

# RECORD Volume 30, No. 3\*

---

Annual Meeting and Exhibit  
New York, NY  
October 24-27, 2004

## Session 133 TS

### Actuarial Research Meets Actuarial Practice: Mortality Risk Modeling

**Track:** Education & Research

**Moderator:** Mary Hardy

**Panelists:** Samuel H. Cox  
Iain D. Currie<sup>†</sup>

*Summary: The manner in which the actuarial profession models and manages mortality risk is undergoing dramatic changes. New modeling techniques have led to better risk management tools. Capital markets solutions are evolving for insurers wishing to protect against unfavorable mortality developments in both life insurance and annuity products. In this session recent key developments in mortality modeling and in the transfer of mortality risk are examined.*

**DR. MARY HARDY:** We have two speakers. Iain Currie is a reader at Heriot-Watt University in Scotland, in the department of actuarial mathematics and statistics. Iain is going to talk about the work that he has been doing on mortality, which has already had a significant effect on the assumptions used in U.K. pricing. Then Professor Sam Cox from Georgia State University will talk about securitization of mortality risk. We're looking at two different aspects of mortality in this session. This is a standing education and research session; we like to bring to the annual meeting sessions that indicate the practical applications of research in the discipline of actuarial science.

**DR. IAIN D. CURRIE:** I'm delighted to be able to present some work that I've been doing with my colleagues, James Kirkby, a student of mine; Maria Durban, who works in Madrid; and Paul Eilers of Leiden University. We met last month at the Research Institute in Edinburgh to put the final touches to this talk.

---

\*Copyright © 2005, Society of Actuaries

<sup>†</sup>Iain D. Currie, not a member of the sponsoring organizations, is at Heriot-Watt University in Edinburgh, United Kingdom.

**Note:** All handout materials are available through the link on the table of contents.

While I was doing some background research for this talk, I thought I would look at some of the history of the problems of longevity. The earliest example I could find for problems in annuities actually occurred in the Old Testament. In Genesis 5, you read that Methuselah lived for 969 years. What is much less widely known is that he had a hefty annuity with Red Sea Life. This company went out of business because it couldn't meet this liability. So already 4,000 or 5,000 years ago, depending whether you counted from his death or from his birth, insurers were having problems with longevity. On the other hand, Shakespeare had a much more optimistic view of the longevity problem. In "Macbeth," Act II, Scene II, Macbeth proclaims, "There's nothing serious in mortality. All is but toys." That was 1606.

A little more recently, the mathematicians started getting in on the act. A famous actuary, Benjamin Gompertz, writing in *Philosophical Transactions of the Royal Society of London* in 1825, made the amazing observation that over a long period of human life span, the force of mortality was geometric in age. This was a surprising, simple observation. Then in your own *North American Actuarial Journal*, which is more or less up-to-date, Lee had a paper entitled "The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications." These two works are separated by almost 200 years.

Gompertz, on the one hand, has a model for mortality for a single year of life. The Lee-Carter model is much more ambitious in some ways. It's a model for a mortality table, so you have a table of observed mortalities that's indexed by age and by year. The two models are different in nature. The Gompertz model is a linear model on the log scale, so it's a smooth model. The Lee-Carter model is a discrete model, where each age and each year gets a parameter. What I'm hoping to do in this talk is try and fit the ideas of Gompertz and Lee-Carter together to give us a smooth Lee-Carter model. Perhaps if there's time at the end, I'll indicate some further extensions, as well.

Here's the plan of the talk. I'll outline some data so that you can have something to fix in your mind. We'll go through Gompertz models and show how we can make smooth models out of these. The idea is basically just simple regression with a modern twist. I'll describe the Lee-Carter model and how we can make that a smooth model. I'll describe some technical difficulties (there may not be time for that), and finally, if time allows, I'll talk about flexible two-dimensional methods.

This is something you should keep in your mind for the duration of the talk. I'm going to be using some data on assured lives' mortality, so the data are claims on assured lives. There are 90 ages (roughly, not exactly), from ages 11 to 100, and we have observations from 1947 to 1999, which is just over 50 years of data. That's the number of claims, and matching that, we have the exposures (exposed to risk). If we divide one table by the other, we can get a raw estimate of the force of mortality at each age and in each year. The big question for pricing is, what is

mortality going to do in the future? It's a large question mark. Because of the requirements of insurers (they are daunting), we are expected to make a guess, an intelligent guess with any luck, about what the force of mortality is going to be in perhaps even 50 years' time. That's an ambitious program. One of the main things I want to say here is that one should be aware of that difficulty.

Let's go back to 1825 and Gompertz' famous equation,  $\mu_x = AB^x$ , showing that the force of mortality is geometric in age. Being a simple mathematician, I'm not comfortable with multiplication, so everything will be done on the log scale. That will turn it into a nice little addition,  $\log \mu_x = \log A + \log B x = a + bx$ . Thus, the log of the force of mortality is linear in age. Once you've got that, you can easily make it more complicated. You could have what I would call a quadratic Gompertz model,  $\log \mu_x = a + bx + cx^2$ , or cubic or whatever you like.

