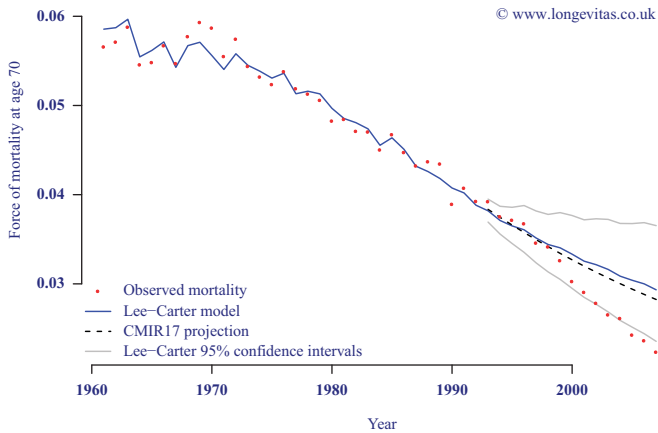


Figure C.

Stochastic models are useful for this kind of work, because they alert us to possibilities which we think are unlikely, and the purpose of insurance reserves is to guard against the unlikely.

Longevity trend risk, in either an annuity portfolio or a defined benefit pension scheme, is a long-term risk. An adverse trend will unfold over a number of years, and so a natural approach is to use a stressed-trend method, which can be used to calculate the extra capital required to be 99.5% sure of covering an adverse trend scenario under the model. This is done for various ages and models in Figure 2 in the paper, and shows that the amount of capital is dependent on the age at outset. Different models produce different results. Model risk exists, but is hard to quantify and requires actuarial judgment.

Figures 1 and 2 in the paper are the result of a stressed-trend approach to capital requirements, whereas modern regulation of insurance companies is orientated around a value-at-risk approach. By this we mean that reserves should be adequate to cover all events occurring over the coming year, barring those with a probability lower than one in 200. The question is then how to take a long-term risk such as an adverse longevity trend and put it into a one-year view.

The framework described in the paper does this in four relatively simple stages below:-

- a stochastic mortality model is used to simulate mortality rates for the coming year;
- those mortality rates are used to simulate the mortality experience of a chosen population;
- the new mortality experience ‘data’ is used to refit the mortality projection model; and
- the updated projection is used to calculate a financial measure, such as an annuity factor.

Figure 5 in the paper illustrates how central projections under a Lee-Carter (1992) model vary when refitted with one year’s new data. Only the first eight projections are shown for clarity; in practice, the process is repeated many times to obtain a set of annuity factors. This set can then be used to estimate the value-at-risk, for example estimating the 99.5th percentile and thus the minimum capital requirement for longevity trend risk under Solvency II. This is shown in Table 5, which again shows the importance of model risk. Figure 6 shows how the VaR capital requirement varies by age in a reassuringly similar manner to that calculated using the stressed-trend approach of figure 2.

Capital requirements in a VaR analysis are quantiles, that is, they are order statistics. As such, there is uncertainty about their value, and this uncertainty can be quantified. Figure 6 in the paper also shows the 95% confidence envelope for the capital requirement under the Lee-Carter model.

Increasing the number of simulations from 1,000 to 10,000 materially reduces the uncertainty, although the central estimate is not hugely changed. However, while reducing the uncertainty over the tail estimate is clearly desirable, it is important not to become too distracted with an ever-increasing number of simulations for two reasons; firstly, because models are only approximations, particularly when estimating the tails; the second reason is that model risk is material and different models must be run. It is considered more important to run VaRs for five different models with 1,000 simulations each than to run a VaR for one model with 10,000 simulations. Since VaR calculations are dependent on the discount function or yield curve used they should be run as frequently as there are major shifts in the yield curve.

The method described has been developed as a solution to the problem of putting longevity trend risk into a one-year view. However, the capital requirements produced will only be minimum floor values for two main reasons. Firstly, a one-year, value-at-risk view will under-estimate the true capital requirement for longevity trend risk because most models will not immediately respond to the beginning of a new longevity trend. This is a desirable property of a model, which should not respond immediately or fully to what might be a mere random fluctuation. Most fitted models are more heavily influenced by the existing data than the newly simulated experience, and so will only respond partially to new data. This is why the VaR capital requirements in Table 5 are usually lower than their stressed-trend equivalents in Table 3.

Secondly, a data-driven approach is unable to incorporate external events which have no precedent in the data. One example of such an external event is the recently announced revision to recent population estimates arising from the 2011 Census. Willets (2012)³ points out that mortality rates above age 90 are disproportionately affected by this.

