Uncertainty of Longevity Projections

Gary G. Venter

Gary.Venter@gmail.com

Abstract

Modeling mortality rates can sometimes give good fits to available data, but a closer look at the best-fitting models reveals that they do not entirely make sense, and have a significant risk of misprojection. Models fit to US population mortality rates for ages 55 and up are reviewed in this light and some simple curve fitting to individual calendar year raw mortality rates illustrates some of the difficulties with the data that create problems for modeling. The trends over time are ambiguous and various assumptions for future trends can give quite different projections, especially by age. The situation changes for modeling for actual pension plans, where there may be less data, the population may be more homogeneous, and more information about the individuals may be available. However the projection issues remain.

Uncertainty of Longevity Projections

Projecting longevity starts with modeling mortality patterns and their changes over time. Hopefully some regularities are then identified that can form a basis for projection.

Example models here have been fit to US male and female population data from the Human Mortality Database. Ages 55 – 89 are fit for years of death 1971 – 2006. The starting point is the LC model from Lee and Carter (1992). That model fits log raw mortality rates, which would be usually expressed as the logs of the empirical q(x). Here the mortality ratio deaths/living for year t and age at death d is denoted as $m_{t,d}$. The LC model for the mean rate can be expressed as:

$$\log m_{t,d} = a_d + b_d h_t$$

Here a_d is the base mortality for age d. There is a different parameter for each age, so in the sample data there are 35 age parameters. Alternative models fit smoothed curves to the base mortality, with fewer parameters. Cubic splines, generalized forms of the Makeham curve, and quadratic fits to logistic transforms of the rates are examples.

The h_t parameter represents the level of the mortality function during year t. As mortality rates have been decreasing, the h's tend to get smaller over time. But mortality trends can vary by age, which the b_d parameters allow for. This is more extreme in the male data, where the trend at the oldest ages has been about half of what it is at the younger ages.

The LC model has fitting problems when the shape of the mortality curve changes in more complex ways than the model allows for. In the sample data this is more of a problem with the male rates than the female rates. A generalization, discussed for example in Renshaw and Haberman (2006), is provided by adding so-called cohort effects to the model. The cohort is the year of death minus the age at death, which is approximately the year of birth, depending on the times of year of birth and death. It is possible that some years of birth were faced with particular situations throughout life that gave them higher or lower mortality rates than people born in other years.

The LC plus cohorts (RH) model for the expected mortality rate is:

$$\log m_{t,d} = a_d + b_d h_t + c_d u_{t-d}$$

Here u_{t-d} is the factor for the t – d cohort, and the c factors allow the cohort effects to vary by age.

Fitting the parameters can be done via MLE by modeling the number of deaths in the t,d cell as a distribution with mean = people exposed (i.e., alive) times $m_{t,d}$, with some residual distribution, such as Poisson or negative binomial. Recent work has found that the Poisson is not dispersed enough for this. Delwarde et al. (2007) found that the negative binomial fit better for the LC model for male and female population data from a few European countries. Venter (2010) found the same for US male and female data for the RH model. For males this study found that the more skewed Poisson-inverse Gaussian distribution fit about as well as the negative binomial. The three-parameter Sichel distribution, which in the fit had intermediate skewness, fit better. For females the negative binomial had the best fit, with the other distributions apparently too skewed.

Venter (2010) also found that the model with cohort effects fit substantially better than the basic LC model, but that there are some problems with the cohort model. If the data is arranged in a rectangle, with rows representing the calendar years t and the columns representing the ages at death d, then the sample data fills out the rectangle. The cohorts fall on the NW – SE diagonals of the rectangle.

One problem is that a good number of these diagonals, especially in the SW and NE corners of the rectangle, have only a few observations. This allows for good fits in these regions no matter what the base mortality and trend parameters are, and so allows those parameters adapt themselves to fit the other parts of the rectangle. This lets the model fit much better to the changes in shape of the mortality curve, but that is not what the cohort parameters are supposed to represent. When they are fit to just a few observations they can indeed change the shape of the mortality curve, but their interpretation as deviations in mortality for the cohort throughout life is highly suspect. This is particularly problematic for the

SW corner of the rectangle, which is the youngest group in the study. The cohort parameters there would be used in future projections for mortality for those age groups, but if the parameters are not really cohort parameters, these projections could have substantial error.