Insurance data are different from population data. With population data, you have large numbers of observations. However, with insurance data, particularly at younger ages, you can often have small numbers of deaths and small numbers of claims, so you ought to use an appropriate model here. An appropriate model would be a Poisson model. The number of claims follows a Poisson distribution, whose mean is given by the exposure times the force of mortality, or  $d_x \sim P(E_x^c \mu_x)$ . If you express that in the log scale, it simply says that the log of the model force of mortality is going to be linear,  $\log E_x^c \mu_x = \log E_x^c + a + bx$ , or quadratic,  $\log E_x^c \mu_x = \log E_x^c + a + bx + cx^2$ , or whatever. That's a regression framework.

What we're interested in is the expectation of the number of claims. It's a linear function of the explanatory variable, year of age. So that's the regression equation, except that it's a little trickier than that because we've got a Poisson variable, and that gives us a generalized linear model. Just think of it in terms of an ordinary regression with a Poisson variable. When I take one year of data—Benjamin Gompertz would have understood this clearly—and we're on the log scale, from roughly age middle 30s all the way up to age 80, the log of the force of mortality lies beautifully on a straight line, which is Gompertz' famous observation. It goes haywire at either end, but for a large span of life, you have linearity. There's nothing particular about saying that you have to do this for a single year of life; you could also do it for a single age. Then you would have a Gompertz model in time. Say we've taken age 65. The log of the forces of mortality at age 65 of the linear Gompertz function is rather unsatisfactory, but if you put a quadratic through it, it works well (see Currie, page 8).

That's the past. The reason you're here is that you'd like to know what's going to happen in the future. The quadratic model seems to do a decent job, so we will just continue it. You have to make a choice here. How are we going to project into the future? This is the choice of forecast function. One classic way of proceeding is to get a good fit to the data and then use the same function in the future. If that's

what is going to happen, that indicates the most extraordinary improvements in mortality over the next 50 years. You don't have to choose a quadratic. Who exactly knows what's going to happen? Instead of continuing with a quadratic, we could just fit the tangent at the last year, and there's a linear extrapolation (see Currie, page 11). Perhaps people are a little more comfortable with that.

Because we're in a nice statistical framework where all the theory has been worked out, we can also calculate confidence intervals. This is interesting, except that it's exceptionally misleading. The usual idea with a confidence interval is that if you can compute a confidence interval, you can be pretty sure that your predicted values are going to lie in the confidence interval. But 50 years out is a pretty confident prediction. If we thought that we could predict things as accurately as that, I think we'd all be extremely happy. What has gone wrong? A confidence interval is always computed conditional on a model being correct, so we have a strong model assumption here. We've assumed that we have a Poisson variable. We've assumed that the force of mortality follows a simple quadratic function. There's a massive assumption going in, and that propagates through into the future and gives us a tight confidence interval (see Currie, page 12). I think that's optimistic indeed.

Let's think a little more about regression. This is a first course in statistics. What is regression? In linear regression you have two basis functions,  $a+bx$ . It's linear combinations of 1 and  $x$ . Quadratic regression is linear combinations of 1,  $x$ , and  $x^2$ , and so on for polynomial regression. You can calculate the means as linear combinations, and you can also express that neatly in matrix form. This is a rather nice idea. Regression is simply a linear combination of basis functions. Then you make a little leap (or perhaps a big leap) and say that we don't have to use simple basis functions or polynomial functions, that we could use any functions and then just use the same ideas.

Here are these things that I've called "B-splines" (see Currie, page 14). I'll tell you what they are in a moment, but it doesn't matter. These are just functions that are replacing 1,  $x$ , and  $x^2$ . We've replaced them with  $p$  functions. We can work out the means in exactly the same way. We may express it in matrix form in exactly the same way. The question is, can we choose these functions in a better way than just using 1,  $x$ , and  $x^2$ ? That's what you always do when you're in your first year of college, but maybe there are other ways of doing this that might have some advantages.

I'm a great fan of using the B-spline functions as the basis functions. A B-spline is well-behaved; it's not an intimidating function at all. This is a cubic B-spline (see Currie, page 15). It consists of four pieces bolted together at dots, which are called knots. Each piece is a cubic. It looks beautifully smooth in the drawings, but it's not perfectly smooth. The first and second derivatives are smooth at the knots. You get this function that's almost a smooth function; it has little hiccups in the higher derivatives, but to the eye it looks perfectly smooth. We have a whole collection of

these, and this is the basis that I'm going to use. I think there are 10 basis functions here (see Currie, page 16). We could say that the force of mortality in the log scale is going to be a linear combination of these basis functions in precisely the same way it would be a linear combination of 1,  $x$ , and  $x^2$ . There is no change here, apart from using these different basis functions.

So we put it into an ordinary regression, nothing fancy (see Currie, page 18). How does it work? Disaster. I've used 23 basis functions, and the basis functions are covering the range of the time axis densely. The regression is able to pick up the bobs and weaves in the actual observations. If you use 30, you will get an even wavier function. If you used fewer, it wouldn't wave about so much. That would be one idea. How could you manage to choose the right number of basis functions? That's one approach, but it's not what I do.

Here's a nice observation. The red dots are the regression coefficients. These are plotted at the centers of their own basis functions. Now you can see what has gone wrong. The reason that the fitted function is bobbing and weaving around is that the coefficients bob and weave around, so if we could stop them from doing that, we would get a nice new fit. My colleague Paul Eilers and his colleague Brian Marx had this idea that they should penalize bobbing and weaving in the coefficients. They have this idea of penalties. Here's a first order and second order of penalties (see Currie, page 19). We're just taking the coefficients, writing them out in a line and defining a penalty function. I don't want the first one to be too different from the second, the second to be too different from the third and so on. That's a first order of penalty, and I can do the same thing with second order of penalties, as well. That's a simple function, and it's a quadratic form, so it could be written neatly in matrix notation.