In practice, therefore, it is anticipated that both practitioners and regulators will view one-year VaR capital requirements as a floor for longevity trend risk. Nevertheless, the framework in the paper is still useful – being able to set a minimum value for the capital requirement for longevity trend risk is a step forward. The framework offers two further practical benefits. The first is that it allows users to explore model risk; it can be used with any stochastic model capable of generating sample paths, and any projection model capable of being fitted to data. When projecting mortality rates, or calculating capital requirements for longevity trend risk, it is essential never to rely on any single model. The framework described here leaves the user free to specify which models to investigate.

The second benefit is that the framework can be used to test models for robustness. A life office will often want to focus on one or two models for everyday use. But a model selected today may cause problems tomorrow. The framework can be employed to test a model's reaction to new data, and to ensure that the resulting projections are stable. As the example in section 10 shows, not every model is robust to the addition of new data. It is useful to know this before investing resources in that model. Cairns *et al.* (2009) described a series of tests for selecting a mortality projection model. The addition of this framework is advocated as a test for model robustness.

³ Willets, R. C. (2012) Longevity and the 2011 Census, <http://www.longevity.co.uk/site/informationmatrix/longevityandthe2011census.html> [Accessed on 12th November 2012]

While the framework itself is complete, refinements are still possible. For example, whilst the method in the paper uses specimen single life annuity factors, it could be extended to allow for dependent spouses' benefits. Similarly, the method could produce an entire portfolio valuation instead of just annuity factors. This would allow for the distribution of liabilities by age and size. There is also the possibility of including a portfolio's own mortality experience.

The authors are of the view that the framework outlined in this paper is a useful step in gaining a one-year, value-at-risk view of longevity trend risk and would be interested in hearing the audience's opinions.

Mr D. J. Philp, F.I.A. (opening the discussion): The paper clearly sets out a practical method for implementing data-driven stochastic longevity models alongside a one-year VaR approach to measuring longevity risk.

In talking about the paper, I will aim to cover the theme of consistency with best estimate assumptions: the items that drive best estimate assumptions should be similar to those that drive changes to the stress models. I will also talk a little about consistency with market stresses, and using judgement.

As mentioned in the abstract, with a one-year VaR approach, we are considering how much assumptions might change over one year as a result of new information. There is a range of approaches in use to deriving best estimate longevity improvement assumptions, for example, Lee-Carter or one of the other projection methods set out in the paper. The approach in the paper for the one-year VaR framework sits logically alongside these. Within the general approach would be included a 'cause of death' model since these models can be fitted to 'cause of death' data.

There are other methods in general use for setting best estimate assumptions, with some companies using methods that rely significantly on judgement about future experience, for example, by considering consequences of recent and forecast medical advances, changes in social behaviour, NHS funding and so on. There are aspects of evolving Solvency II regulation that would encourage a company to adopt such an approach. For example, in the EIOPA Technical Specifications for the Solvency II valuation (dated 18 October 2012), the section on the calculation of best estimates asks users to "[consider uncertainty due to] future developments [which] shall include demographic, legal, medical, technological, social, environmental and economic developments including inflation." Similarly, in a section on deterministic techniques a statement is made that the application of deterministic techniques and judgement can be more appropriate than the mechanical application of simulation methods.

I would suggest that the 'method' of using the CMI 2009-2011 models is not a method at all and that these models are just a tool to express a basis using 'standard' terminology. Judgement is needed to decide what would affect mortality and then use the CMI models as a tool. Whichever method is used to derive the best estimate assumption, a desirable feature would be that the stress recognises the factors that would cause the best estimate to change.

Turning now to the issue with data. The data typically used within longevity projection models is mortality experience. There is an alternative view that we should be modelling data on historic assumption changes to derive a stress. Whether you could practically obtain that data or not is a different issue.

In the scenario in the paper, it is implicitly assumed that the assumptions are derived mathematically from experience data and the stress flows through automatically – not necessarily an inconsistency.

In respect of the section in the paper on the components of longevity risk, basis risk in models is prevalent. For example, people generally apply flat rate adjustments to standard tables, which can create funny rates at extreme ages. This is a form of basis risk. Converting a basis into standard terminology such as the CMI 2009-2011 models, creates risks.

There are also risks created by the requirement to publish results. For example, even if a longevity model gave a certain result, a company might not choose to use the result if it meant its basis looked weak relative to peers.

Also, sometimes companies will retain a best estimate basis until it moves outside a reasonable range. This is another reason why modelling experience is not necessarily a good proxy for modelling assumption changes.

The next section of the paper talks about the stressed-trend approach. It highlights that the natural way to think about longevity is in run-off. We cannot do that so we adopt a one-year VaR approach.

It is worth considering what we mean by one-year VaR and run-off and whether we adopt consistent approaches to market and non-market risk.

A one-year VaR for market risk is answering the question, “How much can the market’s assumption about the value of the future income stream from an investment change over one year?” The equivalent question for longevity is then: how much can a company’s assumption about future mortality rates change over one year? The only difference in approach is who is making the assumption: the market or the person calibrating the longevity stress.