The other problem is that once you add in cohort parameters, the cohort, base mortality, and calendar year parameters can become highly correlated. A trend in mortality across calendar years can impose a trend across cohorts and ages as well. In the Venter (2010) study this was the case for the female model. Using the information matrix from the MLE fit, these three types of parameters were found to be highly correlated, which made most of them not statistically significant. Even though the parameters taken as a whole improved the fit substantially, as measured by penalized likelihood, they were not individually meaningful or significant. They implied opposing trends in cohorts and calendar years, and a rather strange base mortality curve. These would be difficult to project in any useful way.

The RH base mortality parameters from that study are graphed in Figure 1 for the male and female models. The female parameters look rather strange, as discussed.

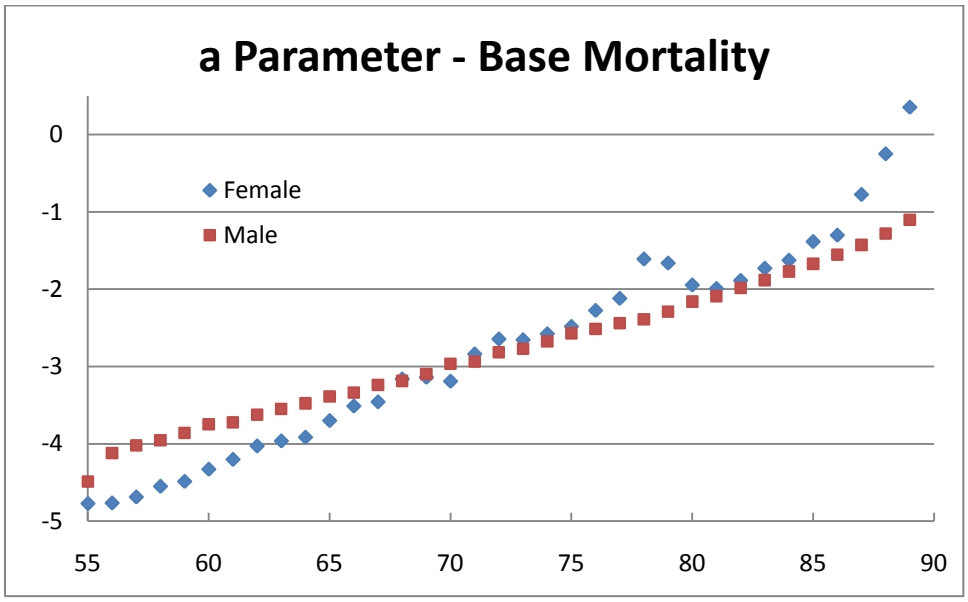

Figure 1 – RH Base Mortality

Figures 2 and 3 compare the actual and fitted log death rates for the male models for several years. However due to the common upward slopes it is difficult to compare the fits, so the

graphs have been rotated by subtracting a linear function of age that represents the slope.