How do you balance the penalties (the smoothness requirements) with the fit? We've got two things that are competing head to head against each other now. We've got the data and the basis functions trying to do a great job on fitting and resulting in bobbing and weaving, and we've got the penalties saying, "Oh no, we don't want that; we want smoothness." How can we balance these two things? We have a balancing function here, which is known as the penalized likelihood function (see Currie, page 20).  $L(\theta)$  is the familiar log likelihood function, and that's the thing that's trying to make the fitted function follow the data.

Then we have the penalty function, the quadratic form and the coefficients. It's trying to make things smooth. It's like a balance. We have fit from the likelihood function and smoothness from the penalty function. The amount of belief you give to one or to the other is measured by the quantity  $\lambda$ , which is the smoothing parameter. If you put  $\lambda$  equal to zero, we just believe the data and get an unsmooth function. If you make  $\lambda$  very large, it means that I want to have a very smooth function. The question has been reduced to choosing this value of  $\lambda$ . The normal equations are applied, but for the Poisson distribution. The Newton-Raphson

algorithm is a famous algorithm that we can use to solve the likelihood equations for the Poisson distribution. The main thing is that it's least squares. A nice mathematical point is that because the penalty has a quadratic form, it slips beautifully into this standard algorithm, and we stay within these squares.

That's the technical side of things. How does it work? This  $\lambda$  has to be chosen by some selection criteria. There's Akaike information criterion (AIC), Bayesian information criterion (BIC), generalized cross-validation (GCV) or a host of other things, as well, that will try and make a proper balance between following the data and following the likelihood and giving us a smooth function. How does it work? That looks a lot more comfortable (see Currie, page 21). Here, I think I've used BIC. This is age 65, and the penalty has smoothed out the function in what seems to be a reasonable, acceptable fashion. We're not fitting a cubic or a quadratic; we're trying to follow the data but follow them in such a way that the smoothness that we get reflects the variability in the data.

You can do forecasting, as well. The way the forecasting works is to take the penalty function and say that you don't want to stop the coefficients anymore. In this case, if we're using a quadratic penalty function, that leads us to a linear extrapolation of the coefficients and a linear forecast. That's the basic method of forecasting with penalized splines. We can still compute confidence intervals. This is striking (see Currie, page 23). When we did the quadratic Gompertz, we had tight confidence intervals. This is a bit of a slap in the face, but it does look a little more realistic. It says that the farther out you go, the less confident you are; the funnel of doubt is widening out. I'm comfortable with this kind of picture, which I think does reflect more truly the kind of uncertainty that we'd expect to have so far into the future.

That's the end of the first part of the talk. That's Gompertz with a modern twist. All we've done so far is look at a single age of data or a single year of data and, instead of using Gompertz' linear function, which works over part of the range of age, we've used a general smooth function that works over the whole range of age or time.

Let's move up to a mortality table. I'm now going to describe the Lee-Carter method, which, as I mentioned at the start, has already been described in your own journal about four years ago in Lee's paper. That's what it looks like (see Currie, page 24). A few symbols are there, so I'll try to take this a little more slowly. It's a model for a table, so the variable is  $d_{x,t}$ . It's indexed by age and by time, so  $x$  gives you the row, the age, and  $t$  gives you the column, the time. It's still a Poisson model; that's the same. It's a bilinear function. The  $\alpha_x$  is a measure of the overall force of mortality by age. The  $\kappa_t$  is an overall measure of the force, the time-dependent element, and the  $\beta_x$  is a moderating function that moderates the time-dependent element by age. The way to think about it is that for a fixed time it's a linear function, and for a fixed age it's a linear function. I'll show some graphs in a

moment that will make this clearer.

One point of note is that we've got a lot of parameters here. Each age has got 90 parameters times two, and each time has got 53 parameters. There's a bit of redundancy in here, which takes off a couple, but we've got about 230 parameters. That's a lot of parameters. How do you fit this model? We're mainly concerned with population data, not insurance data. I think that's an important point. They just use the least squares method on the log scale, so that's one way of doing it. Then a nice paper came out by some Belgium people in 2002 that made the obvious observation that because this is a Poisson model, you can write down the likelihood and therefore estimate the parameters using maximum likelihood. They did that, and that works well.

I'm a statistician, so I have my own take on this. I like the idea of conditional generalized linear models. I think this makes the Lee-Carter process clear. Let's suppose that we knew the kappas. If we knew the kappas, we have a generalized linear model in age. The kappas now have got little hats on them to indicate that we know what the value is (see Currie, page 25). That's just a linear function now, so everything is simple, and I can fit that. I can work it the other way around, as well. If I know the alphas and the betas, I have a linear function in kappa, so I can estimate that. I can just jump back and forth between the two, and that will converge quickly. That works well.

What does it mean? This explains what the Lee-Carter method does all together (see Currie, page 26). We've seen the alphas already; that's just the force of mortality by age. There we've got our Gompertz function going from about 40 up to 80 or 90, which is familiar. Notice what I've plotted here. These are discrete. These alphas are discrete, so there's one alpha for each age, but I've joined them up. Look at the graph of the kappas; this is the time-dependent element. You people are interested in this. The kappa is the key to this. The kappa is coming thundering down, and that means that the force of mortality is improving rapidly. The improvement varies with age.