If we follow this logic, this would lead a company to adopt the one-year VaR framework set out in the paper.

Section 4.7 shows that a stress trend taking the 99.5th percentile in run-off is onerous. The impact is higher if the 99.5th percentile is taken at every point in time. Even so, we would still need to use a lower probability in run-off to allow for this. As well as the method in section 4.7, there is a similar method, which was set out in GN46.

It is hard to argue that the shock approach is a robust method. The authors suggest this approach produces capital requirements that are too low at young ages and too high at old ages. Different portfolios should generate different stresses, depending on what the age mix is, and this method does not really do this.

The VaR approach described is consistent with the way projection models are derived and, if the underlying models are viewed as appropriate, the approach will produce capital requirements that are appropriate for any given portfolio of pensions and annuities.

The limitations of the general approach may not be as significant as the section 7 about model choices implies: the method does not require models that generate full sample paths (in a time-series sense). It only requires a range of probabilistic outcomes at a one-year time horizon and the

capability of re-fitting the model allowing for the extra year's data and generating a central projection that reflects this. That is brought out in the following section 8.

The general approach of considering how much assumptions could change over one year based on new information is also extendable to models where judgement is applied regarding the extent to which past data is a guide to the future.

The approach of being able to separate out trend risk and volatility is also a useful tool for sense-checking the output from the model.

As expected, the results for the one-year VaR approach are lower than for the stressed-trend approach. I would be interested in the authors' views about which model they prefer out of those tested, and the rationale for it.

Figure 1 mentions the AIC (a test of relative fit). A model should pass a test of goodness of fit and tests for absence of auto-correlation in residuals. Rather than plotting $\log(\text{mortality})$, which changes slowly, I have found it useful to plot the improvement rates themselves to help convince others that the projection pattern is consistent with the historic data. The difference between the ARIMA(0,1,0) model and the ARIMA(3,1,3) one in figure 7 would be even clearer on this basis. Figure 2 is an excellent illustration of the possible impact of model risk.

In conclusion, the paper presents a logical method for deriving a longevity trend stress for companies that are using data-driven projection models with no judgements applied post-fit. In practice, I would expect many of the large UK annuity providers to consider a range of approaches to setting best estimate assumptions and that many of these approaches require judgement post-fit. Even where the general approach is not to apply judgement post-fit, respectability considerations, benchmarking and audit may make it difficult to prevent doing so.

Mr A. J. Clarkson, F.F.A.: I congratulate the authors for an accessible paper. I agree with much of what Dr Richards said, in particular, how important it is to consider a range of models because you will get a range of answers.

I understand the reasons for developing a one-year VaR model for longevity risk. Nevertheless, I am not convinced that it is the appropriate framework for measuring and managing the risk relating to a long-term liability, such as longevity risk, for two reasons.

Firstly, the board and senior management's ability to understand what the key assumptions are in fitting a stochastic mortality model and to be able to apply appropriate judgement as to which one they are going to go with in the end. Secondly, – as Dr Richards touched on in his opening remarks, the key risk over the long term for longevity risk is a trend risk. A one-year VaR, as the paper points out in section 6.5, is essentially going to be dominated by volatility rather than trend.

In reality, I am not convinced there is a deep and liquid market for longevity risk. In that environment, companies need to consider the long-term nature of the risk.

I prefer a model based on future assumptions about medical advances, and so on, although this would be more difficult to demonstrate as compliant with quality standards under Solvency II and it is judgemental in terms of assumptions. However, it should be possible to persuade the regulator

that you can use such a model for your capital requirements under the ICA and also when applying for internal model approval under Solvency II.

In reality, do any of us know what a 1:200-year event is? The answer is clearly no. I do not believe a stressed-trend approach is any more or less likely to be right than a one-year VaR approach. A board is more likely to understand the key assumptions it is making when it is coming up with a stressed-trend model; in particular, the assumptions about what future medical advances the company might be able to survive given the capital it will be holding.

Adopting a stressed-trend approach for longevity alongside a one-year VaR approach for market risk or credit risk is, I suspect, mathematically incoherent. Nevertheless, it is a more appropriate approach in terms of practical management of risk. I would be interested in the authors' views.

Prof A. D. Wilkie, F.F.A. F.I.A.: There are a number of points that have not been mentioned in the introduction or discussion.

If you are looking at your own company's portfolio, and if there are unusually few deaths in the portfolio in this coming year, you have more survivors to value at the end of the year, regardless of what your assumption is. If, additionally, the fewer than expected deaths caused you to change your assumptions, you must, in a sense, double change. That may automatically be taken into account in the authors' methods, but they do not mention it. The long term would be considered in these adjustments.