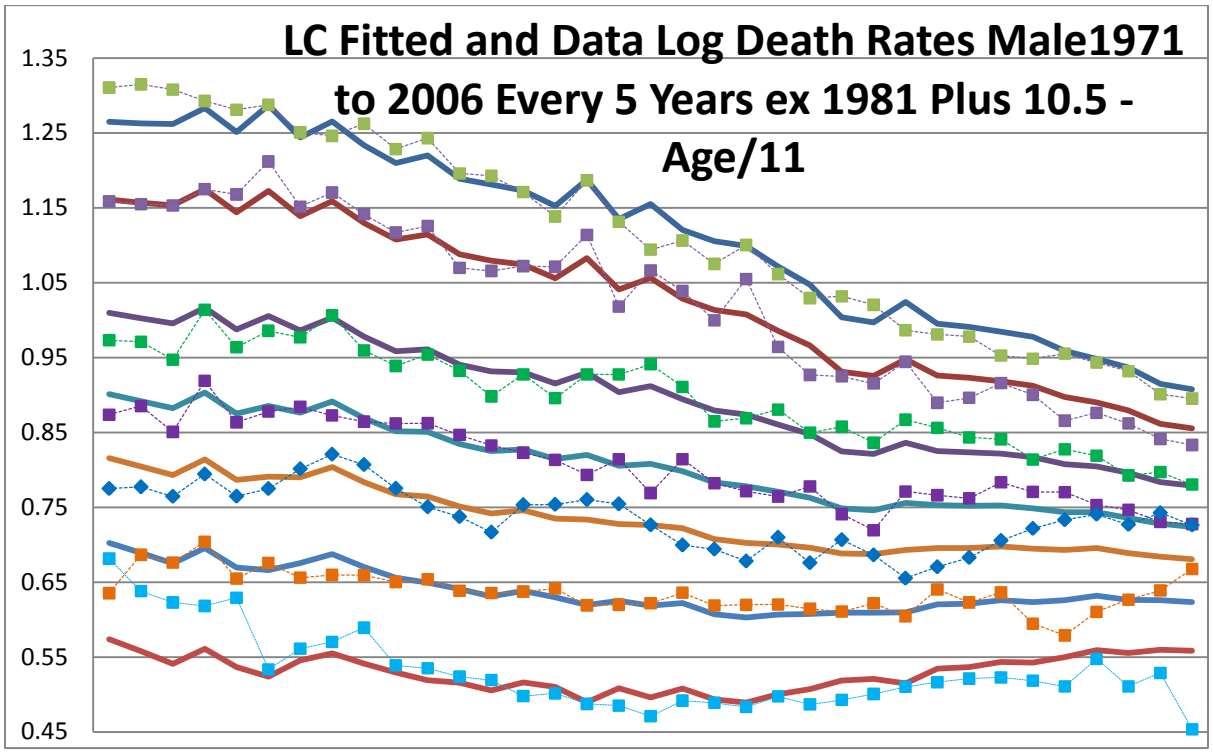

Figure 2 – LC Male

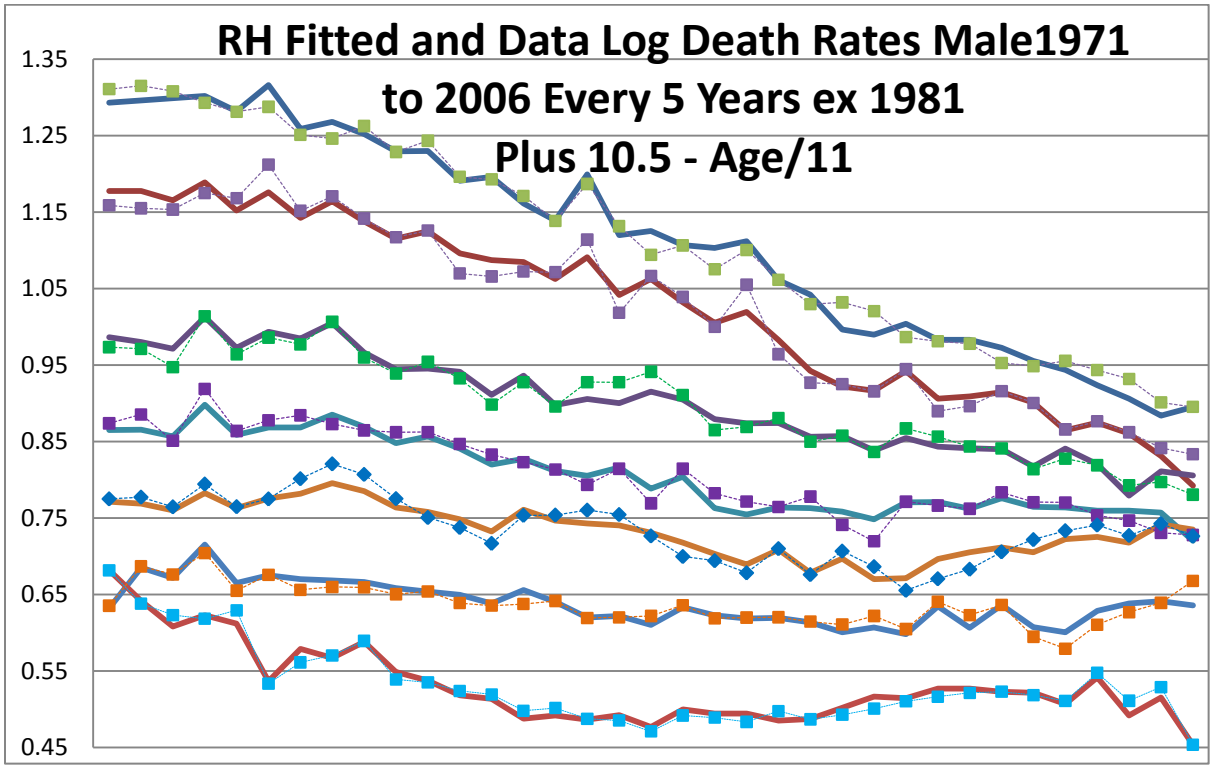

Figure 3 – RH Male

The solid lines are the models and the dotted lines the data. The LC model is seen to fit fairly poorly at younger ages in the earliest (top) and latest (bottom) years, as well as for older ages in some of the years. This is because the mortality curve is changing in ways that the LC model cannot handle. The RH model responds quite well to those areas, fitting virtually every twist of the data almost precisely. In fact it even matches the higher mortality for age 55 in 2006 than in 2001. As discussed, it is not plausible that these improvements come from actual cohort effects that hold over the whole lives of the cohort.

The greater improvement in mortality at the younger ages compared to the older ages is also apparent in these graphs. Even with the linear transforms, the distances between any two lines remains constant, as the same linear function is subtracted from both lines.

For the female models the improvement in fit by adding cohorts was comparable, but there was not the same age difference in mortality improvement, so the shapes of the mortality curves did not change as much as for the males, and the LC model fit better to start with. There was a similar lack of improvement at age 55 from 2001 to 2006, however. The female rate of improvement in mortality at all ages was closer to that of the older men, so was less at younger ages than for males. However it was also less steady, with fairly long periods of little or no improvement alternating with short periods of great improvement.

Curve Fitting