Look at the moderating function, the beta that moderates the kappa by age. On the bottom right is the cross section of the surface for age 65. The key observation for me is that if you look at the bottom left and the bottom right, it's the same picture. Of course, that happens because the bottom right has got  $\alpha_{65}$  plus  $\beta_{65}$  times  $\kappa$ , so all you're doing is scaling the kappa and shifting it. The way to think about the Lee-Carter model is a heavy assumption. It takes the kappa function and scales and shifts it for every age. That's made clearer where you get ages 50, 60, 70 and 80. If you took the scales away, you couldn't tell them apart. But the scales are different; they're located differently and scaled differently, so the slopes are different. They only look the same because of the way they're plotted. That's a heavy assumption. What you get is nice.

There's a great thing about Lee-Carter. It works so well because you've got a strong signal in time. The argument here is a subtle one and a strong one. We're not too worried about how good a model this is of mortality. What we want to get is a strong signal in time so that we can get a clear projection. That's the thinking. Here we are forecasting the kappas (see Currie, page 29). The usual way of doing this is with time series. That's how Lee-Carter works. There are the projections in 10-year age bands. You can see that they're not parallel because the moderations are different, and they're at different heights because the alphas are different.

Can we now put these two ideas together? We've got Gompertz in 1825 and Lee-Carter in 2000, and we put these two things together. We can do this rather nicely. Here is the Lee-Carter model, and all I'm going to do is say the  $\alpha_x$  is our description of the overall mortality, so I can replace that with any function I like. In particular, I can replace it with a Gompertz function. This is what I've called the Gompertz-Lee-Carter model (see Currie, page 30). I've replaced  $\alpha_x$  with  $a_1 + a_2x$ , a linear function. I've imposed a structure on the overall mortality, a smooth structure. You can see what's coming, of course. We're going to come to general smooth structures, but at the moment this is just Gompertz. This is Lee-Carter with an overall Gompertz smooth function in age. I've printed both fits (see Currie, page 31). The red fit is the original discrete Lee-Carter, and the black fit is the Gompertz-Lee-Carter, where the overall age has been replaced by a linear function. This has a dramatic effect for the moderating function in particular. It has changed a lot. It has little effect on the kappa function and not much fit on the cross section at age 65.

Now we've done all the preparatory work. We know how to fit a smooth function for any given age using our B-splines, so instead of replacing  $\alpha_x$  with a linear function, we'll just replace it with a smooth function. Why stop there? Let's replace the betas and the kappas; let's make them all smooth. I think this is a nice idea. As a statistician, one is trying to simplify things. The Lee-Carter model is a complicated model because of its large number of parameters. Let's try and cut that down. Maybe we can get a stronger signal and a more stable model. It looks less, but it's simpler. I've replaced the alphas with a smooth function, the betas with a smooth function and the kappas with a smooth function (see Currie, page 32). I'm going to use this jump-back-and-forward method, which is simple conceptually and fast to fit. Of course, I use penalties. The penalties are imposed on the alphas, or the a's, the b's and the  $\kappa$ 's. That ensures two things; it ensures that the function is smooth, and will also allow us to make projections.

Here they are (see Currie, page 35). There's a lot happening. The red line is the original fit for the discrete Lee-Carter. The blue line is the smooth one. I've even put in the regression coefficients. They follow the functions that are used in the model. You can see that they're nice fits there.

The reason why things look smooth on the alpha (see Currie, page 34, top left) is that we have an enormous scale. This is going on the log scale from -8 to -1. It's a

colossal scale, so things do look smooth. There's not much change there. The big change is on the betas. Their dashing about early on has been smoothed out. For the kappas, instead of having a function that's bobbing around to some extent, we've got a completely smooth function. You might be a little more comfortable with projecting using a smooth function. I'm not claiming that this is necessarily a better method here, but there are some advantages to it.

We're not going to use a time series method to project; we're just going to use a simple penalty method. You saw the comparison of the two fits. The blue fit is a smooth fitted function, as opposed to the original Lee-Carter method, which gives you a slightly raggedy function. In projections, the alpha term essentially remains the same, and the beta remains the same. The kappa is the same, except it has been continued using the penalties. The projections are not the same. Should we worry? I don't think so. These are two different models. There are big differences.

These are not the only two ways that you can do forecasting. People often say to me that they're looking forward to hearing about this great method of forecasting the future mortality. They seem to think that somehow I know how to do this. I cannot tell you that I know how to forecast mortality. I can tell you that I don't know how to do it. That's an important message. That's a serious message. You're going to have to think about your pricing in such a way that you don't assume that you know what is going to happen in 50 years' time. That might have some serious implications for your product design, for example, but that's something you know about and I can only comment on from the wings.

**FROM THE FLOOR:** What about confidence intervals for those?

**DR. CURRIE:** I can compute those, but I haven't shown them here. Lee-Carter confidence intervals are usually computed on the time series element.

We are running a little short of time, so I'm going to skip through the next part; it's a technical thing. I want to say a little bit about my own work that I've been doing over the past few years with Paul Eilers and Maria Durban. You might have gotten the impression that I'm not completely sold on the Lee-Carter method of modeling. I said that four years, four cross sections for four different ages (ages 50, 60, 70 and 80), all had the same form. Somehow or other that doesn't sound right. We would expect the pattern in mortality to be varying for different ages. That would seem to be almost self-evident, and yet the Lee-Carter model doesn't allow that. It may give good forecasts, so that's a different issue, but in terms of a model, it's not satisfactory. What I'm going to try and explain briefly is how you might be able to get a completely flexible approach to modeling, a genuine two-dimensional method.