Next, if at the beginning of the year you make a best estimate of the long-term assumption and allow for 99.5% variability, you will have two figures. In a year's time you might change your model, your best estimate and your 99.5% level. I am not quite clear whether we are measuring over one year the 99.5% amount of the change in the best estimate, or the best estimate of your change in the 99.5% long-term amount, or the 99.5% change in the 99.5% estimate, which may be double-counting.

It may be that the authors are entirely clear what they are doing and can explain it.

You do not have to use the same model for best estimates and for your variance: there is a consistency if you do. But you could, for your long-term best estimates, use something like the CMIR 17 or CMI projection models and then wrap round these a variance that you have taken from a Lee-Carter model, or some other model.

One approach is to choose the CMI data from some time back, consider the CMI forecasts based on the 1968 experience, the 1980 experience and the 1992 experience, and see what the variability in the actual experience was over ensuing years. For some of those periods the CMI has published figures. It would be useful to have them available to see how far away the experience was from the last forecast made. From these you could estimate the variability.

The simple model that we thought of some years ago was a bit like the model in formula 9, paragraph 8.2, where there is a single κ in the Lee-Carter model that changes from year to year.

Again, I am not quite sure what the authors are doing here. I would put in a stochastic drift with a random walk element and an annual blip that is not carried forward. The blip might be upwards or

downwards. However, for an insurance portfolio, unusual changes in mortality are more likely to be as a result of a serious Spanish influenza epidemic than be a sudden large reduction in the number of deaths per year. It is easier for the mortality to increase than to come down suddenly.

So, a two-level element, one of the stochastic drift, plus the unusual features of the year, might be a useful method.

The authors do not explicitly say that they simulate the parameter values. I would like to use, for the purposes of this discussion, what I would call a 'hyper model'. You use a model, estimate the parameters and then, instead of using those parameters in simulations, you simulate the values of those parameters within each simulation. The parameters are only estimates and you do not know their values exactly from the past data. You have a confidence interval. If you have used maximum likelihood estimation, you have a full covariance matrix of the parameters. Therefore, you can think of the parameters themselves as possibly being multivariate normally distributed. You may have to make adjustments for ones that can be positive only or are in the range (0,1) or something similar. You can thus explicitly allow for parameter risk within each of your models.

It is harder to weight the model risk, although it is important. One of the problems is that, if you look at five models, the cautious actuary would use the most conservative one and the manager whose company is a bit short of cash would use the least conservative one. It is purely subjective as to which one is used.

Dr I. D. Currie (responding): Prof Wilkie's point about the variation in the parameter estimates suggests, as an example, simulating from the normal distribution of which they are realisations. In effect that is what we do, because the confidence intervals for the best estimate for the trend risk are computed exactly, taking proper account of the covariance matrix of the estimates of the underlying coefficients.

Mr A. D. Smith (student): This paper is especially helpful because of the way the authors have reduced several models to a consistent mathematical form and have tabulated the results consistently across models.

As Dr Richards has already pointed out, much of the longevity work until now has focused on the emergence of trends over long future time periods. The main contribution of the paper is the ingenious method of producing a one-year VaR starting from those much longer projections, which is essentially to fit the model with and without the last year of data, and then randomly re-simulate that last year of data to see how the model could have turned out had the last year of data being different.

That gives you the distribution of possible basis changes but would not have captured many of the real reasons for basis changes that have happened in the past. Those big basis changes were often due to, for example, introducing two-dimensional rather than one-dimensional mortality tables or recognising the cohort effect. In life rather than annuity business there were changes due to AIDS tables which came and then went. Not one of those was actually driven by changes in an extra year of data. So we have to be aware of the limitation that these extra sources of changes may not have been picked up in the method the authors propose.

Table 5 gives some interesting indications of the model risk. The stressed values of the Cairns-Blake-Dowd models were lower than the base value for the two models that incorporate a cohort effect.

This shows that if you had just focused on the Cairns-Blake-Dowd models and used them to construct your 99.5% confidence interval, you would have been missing out on a quite plausible possible change of model that your actuary might have decided to apply next year. That highlights the model risk.

I looked at table 3 and table 5 to compare capital requirements on a one-year basis with those on a run-off basis, and was surprised at the results. I expected the one-year numbers to be much smaller. With the exception of the Cairns-Blake-Dowd models, according to this paper, the resulting stress on a one-year basis is more than three-quarters of the size of a stress on a run-off basis. So you need a pound's worth of capital to run off but you need 75 pence more just to survive the first year. This surprising result deserves more explanation. For the *P*-spline model, the results set in the paper seems to suggest that you need more capital to survive one year than you need to survive forever.

Another kind of run-off model, which would be consistent with the authors' one-year VaR approach, is to start with a model fitted to past data, generate a random new year of data, refit your model, simulate the next year's data from the new model and then refit a third model at the end of that year, so your model is evolving over time. If you compare the capital required under this iterated approach to the one-year VaR, you might end up with the situation that you need three to five times as much capital to run off as you need over one year.