As noted, fitting curves to mortality rates has been explored as an alternative to LC models. Here a generalized Makeham function is fit to the raw death rates, although fitting to force of mortality is more typical. Richards (2008) discusses some such generalizations, based on earlier work by Beard (1959) and Perks (1932). Using a curve to fit the a_d parameters requires a log transform, and the form used here takes 4 parameters α, β, θ, γ:

$$a_d = \theta + \log[(1+\alpha\beta^d)/(1+\gamma\beta^d)]$$

Fitting such curves with four parameters to the log death rates in each year 1971 – 2006 results in the use of 144 parameters, compared with 104 for LC and 207 for RH.

Using a form of the negative binomial (see appendix), the following values of the negative loglikelihood (NLL) were produced:

	NLL		Parameters	HQIC Needed	NLL Improvement	
Model	**Female**	**Male**	**Added**	**Improvement**	**Female**	**Male**
LC	9444	9652				
Curve	9482	9553	40	79	-38	99
RH	8748	8972	63M, 103F	124M, 202F	696	581

The goodness of fit test here is the HQIC, which requires an improvement in NLL of log(log(sample size)) for each extra parameter, where here the sample size is 1260, requiring an improvement of 1.96555 per parameter. This is intermediate between the AIC and BIC. As can be seen, every model fit the female data better than the male data, and the RH model gave the best fit to both data sets, even though it is of dubious interpretation here. The generalized Makeham curve fit better than LC for the males, where the mortality curve was changing more over time, but LC fit better for females.

Nonetheless, for both males and females, the curves provide smoothed versions of the mortality functions for each year which are smooth enough to show all years on a chart, thus providing some insight into what the changes in the mortality functions have been.