I'm going to ask you to switch on the imagination here. We have one B-spline basis in age ( $B_a$ ,  $90 \times 13$ ), and we have another B-spline basis in years ( $B_y$ ,  $53 \times 10$ ). I'm just going to multiply the two together. Every age is going to be multiplied by every

year. So we've got all these hills sitting along here, and all these hills sitting along here, and I'm going to multiply the two together. This is simple arithmetic. What do you get? That's what you get (see Currie, page 37). That's the basis that I'm going to use in two dimensions. It's an exact analog of what you would get in one dimension. It's easy to construct. Every slice in age is multiplied by the basis on the other side, and you get that picture. That's a small basis. It's a 3 x 3 basis. We're going to be using a big basis here. You might have 25 basis functions in age and 25 basis functions in year. This will give us 500 basis functions. This is regression with 500 variables.

With each one of these little hills, we're going to associate a regression coefficient. That's going to enable us to follow the bobs and weaves of the mortality surface. I'm going to use penalties to make sure the thing doesn't go crazy and that we don't over-fit. We hope to get a general fitted surface that will follow faithfully, taking due account of the variation in the data, the mortality surface. There's a nice function that does all this. The thing that's done in the imagination is the Kronecker product of these two basis functions. That's the technical term for that. For penalties, we're going to have a penalty that's going to be associated with each regression coefficient. The coefficients live in the age-year plane. I'm going to have penalties on the age direction and I'm going to have penalties in the year direction. That should give us a smooth function (see Currie, page 39). There's what you get (see Currie, page 40). There's the fitted smooth function. It doesn't tell you much. The scale is large and cross sections are a bit better.

What am I doing next? This is the acid test. This is a calibration exercise. I've used the first 25 years or so of data, and I'm going to forecast the next 25 years of data. This is not forecasting the future; this is forecasting the past. We know the answer, and we can stand up and be counted here. The forecast for age 35 doesn't do too bad a job (see Currie, page 41), but the important point is that the data lie comfortably within the forecast funnel. We're not going to get it right, but at least we're well inside our confidence interval.

Here's age 65 (see Currie, page 42). This is what all the trouble is about. The mortality at ages 65, 70 and 75 has been increasing incredibly quickly, particularly over the last 10 years. We're not sure why that happens, but it's a fact. The forecast gets this wrong. My attitude would be that if we're sitting in 1975 and thinking about where mortality is going to be in 25 years time, you'd be doing well to make a correct guess. We didn't get things absolutely right, but again, we're comfortably within the confidence interval. If you'd done some kind of discounting for the risk that's involved in forecasting, you would have protected yourself against the particularly rapid improvements in mortality that occurred.

Looking at projections up to 2050 (see Currie, page 43), I'm sure some of you are wondering what has happened at ages 30 and 40. That's a striking thing. We've got this crossing. You may or may not be comfortable with this. What has happened to

age 30 over the past 20 years or so is that there has been a dramatic external impact from AIDS on the force of mortality at age 30. The force of mortality has become becalmed around age 30, but is continuing to fall around age 40.

If you didn't like the resulting overlap there would be ways around this. You could use confidence intervals to shift things down if you wanted, or you could try fiddling with the penalty. But as far as the data are concerned, there's a strong suggestion that we may get crossing over in the future. Again, I would say that what's more likely is that you want to have a product that's not going to be so dependent on forecasting 50 years into the future.

I have looked at the patterns by gender. They are different. This is a male pattern.

Here are the ages in which you're particularly interested, ages 60 to 70 (see Currie, page 44). You can see that we haven't got power projections here. Here are four ages, adjacent ages, from 70 to 73 (see Currie, page 45). These are slightly different. They're not scaled and shifted; they're just slightly different, coming from the two-dimensional surface that we're fitting here. Projection at age 35 to 2050—we're not sure what's happening (see Currie, age 46). Projection at age 65 to 2050 is also shown (see Currie, page 47).

This is a comparison of all three methods that I've shown you (see Currie, page 48). We've got the discrete Lee-Carter in red. We have the smooth Lee-Carter in blue, and we have the two-dimensional method in green. One can make some observations about things that are happening in that graph, but I don't know that any of them are profound. I think it so happens that for most of the ages, the two smooth methods give the same answer. I think that's a coincidence. The other fairly obvious point is that the discrete Lee-Carter is forecasting much faster improvements than either of the smooth methods. I think that's probably a general property of forecasting with a model with a lot of parameters in it. My methods are taking the number of parameters back, hugely. Although I have a large number of fitted parameters, somewhere on the order of 500, by the time we've put the smoothness on, we come back to about 70. The effective number of parameters is 70, so I'm using far fewer parameters than the Lee-Carter method, and I think this is moderating the forecast. I think that's why that is happening. There might be a tendency for the Lee-Carter method to give more extreme forecasts simply because of its large number of parameters. That's a kind of speculative remark.

I hope you're interested in what I've been saying. You might even be sufficiently interested to look at some of the papers. There are references in the handout that you can download from the Web site, or from my Web site, on a number of different aspects of mortality modeling.

