Session 3B: Mortality Modeling II – Other Methods

Discussant: Thomas Edwalds

Presented at the Living to 100 Symposium
Orlando, Fla.

January 5-7, 2011

I do think these are two very good papers, and it's my pleasure to have a chance to discuss both of them. The common theme of these papers is projecting mortality and estimating confidence intervals around mortality projections.

Projecting mortality at older ages is a very difficult task. First of all, it's difficult to measure the current mortality at old ages because data is often sparse or even outright lacking and the data that exists is often of poor quality, so even knowing what the current mortality rates are can be very difficult at the older ages. It's even more difficult to estimate the trends in older age mortality because you're not certain about what it is currently or what it was in the past. Furthermore, if you are using population life tables or industry mortality tables for your history of past mortality rates, the tabular rates are often formulaic extrapolations of younger age data. They may not be based on actual mortality experience at the older ages at all. So projecting future rates of mortality at older ages is a difficult task to begin with, and both of these papers have taken on the challenge, not only of estimating the projected mortality rates, but of estimating confidence intervals around those projected rates. This is quite an ambitious task in both parts, and they've done a good job of it.

Both papers generally agree that there are seven easy steps to getting confidence intervals around your future projections of mortality rates. You've got to find a data source, choose a data series to model, and choose a model form. Then you've got to estimate your parameters and your error term, and from that model you can stochastically simulate however many possible future series values you want. Depending on the type of values that you choose to model, you may need to calculate the q_x values from the values you have stochastically simulated. Then you rank them and get your desired confidence interval. So it sounds simple, but it's not an easy task.

The papers have taken a somewhat different approach in how they execute these seven steps. Both papers use the Human Mortality Database as their data source, but they take different slices of it. We heard comments in a previous session at this conference about how perhaps some of the higher age data in the Human Mortality Database might be overly smooth, but that's not really a problem for either of these papers.

The change in the discount sequence ratio is the quantity that is projected by Wang and Yue. This quantity is the main focus of the paper, and they take a lot of care to establish that it is, in fact, worthy of projection. There are several different versions of the discount sequence ratio they refer to. The one they ultimately project is the discount sequence ratio of the number of survivors from the life table, using quinquennial ages, i.e., $(l_x / l_{x+5}) / (l_{x+5} / l_{x+10})$

Li and Chan project the logit or log-odds of q_x, i.e. $\ln (q_x / (1-q_x))$. That's not the main focus of the paper, and their main point about simultaneous prediction intervals could have been made regardless of the underlying model used in the projection.

One of the things that's different about these papers is the type of model that is estimated and projected. Wang and Yue are using a Brownian motion stochastic differential equation. They show that the discount sequence ratio has a certain mean reversion characteristic and model the change in the discount sequence ratio as the sum of a mean reversion term and a diffusion term. The model is parsimonious, with only two parameters to be estimated from the data. They use this model to simulate trajectories of values of the discount sequence ratio and then estimate point-wise confidence intervals from these simulations. As Li and Chan point out, the use of point-wise confidence intervals for these simulated values is standard practice. They generate stochastic simulations for the same historical time period as the data from which they fit the model and then present the estimated confidence intervals around the past experience.

Li and Chan model mortality rates using the Cairns-Blake-Dowd method with cohort effect, which has three time-dependent coefficients, plus the cohort coefficient. The three time-dependent coefficients are the ones that actually get modeled as random walk with drift. They use this model to stochastically simulate future trajectories of mortality rates, and then present the results of the simulation as mortality fan charts. The fan charts help you visually see how the uncertainty of the mortality estimate increases over time as you project into the future.

Wang and Yue take a great deal of care with their model. I particularly like the fact that they demonstrated graphically in the paper the finding that there's actually white noise when you get down to this change in the discount sequence. Of course, what they demonstrate in the paper is the white noise of the discount sequence ratio of life expectancy, not of survivors, which is the one they model. Still, it makes a good visual impression in the paper to show that, in fact, the discount sequence ratio is a good candidate for a projection tool. They demonstrate historical values of a quantity, the discount sequence ratio, and this history appears to be white noise. So if we model this quantity, we should have a good estimate of the error term. Then if we project it assuming Brownian motion in the future, these projections will have a solid foundation.