That iterated approach leads to what some people have called forward rate models where you are studying the whole curve of forecast mortality rather than just a number of deaths in a given period. Such models allow external valuations that have not just come from data. Various others have been mentioned already like changes in causes of death, and so on. Mr Phillip Olivier and I published a model in 2004. Andrew Cairns, in 2006, put together a synthesis of all the forward rate models that were available and I am disappointed not to see reference to those models in the paper. By design they naturally address a one-year VaR since you do not have to go through some contortion to turn a long-term projection into a one-year VaR. As you might guess from the name, 'Forward rate' models are similar to the models that will be used to address, for example, interest rate risk. So it is easier to show consistency between the longevity and the interest rate model components.

Mr P. J. G. Ridges, F.I.A.: The authors have simulated one year's data, and to estimate the one-in-every-200-years scenario have then looked for the one-in-200 event. If instead of one year, five years of data is simulated and then balanced that by looking for a five-in-200 scenario, would that give broadly the same amount of capital required?

Dr Richards (responding): I would not know until I ran the calculations. However, with the way the framework is set up, there is nothing to stop someone from simulating five years' worth of data and performing a five-year VaR calculation.

Mr. Smith asked why the 2DAP model produced one-year VaR capital which was higher than the stressed-trend one. The reason for that is that the 2D *P*-spline models are more responsive to an extra year's worth of data than most of the other models (and possibly too responsive).

I agree with the points made about the model not capturing 'real world' life office changes in bases. There are other reasons for bases to be changed, which is why this approach is regarded as appropriate for a minimum capital requirement.

Prof Wilkie made a point about an adverse scenario over one year having two components: a stronger trend leading to a capital strengthening and the problem with more people surviving. This is idiosyncratic risk in table 2. What he was talking about would be the risk of adverse experience over one year, which would be handled as a separate item in the ICA or in Solvency II from the trend risk that is the subject of the paper.

Prof Wilkie was also right that there is no reason why you have to use the same model to generate the sample paths and refit the models. There is nothing to stop you from generating the sample paths with, say, an Age-Period-Cohort model, but using a Lee-Carter model to do the fits.

Prof Wilkie asked which one of three definitions of the 99.5th percentile the authors used. It was the first one he itemised: we are showing the 99.5th percentile of the change in best estimate. Mr Clarkson mentioned cause-of-death projections as a possibility. I have a number of issues with such projections, and some are itemised in Richards (2010). One of the biggest problems is the handling of the correlations between the various subcategories. This applies irrespective of whether or not it is a cause-of-death-based model, or else a model somehow seeking to explain future medical investigations. The issue is basically mathematical. Carriere (1994)⁴ showed that a proper cause-of-death elimination or projection is impossible without knowing the correlation structure between the various sub-categories being used. Carriere (1994) then went on to show that it is impossible to know this correlation structure. This problem applies to any attempt at disaggregation of mortality statistics. In practice, I often see the assumption of independence between the various causes being used. This seems to be an assumption that is made for the convenience of the model builder rather than there being any evidence to back it up. This is an issue if there are positive correlations between the sub-groups, because the model will understate the uncertainty over the future mortality trends and could therefore lead to inadequate capital requirements.

Mr W. D. B. Anderson, F.I.A.: Mr Philp mentioned some philosophical points at the beginning of his remarks about what it was that we were trying to model in the first place. In the spirit of the Solvency II regime, the attempt is to try to make sure that the insurance company has enough money at the end of the year to offload its liabilities to a third party.

Conceptually, what we are trying to model is a step change function in what the market consensus view is of actuaries' expectations of how the future trends are going to look. The actuarial tables that the Institute and the Faculty has issued have been characterised by a series of step change effects.

It may be spurious to spend time on the mathematics of the models when there are going to be big changes because of medical breakthroughs or, perhaps, evidence that trends in mortality improvements are not decelerating at the rate that the markets assumed. It is difficult to see as a board member why you would invest money in this over and above other priorities. Even though the mathematics of cause-of-death modelling might be awkward, philosophically it is much easier for non-specialists to comprehend.

Mr M. Selby: This is a question for the authors: at the very tails of the mortality distribution, that is, the worst one-in-200-year event, is there any merit in thinking about how mortality itself might be correlated with the economic scenarios?

⁴ Carriere, J. F. (1994) Dependent decrement theory, *Transactions of the Society of Actuaries*, Volume XLVI, 1–21.

If you have the worst one-in-200-years event from an economic point of view you are probably not going to have your best one-in-200-years situation from a mortality point of view. There might be some correlation, especially at the end.