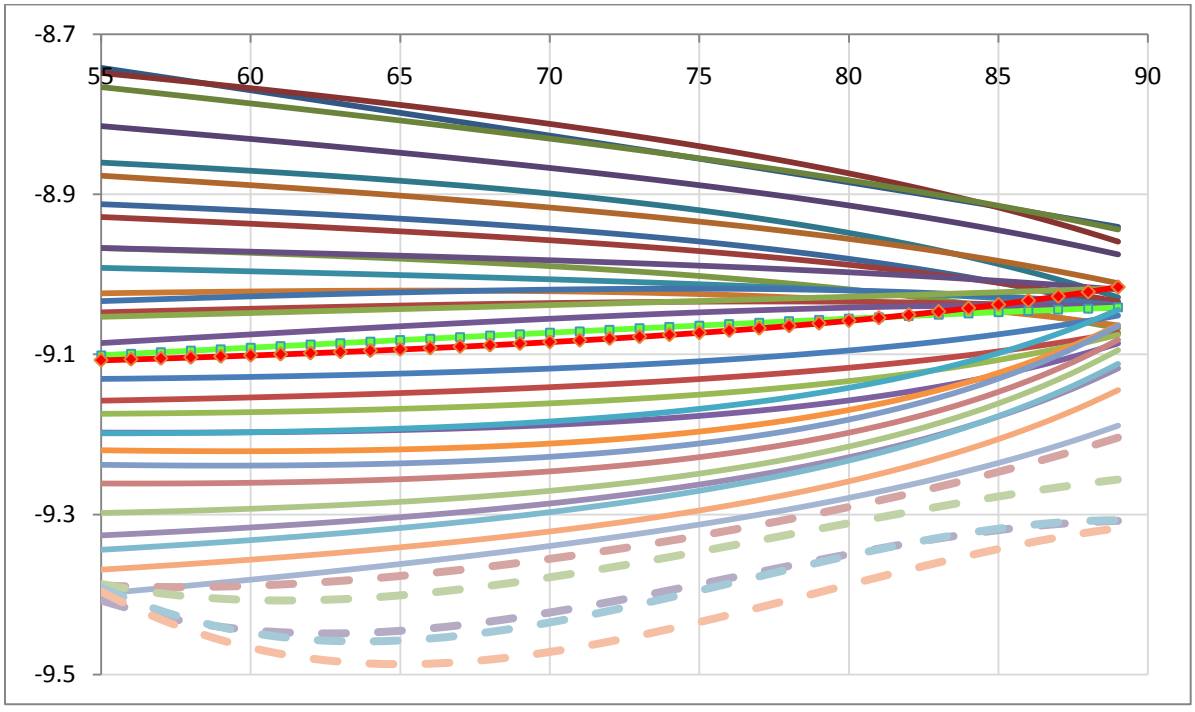

Figure 4 – Generalized Makeham Male

The male curves in Figure 4 (with age/12 subtracted) actually divide into three periods. First for 1971 until 1987, which is the light-green line with the square markers, the curves are straight or downward-curving. Then starting in 1988 (red with diamond markers) the curves bend upward. Until around 2001 or 2002 (first dotted curve) the mortality at age 55 is steadily improving, but the improvement at the other end of the curves is slower and sometimes non-existent. Then somewhere around 2000 to 2002 the improvement at age 55 stops and the improvement at the older ages accelerates. The last three years show a different shaped curve from the earlier years.

The changes in shape show why LC had problems fitting this data, but the fact that the biggest changes were at the ends of the lines shows why RH can make a big, albeit artificial, improvement in the fit. The graph also suggests that projecting future changes in longevity has a high degree of uncertainty involved. Should you just project the last five years, or from 1988 on, or average improvements in mortality over all the data? This could make quite a difference, especially at some ages. The recent lack of improvement at age 55 is particularly problematic. That could be related to the recent reduced access to health care in the US for people under 65. If so, you would expect it to eventually improve over time as the new insurance regulations take hold. At the other end of the curve, it might be reasonable to assume that the older ages will improve at the same rate as most of the curve, as there seems to be a trend in that direction over quite some time. Nonetheless this is an assumption imposed on the projection process and thus adds to the projection uncertainty.