Wang and Yue discuss the regularity condition at length, but it does not appear to be necessary for showing that the discount sequence ratio is a good projection tool. The regularity condition is actually relevant to a separate issue that was addressed in the paper about missing values. If some values are missing from your historical data, you can actually use the regularity property of the discount sequence and start filling in those missing values. This makes it very valuable. I did note that the discount sequences for life expectancy and for l_x values were the ones that displayed the regularity conditions. Some of the other ones that they showed, like the number of deaths, actually didn't seem to satisfy the regularity condition. The discount sequence for the number of deaths approached the regularity condition, but it didn't actually display the property historically. That ultimately is not important, since they chose to project the discount sequence ratio for l_x values.

For Li and Chan, the main point is not about the Cairns-Blake-Dowd model or the estimate of future mortality rates. The main point is about effectively communicating the uncertainty in future projections of mortality. I certainly like the fact that their method of simultaneous prediction intervals results in wider confidence intervals, because that just makes more sense to me. I get very nervous about narrow confidence intervals on future projections because I just don't believe them. I don't think we know that much about it.

The envelope concept, which is used in the Chebyshev bands method and describes how you would contain a subset of future samples, is very useful. One thing I particularly liked about the Chebyshev bands method is that it provides a metric on the entire projection, so that you can actually rank the projections as unified entities. The metric is the maximum difference from the mean in terms of number of standard deviations over the entire projection period, so the ranking is based on whichever point on that particular trajectory is farthest from the mean in a relative sense. The Chebyshev method uses this ranking to determine which simulated projections are in the confidence interval to be estimated, and which are not. Because of this, you can always have exactly the right number of simulations in your estimated confidence interval. For example, if you have 10,000 simulations you can pick exactly 9,500 of them to be in your 95 percent confidence interval. The adjusted intervals method gives you a smoother result than the Chebyshev method, but because at each step you're going to add one point on each side of the previously proposed boundary for the confidence interval, you may end up with more than 9,500 out of 10,000 simulations in your 95 percent confidence interval. You don't necessarily add one simulation at a time, because it's not a ranking that goes one by one.

In the introduction to the paper, Li and Chan mentioned that their method handles parameter risk. I didn't actually see that in the paper, and maybe it's not there. It would be a valuable addition, and I would encourage the authors to keep pursuing that.

One general point I would like to make is that any estimated confidence interval based on a stochastic simulation is completely determined by the model selected for the simulation. The simulations will estimate that confidence interval based on what the model is predicting, and the error term drives the entire dispersion of results. Li and Chan use a matrix C that's multiplied by a three-dimensional standard normal vector, and that's where all of the variation in the future projections is actually coming from. If you use a different estimate for matrix C, you get a different width of your confidence intervals. For Wang and Yue, the diffusion term of the Brownian motion drives how much variability there is. So when you are trying to estimate a confidence interval using stochastic simulation, you really have to understand how and why you got your error term, and what it means for your simulated projections.

Finally, I want to highlight the point Li and Chan make about the difference between the point-wise confidence interval versus the simultaneous confidence interval and raise the question about when do you actually need the simultaneous interval. As I said earlier, I like the simultaneous interval because it gives you a wider confidence interval, but it is important to consider what uncertainty you are trying to communicate. One of the questions that came up at an earlier session today was how many people in the United Kingdom are going to be at least age 100 in 50 years. That's actually a point-wise question, so you want a point-wise confidence interval around that particular answer. But for most of the types of longevity protection devices that are being contemplated, such as longevity swaps, you actually do need the simultaneous interval because you'll trigger a claim event if future mortality falls outside of the defined band at any point along the way, not just at a particular point in time.

I do want to just congratulate the authors of these papers. I think they’re excellent papers and a great contribution to this conference, so thank you.