**DR. HARDY:** Because the two talks are different, we'll take questions on Iain's talk while that's fresh in your mind.

**MR. SAM GUTTERMAN:** The basic method assumes that the same forces of change are going to be in effect in the future that were effective in the experience period. That's a rather naïve type of forecasting method. Forces in the future are likely to be different by age and by gender. How can you utilize this method if you assume that the past is not necessarily indicative of the future?

**MR. CURRIE:** I don't think you can. That's an easy question to answer. This method is absolutely dependent on saying that our best guide to the future is what has happened in the past. I understand what you're saying. You're getting at things like future medical advances. For example, what effect are statins going to have on the force of mortality? That might improve things out of all imagination. On the other hand, there are some other problems that might be building up. The obesity problem is a big issue in the United Kingdom, and I believe it's also an issue in the United States. That could drive things in the opposite direction. How can you balance these two things? I don't know the answer to that question. They might find a cure for AIDS. There are lots of things that you might want to try and build into a model of mortality. I think it's difficult to do that. I've hung my hat on the pole that says that the past is the best indicator of what we've got of the future. If we knew what effect statins were going to have and if we knew what effect obesity was going to have on the force of mortality, we could build it in. But I don't believe that we know that. It's a big issue. How do you get in these particular forces into any model? It's not an easy problem.

**DR. SAMUEL H. COX:** I want to thank Iain for bringing his work here and presenting it. It's interesting. As Mary said, what Yijia Lin and I are working on is related in that the motivation for a lot of the risk management techniques we're going to talk about comes from the problem of projecting mortality. Maybe we'll have some things to say about that toward the end. The work that we're presenting today is going to appear in *The Journal of Risk and Insurance* some time, maybe early next year.

We started this a couple of years ago. Yijia is a Ph.D. student at Georgia State. A couple of years ago, she came to me and said that she wanted to do her dissertation research in life insurance. She's not an actuarial student, so I told her that she didn't know enough mathematics and she would have to take a life contingencies class with our actuarial students, so she did that. Each year we give a prize for the best student in actuarial mathematics at the end of the year, and she won that prize. She picked it up quickly. Then last spring she wrote the first draft of this paper, which we've distributed to a number of people, and we've presented it to seminars at universities like Waterloo. We've gotten a lot of valuable comments and ideas, and we're grateful for them.

What I'm going to talk about is generally about mortality-linked securities and mortality deals, in particular one deal we know about. I'm also going to talk about an idea that we got from Shaun Wang about how you might price mortality

securities.

I'll give a general idea of what we're calling "mortality-based" securities and, in particular, the Swiss Re deal that we learned about in December 2003. Then there may be a discussion of the other aspect of mortality risk, longevity risk, and describe how we think that might be the basis for a mortality bond. This idea of securitizing mortality risk has been around a long time. The early securitizations of catastrophic property risks were in the early 1990s. In the mid-1990s there was some talk about mortality deals that would be similar to that, but the interest in it seemed to drop until a couple of years ago. We started hearing rumors of these deals, and finally one was announced by Swiss Re.

There's a paper I wanted to mention that surveys a lot of life- and annuity-related securitizations. It's a recent paper by David Cummins. It's on his Web site at Wharton. He discusses five different classes of securitizations that are related to life insurance and annuities. All of them have mortality risk to some degree or another; some of them also have also lapse risks. What we're looking at is a pure mortality transaction, not some of the other aspects. What we're thinking of is a way to hedge mortality risk and isolate it from other risks that are included in these other kinds of securitizations. You can think of most of the other securitizations as being more similar to asset securitizations in the sense that an asset securitization, like a securitization of a bond portfolio or a loan portfolio, converting it and using this cash flow to support a bond is similar to these securitizations in that the part of the cash flow from a book of life insurance or annuity policies was dedicated to fund a bond, and that was sold to investors. That's similar to asset securitizations. The catastrophic risk securitizations and what we're thinking of as mortality securitizations are different in that they're based on the pure risk rather than forecasting cash flows. It's a little different from the other things he has in mind.

This is the structure that we're looking at, and it's the one that Swiss Re used (see Cox, page 12). It's the same basic structure that was used in the catastrophic risk securitizations. A reinsurer, or in some cases an insurer, will form a special-purpose company. One of the early diagrams I saw like this involved USAA's securitization of its property risks. It formed a special-purpose company in the Cayman Islands. You can think of the special-purpose company as being a reinsurer itself but formed just for the purpose of issuing these bonds and providing the insurance to the insurer or reinsurer.

Initially, the reinsurer pays the premium. The investors are going to buy bonds, so there will be a contribution from them. The total is used to buy collateral bonds, default-free bonds. Later, the income from the bonds, either in coupons or in redemptions, is used to fund obligations to the reinsurer and investors. In some years there will be benefits paid to the reinsurer, and in some years there will be benefits paid to the investors, in the forms of, say, bond coupons or maturity values. That works without any default risk. The program is set up so that the cash

flows from the bond portfolio always exactly fund the benefits to the reinsurer and the benefits to the investors.

The way this structure works is that there's no default risk. The only risk is the risk that's written into the bond indentured in the case of catastrophic property risk. At USAA, for example, it was its own portfolio of losses in the Gulf coast and Atlantic coast on its own portfolio, so some measure of loss in that portfolio would be put into an algorithm, a function that was defined in the bond indenture. That would trigger a payment of B to USAA from the reinsurer, and then a decrease in D, correspondingly. The investors would not get the full coupon or the full redemption value. We're thinking that the allocation of B and D would depend on perhaps a mortality index or something like that.