The Chairman: In section 10, ‘a test for model robustness’, provides for a test of resilience or volatility of an individual model and that it helps give a feel for the robustness of the calibration. It is difficult to see how it gives information about the appropriateness of one model versus another, unless there is perhaps more inherent instability in some types of model in than other types of model. For example, we heard about the Age-Period-Cohort model that seemed to give anomalous results for the one-year view versus the trend view.

I would be interested in the views of the authors as to whether there were any models that had greater instability and therefore whether the one-year VaR approach could give some sensible information that would help inform model choice.

Dr Currie (responding): This is quite heavily connected with the idea of model complexity. Some complex models have many parameters and involve forecasting many different components. It is almost always the case that, as the model complexity builds up, the model fit will improve but the volatility of the forecasts is liable to explode, which is a downside to improving the fit of a model.

Mr R. Austin, F.F.A.: It is probably useful to highlight the importance of expert judgement, which is one of the key areas in terms of Solvency II for many firms.

One of the things that Solvency II is bringing out is that, in the past, such judgements have been hidden in a black box of calculations. This is a good example of where the rationale behind the judgement should be captured. While we can use a data-driven approach to come up with longevity assumptions, it is important to consider the potential different sources of information that are used in forming a judgement. One of them will be the data-driven approach.

Dr Currie: Expert judgement has a poor record in terms of forecasting mortality. There is a famous paper which was published in *Science* 2002 by Oeppen & Vaupel.⁵ The authors plotted the maximum female life expectancy or record life expectancy, over all the countries in the whole world from 1840 to 2000. In 1840 the record life expectancy was 45 years, which was held by Swedish females. In the year 2000 the record had risen to 85 years and was held by Japanese women.

This plot was almost precisely linear, so there was an inexorable rise in this record life expectancy.

On this plot, they also marked the location of various predictions for the maximum future life expectancy and the dates when these predictions were exceeded.

They commented at the end of their paper that “the linear climb of record life expectancy suggests that reductions in mortality should not be seen as a disconnected sequence of unrepeatable revolutions, but rather as a regular stream of continuing progress.”

Prof Wilkie: I recall an earlier paper, possibly Eileen Crofton, which looked at the distribution of causes of death among cohorts which had more or less completed their lifetime, so starting probably with 1840 births.

⁵ Oeppen, J. & Vaupel, J. V. (2002) Broken Limits to Life Expectancy. *Science*, Vol 296, 1029-1031

Results revealed that the proportion of people who had died from respiratory problems had remained reasonably constant over that time. In the past it had been tuberculosis and at the time she was writing it was lung cancer.

The point that Dr Richards was making about the correlations is that if people do not die from lung cancer, are they going to die from some other respiratory problem fairly soon. That is the sort of difficulty that one cannot deal with. Causes have varied so much over long periods, for example, nobody dies nowadays from consumption or apoplexy. The terminology has changed over the years, so you cannot use causes over long periods of time.

Trying to assess the one-in-200 level is exceptionally difficult. We should be telling the FSA that this cannot be done practically.

The authors may have used a normal distribution, as I would, for many of the estimations. If you were to use a fat-tailed distribution instead, the 99.5% levels might go much further out.

Taking the other example that I mentioned earlier, the possibility of a serious Spanish influenza epidemic that various vaccines did not prevent, somebody might estimate that the chance of it is about one in 190, and so will include it. Somebody else might say it is really one in 210, and so the answer is zero.

I had, in a general insurance context, a problem where, depending on the values of the parameters I used, the one-in-200 reserve was either zero or 2 billion; the difference in the probability was quite marginal, but the difference in amount made a difference.

The Chairman: I now invite Dr Richards to respond to the discussion on behalf of the authors.

Dr Richards (responding): Mr Selby posed a question regarding potential correlation between extreme mortality scenarios and extreme economic scenarios. It is easy to imagine that the two may well go hand in hand, but that is not the subject of our paper. Mr Makin may have some comments about the correlations that you may have to assume when aggregating the capital that needs to be held for these individual risks.

Mr Wood asked about the VaR process being able to recommend models. It cannot tell us if a model is any good, only if it is robust to new data or not. I recommend the approach set out in Cairns *et al.* (2009), which gives a series of quantitative measures for assessing how useful a model might be. We would suggest just using the value-at-risk framework as a way of testing robustness in addition to the sorts of tests that Cairns *et al.* (2009) proposed.

On the question of stochastic models and judgement, all models are approximations. An important element of the paper that we have not discussed so far is that all of the models that we have used are published models from peer-reviewed academic journals: Lee & Carter (1992), the Age-Period-Cohort model, Cairns *et al.* (2006) and the *P*-spline model from Currie *et al.* (2004). Each of these models has its own strengths and drawbacks. None of them is uniformly better than any of the others in every single area of consideration. However, the benefit of these models is that their strengths and weaknesses are all relatively well understood.