Let's look at the only actual deal I know about, although we've heard rumors of others. The Swiss Re deal was issued in December 2003. It's described in some of the literature that was published then as a four-year bond, but it matures in January 2007. I think the reason it's called a four-year bond is that there are four years of exposure to the investors for mortality risk. There are no coupons at risk. That is, the  $D_t$ , the coupon or the cash flows to the investors, is fixed and certain until the bond is redeemed, and then the investors may lose a portion or all of the redemption value. It was priced to sell at 1.35 percent over the London InterBank Offered Rate (LIBOR). The LIBOR rate at the time was low, so maybe as a percentage of the LIBOR rate, the premium might not look so low.

The structure, that is, the rule that's written into the bond indenture that defines how the investors get their redemption value back (get the maturity value), is based on a population index. The index is a weighted average of population indices created by independent government agencies in the United States, United Kingdom, France, Italy and Switzerland. All of that is public information. The weights are in the bond indenture but were not published, so I don't know exactly what the weights on the country indices were. We did talk to some people at Swiss Re, and it was structured to reflect their own exposure in those countries. The triggering mechanism, then, is based on the maximum of those. If something goes wrong and the index goes wrong in the sense that mortality rates are higher than we would think, the bond investors are at risk for that event, and they'll lose a portion of their maturity value. The base level is the 2000. They calculate the 2002 index with those same weights, and then in 2003, 2004, 2005 and 2006, they recalculate the index. In any of those years, the Swiss Re could calculate the maturity value and reduce it if one of those values comes up high enough. If it stays below 1.3 times the base rate, that is, if it doesn't rise 30 percent above the base rate, there's no reduction in the redemption value; it pays the \$400 million back. If it goes over that, the investors start losing the redemption value until the index goes to 150 percent of the base value, and then they lose everything. That's the structure in it. It is a pure mortality deal.

We asked the company why it was doing this and what it had in mind. There were two main reasons. Swiss Re does a bit of securitization of insurance risks in the property area. It had in mind doing something like this for a while. Part of the motivation is to break the ice, to get investors may be comfortable with the so-called catastrophe bonds, property risk, to get used to the idea of accepting some mortality risk. That was one reason. The other reason is that this is a hedge against an epidemic in the next three years. A shift at this range is an extreme shift, but it could happen with an epidemic, something like a 1918 flu epidemic.

There's another approach other than using bonds, and we've heard that this is being considered. A swap approach would not have to be public. This could be a private transaction or one arranged through a dealer, broker or reinsurer. In that case, you replace the collateral by annualizing the premium from the reinsurer and the purchase payments from the investors. Think of it that way, but each, the reinsurer and the investors, pay fixed and get floating benefits. At each payment date, the fixed payments add up to the floating payments. The fixed payments are given to the broker or dealer. They apply the terms of the contract and decide how that's allocated as a benefit to the reinsurer to cover a mortality event or as a coupon to the investors as a return on their investment. This might be convenient in some cases. The disadvantage is that there's a default risk because the payments aren't collateralized. You're taking a risk that the other party might default.

Here's the other side of the risks. Instead of focusing on a sharp increase in the mortality rates, how might you securitize risk of missing a projection of what Iain was talking about, the improvement in mortality? We're thinking of maybe a long-term bond that has a way of allocating the risk of missing the projection to investors. The idea is to set a strike level for each future year. What we did was project into the future what the improvement in the force of mortality has been and get a projection (see Cox, page 21). Maybe you could do something more sophisticated along the lines Iain suggested. Set a projected value and then use that as the index,, like the index in the Swiss Re deal. In the case that the projection is conservative enough and that in the given pool that you're looking at, enough of them die so that the survivors to year  $t$  stays below the projection, there's no benefit to the reinsurer or to the originator. If you miss the projection and rise above that strike level, you start paying benefits at that point. The investors will start losing investment income until you reach a maximum. At that point the coverage is fixed; you get the maximum benefit provided by the bond. It's like a call spread, but one in each year.

This is the net benefit to the insurer (see Cox, page 22). The number of effective annuitants—you can cast it that way—is what they projected if it's below the strike level, and then it stays at the projected value until you rise up to the second strike point, when you're at risk again. It could increase after that.

What the investors get is the other side of it (see Cox, page 24). They get their full

coupon as long as the number of survivors stays below the first strike level, and then they start losing coupon as the forecast turns out to be not conservative enough. Eventually, if it is bad and rises above the second strike level, they don't get any coupon back. I say coupon because that's the way we've set it up, but you could also make the redemption value at risk also. In each year we'll have it split, so that the coupon amount exactly pays either investment return to the investors or benefits to the insurer. To do that, you've got to be able to convince the people involved of what the prices are so that you get enough premium from the insurance company and purchase price from the investors so that you can buy the collateralizing bond. What you need to buy is on the right side of equation (2), so you need to get that much (see Cox, page 28). The  $d$ 's there are just risk-free discount factors in the market at the time that you buy the bond. You're going to buy a collateralized bond that is a risk-free bond.

For the pricing, we used Shaun Wang's idea of effectively transferring the price of risk in one market to another. This is what he calls a distortion operator. It's based on the cumulative density function (cdf). This is how it works (see Cox, page 32). You transfer from one market to another, from one market where we can observe prices—that's, say, the immediate annuity market—into a bond market where we don't know the prices. Our idea is to make the pricing analogous. In the one market we'll think that the cdf is the base, or the physical, actual mortality that you might get. We're going to use the 1996 Individual Annuity Mortality Table for the original cumulative density function. Then there's a market price of risk,  $\lambda$ , that shifts that, and you bring it back to a new distribution,  $F^*(x)$ , that incorporates the market price of risk. Our cdf is  ${}_tq_{65}$ , doing it for age 65.

We have a sample of market prices. The prices include expenses, so we use an expense assumption of 6 percent. In the paper there's some investigation of that, but if you were actually doing this maybe you'd have more access to the appropriate expense assumption than we did, so you might make a better application of this technique. What you do then is take the market prices and set them equal to the model prices, using  $\lambda$  as a parameter (see Cox, page 37). On the right side, there's an error; the 7.48 is monthly, so it should be multiplied by 12. The left side is market price. The right side is model, and we tune the model until we make them equal. That will give you a numerical value of  $\lambda$  so that your model reflects the market prices. That's all in the annuity market, so in effect what we're doing is calculating what insurance companies demand for a market price of risk in their individual annuity market.

The base (or physical or actual) distribution without regard to risk is the 1996 Individual Annuity Mortality Table. After you calculate the market price of risk and then calculate the other distribution, it's shifted up and to the right, which is conservative, meaning that the people selling the annuities are anticipating people living longer than projected in the 1996 table (see Cox, page 39). That's all in the annuity market.

Now what we're going to do is take that distribution and apply it to price the bond. We set our strike levels by projecting improvement. We wind up calculating the price of the bond and the price of reinsurance premium (see Cox, page 41). They're relatively low. I think we ought to consider revising the paper. The reason they're low is that the spread between the strike prices is narrow. I didn't realize that until after the paper was accepted, so I haven't done anything about it yet. The second strike levels,  $X + C$ , needs to be adjusted because this is a light coverage the way we have it now.

That's some indication of how things move. These are not projections; these are "what if" things. What if the mortality curve is not as you projected? We just used our own ideas of what might happen and how things might shift. You can do that and illustrate how investors' or reinsurers' positions would change.

We're getting close to the end here. Our idea is that this could be used to increase capacity to take on longevity risk in the same way that Swiss Re used its deal in December to increase its capacity to take on life insurance risks. The security markets can provide increased capacity. They did it in the property risk area. They could use mortality bonds. I also showed how swaps might be used. We illustrated how Wang's transform can be used in pricing. There are other methods, and we're looking at some other methods for essentially modeling the market price of risk. There are other approaches. These are some of the papers we used in getting to this point (see Cox, page 48). I mentioned Cummins' paper as a good overview of securitization in general. The papers by Shaun Wang describe his method of transforming distributions and applying them in a number of different ways.

**MR. LAWRENCE S. CARSON:** Did the Swiss Re mortality bond have any exclusions for terrorism, nuclear, biological or chemical?

**MR. COX:** No, there are no exclusions. Just calculate the public indices, public available information. If they rise for any reason, the bond indenture is applied. It doesn't matter why they rise. It could be a terrorism event; it wouldn't matter. There's also a feature in the bond that allows it to delay the redemption. Let's say that in 2006 there's an epidemic, and the bond matures on January 1, 2007. The public information won't be available then; it will be some time later. It has an option to delay payment of the redemption value in that case. It can wait until the data are available.

**FROM THE FLOOR:** That was similar to a question that I had. I know that the population data get revised historically, so the question is, how do they handle that on a contract?

**MR. COX:** These are such extreme levels that everyone would know if there's going to be something worth delaying repayment of the amount at risk. The one that the

interest didn't delay in was Swiss Re. It has to continue making the interest payment if it continues it, but it can delay it up to three years.

**DR. HARRY PANJER:** Iain, you used the B-spline approach for both age and time. It seemed to me that the functions originally were relatively smooth. You could have used smoothing splines or something like that, just cubic splines, for each of them and gone through the same kind of approach. I didn't have a clear insight into the number of parameters and dimensionality of the problem that was introduced by using B-splines. I wonder if some other simpler methods might produce the dimensionality. I suspect that the answer is no because you're assuming the B-splines and those are fixed quantities and there's only one parameter per B-spline.

**DR. CURRIE:** You always have the problem of determining the dimensionality of your problem. The way that we approach it is to over-parameterize to a huge degree and then use the penalties to scale it down appropriately. But you're absolutely right, and I did indicate in the talk that there is another approach. You can try and fiddle around with the level of parameterization. That was a popular approach in the 1970s and 1980s, which essentially involves choosing the knot positions. You can certainly go that way if you're more comfortable with that approach. I favor the way that we approach it because you remain within a regression framework, and mathematically it is an attractive scheme, whereas knot selection is not easy mathematically. It's a difficult optimization problem.

**MR. ANDRE CHOQUET:** My question is for Sam. I was wondering if you had experience with the application of mortality bonds for pension plans in the sense of hedging mortality risk, or if you had an opinion on how it could be used for that?

**DR. COX:** The example I gave was insurer/reinsurer, but certainly a pension plan, say a large corporate plan, could do the same thing. A lot of firms have sophisticated hedging activities in other areas of risk, so it seemed like a natural thing that a private corporate pension plan with long obligations could do the same sort of thing. However, I haven't seen it. The only other things we've heard about are swaps within, say, a family of companies, where the swap is used to shift mortality risk from one insurer to another, but it's not taken to the public.