It is human nature for people, even actuaries, to lay great store by their own judgement, but it is tempting to downplay the likelihood of alternative adverse scenarios, especially when those adverse

scenarios are associated with extra capital requirements. For us, one of the benefits of using stochastic models is that they limit the scope for false certainty when there are strong financial pressures demanding it. As a caution, I quote Booth & Tickle (2008): “The advantage of expert opinion is the incorporation of demographic, epidemiological and other relevant knowledge, at least in a qualitative way. The disadvantage is its subjectivity and potential for bias. The conservativeness of expert opinion with respect to mortality decline is widespread, in that experts have generally been unwilling to envisage the long-term continuation of trends, often based on beliefs about limits to life expectancy.”

Booth and Tickle go on to list several examples where expert opinion has underestimated mortality improvement in the past. Expert opinion can be valuable, but it has to be used carefully and sparingly. A particular point to remember here is that the purpose of the VaR method is to estimate tail risk. This could not be done if the model being used contained any kind of built-in limits to cumulative mortality improvement, or any limits to human lifespan. If a model contains this kind of assumption, then the tail risk will merely reflect the assumptions of the modeller.

Dr Currie has already mentioned a paper by Oeppen & Vaupel (2002). They wrote about the dangers of expert judgement: their final words are: “Experts have repeatedly asserted that life expectancy is approaching a ceiling: these experts have repeatedly been proved wrong.”

We can see the dangers of relying too much on expert judgement. However, this is not to say that there is no role for expert or actuarial judgement, as this judgement is still much required. It is just a question of at what point in the process it is used. For example, for a user of the framework in our particular paper there is still plenty of judgement required. Which models do you want to use, and how many simulations are you going to use? How do you go about collating the outputs and how do you come to a decision given that different models give different results? Which capital requirement are you going to pick, bearing in mind the different plus and minus points of the various models, and other aspects like model risk and basis risk? There is plenty of scope for actuarial judgement, even within this particular framework.

Mr S. J. Makin, F.F.A. (closing the discussion): The paper set out to answer two questions. The first was how might expectations of future mortality improvements change over a one-year period. The framework presented was information-theoretic in nature and set out to consider how we might respond to new information that emerged over a one-year time period.

The second question was what financial impact do these changes have. The resulting figures are set out in the paper and have been discussed here with Mr Smith going into some detail.

All of the figures were calculated by reference to a VaR risk measure. This approach generalises and can be used with other risk measures, such as conditional tail expectation. Particularly interesting would have been the increase in the number of simulations needed to bring the simulation error down to some sort of tolerable level on other risk measures.

On that subject, the paper puts forward 1,000 simulations as the minimum tolerable number for 99.5th percentile value-at-risk work. My own take on that is that 1,000 is not enough. 5,000, perhaps, might be closer to the mark, particularly if considering a number of risks, whereas the 10,000 simulations that the authors showed in their supplementary chart is perhaps closer to the mark for a single risk factor.

One of the comments made by Dr Richards in his opening remarks was that it is important not to spend too much time worrying about the number of simulations. I don't agree with that. Given a model that is to be used to calculate capital requirements, companies want to have some certainty on what the resulting capital requirement will be and, in particular, they do not want volatile capital requirements simply because of simulation error. I would distinguish that point, which relates to simulation error, from the effect of model risk, which does not.

One of the sentiments coming through from the paper, and also Dr Richards' introductory remarks, is that the framework requires a model-driven approach. I do not quite agree with that either. As presented, the framework requires a one-in-200 one-year view of a change in mortality experience together with a means of deciding how the trend assumptions would respond to that, and judgement can unquestionably fulfil a role here. To reinforce that in a Solvency II context, the emerging rules do not require the use of a data-driven stochastic model to generate distributions.

For those who do favour judgement, but are perhaps sceptical as to the role of data-driven techniques, any judgement cannot be truly robust if it does not have regard to the sort of modelling techniques that have been presented. On the one hand, we should have the application of extrapolative approaches informing judgement, but on the other we should have judgement informing the application of extrapolative approaches. The authors pointed out that a data-driven model cannot incorporate external events which are not referenced in the data. That is true, and I therefore like their resulting stance that the results can be adjusted for concerns in data or perhaps having regard to medical advancements. Dr Richards suggested the results were perhaps an underpin, and for practical purposes in a 99.5th percentile regulatory VaR context, that is probably right; any adjustment would probably be a strengthening.

As Mr Philp, opening the discussion, pointed out the method as presented does not require models which generate sample paths beyond one year. That is true in the context of the authors' approximation. Solvency II considers best estimate liabilities to be a probability-weighted average of all future outcomes, so a technically pure implementation of the one-year VaR framework requires nested stochastic calculations with projections in run-off. However, the authors' approximation, deliberate or otherwise, is entirely proportionate.

Mr Anderson made a point about the philosophical purpose of Solvency II and what it was trying to achieve. The purpose of the risk margin in Solvency II, which we have not touched upon notwithstanding that longevity trend risk is arguably the most non-hedgeable of all risks, is to turn a best estimate liability into something of a transfer value. It is then the purpose of the capital requirement on top of that to try to ensure adequacy 199 times out of 200 over the year ahead.

There was a sense in the discussion that looking at past assumption changes is a good idea. We heard some of the reasons why assumption changes might not be entirely pure and so we need to be wary of this in making any inference about the future from past assumption changes. Solvency II is founded on sound, technical provisions and that is a more preferable starting position when considering capital requirements than the consideration of changes in historical, and as a result possibly impure, trend risk assumptions.

On Mr Selby's question on correlations between longevity and market risks, there are two separate considerations. In the context of the framework and models presented, what drives the trend risk is a volatility process, and that process, and hence trend risk itself, is not correlated with anything.

The second consideration is in run-off, over which period it is easier to imagine a correlation between longevity risk and market risk. An argument commonly put forward is in connection with credit risk, where an increase in longevity creates pressure on companies' finances – and hence default risk – through more expensive defined benefit pension scheme obligations.

The various innovation processes and error terms described in the paper are all normally-distributed. That comes from the maximum likelihood approach taken where, through a combination of ease and habit, all of these were taken to be normally-distributed. However, the normal distribution has had a hard time recently, including in a recent speech titled 'Tails of the Unexpected', by Andrew Haldane and Benjamin Nelson at the Bank of England⁶, which concludes: "Normality has been an accepted wisdom in economics and finance for a century or more. Yet in real-world systems, nothing could be less normal than normality. Tails should not be unexpected, for they are the rule."

This does not mean that you should delete all normal distributions from spread-sheets and other applications, but that you think hard about the appropriateness of the normality in future quantitative work. The framework as presented can easily be extended to do that. One approach is to directly use non-normal innovations in the maximum likelihood process. However, there is an alternative approach, which is to use what are sometimes called quasi-maximum-likelihood techniques. The approach is to first fit a maximum likelihood model, and to then go back for a second pass round to try to formulate a better description of the errors that are unexplained by the first-pass fitted model. The reason why I like this approach is that it gives proper prominence to the role of innovations and the importance of describing errors accurately.

Returning to the information-theoretic nature of the paper's framework. The context is still a longevity-related one but that of longevity mis-estimation risk as opposed to longevity trend risk. Uncertainty exists over portfolios' actual underlying mortality rates since these can only be estimated with a degree of confidence linked to the scale and richness of the data. So there is a so-called mis-estimation risk, that is, the risk of mis-estimating current – or period – rates of mortality owing to statistical fluctuations.

The issue with longevity is that it emerges in run-off. We cannot know, as Mr Clarkson pointed out, whether movements in observed experience data are because we just have observed the one-in-200 event, as a result of mis-estimation risk, trend risk, or volatility risk, or as a result of any of the other elements of risk pointed out in the paper.

The best solution for setting best estimate liabilities may be to use opening mortality assumptions as opposed to trend risk assumptions and smooth year-on-year experience variances by averaging what has been observed over recent years, say the last three or five years, to arrive at a point-estimate of current rates. That seems sensible and is fairly common practice.

More care will be needed when determining the capital requirements. A direct read-across of the paper's framework might be to determine the one-in-200 one-year change in mortality as a result of mis-estimation risk, and then calculate how much assumptions would change in response to that. If the one-in-200 mis-estimation event did happen we would not know it, and so technical provisions would move in line with either one third or one fifth of that change, depending upon how

⁶ Source: <http://www.bankofengland.co.uk/publications/Documents/speeches/2012/speech582.pdf>

one smooths the experience variance. It could then be argued that that is acceptable as a means of determining a capital requirement for longevity mis-estimation risk because certainly the definition of the Solvency II Solvency Capital Requirement, is the one-in-200 one-year change in net asset value.

I have a difficulty with that at a philosophical level, and my point, which relates to the information-theoretic concept is the mis-estimation stress contains much more information than just a one-in-200 one-year change in the level of mortality rates. By hypothesis, it also contains the information that that change is permanent because, if it did not, the change in mortality would be a transient volatility stress.

To ignore this information would be wrong. The capital requirement for longevity mis-estimation risk should be determined by taking through the full impact of the assumption change and not just one third or one fifth of it.

I agree with the view that it is appropriate to consider how one's assumptions would change over one year according to new information; but the point to be clear on is what information it is that we are talking about. In the case of mis-estimation risk, do not overlook its permanence.

Chair: It remains for me to express my personal thanks, and the thanks of all of us, to the authors, the opener and closer and all those who participated.