The generalized Makeham model did not fit as well as LC for females, but the fits in Figure 5 still provide some insights. Here age/10 was subtracted to remove the upward trend. It is apparent that there has not been so much change in the shapes of the curves as in the male model. What does stand out, however, is variation in the rate of mortality improvements across the age groups. For instance, for ages 75 and above, there were fairly long periods with very little improvement in mortality, punctuated here and there with years of substantial improvement. Ages 65 and below, on the other hand, had much more

steady but generally small improvements. As with the male data, there has been little improvement at age 55 in the latest few periods. Also since about 2000 there has been somewhat similar year-to-year improvements in the male and female graphs, even by age.

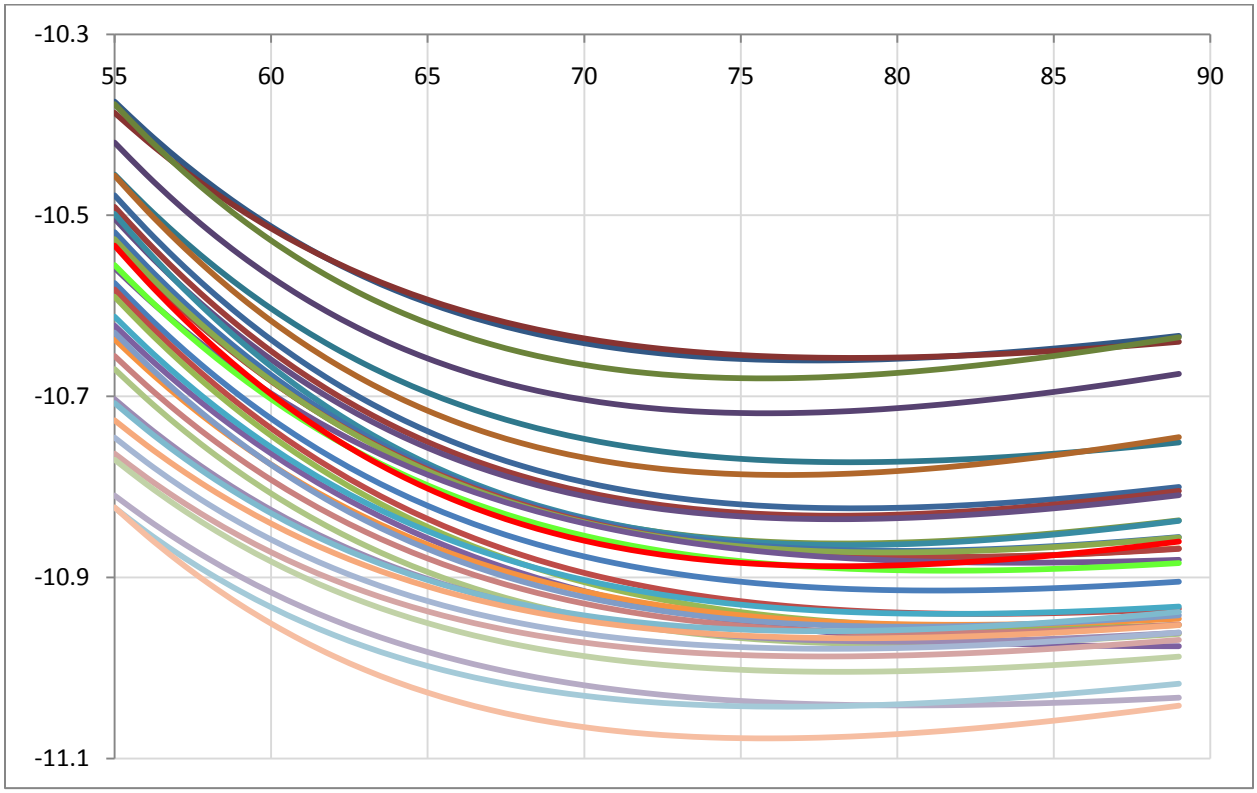

Figure 5 – Generalized Makeham Female

As with the male model, this graph brings out some problems in projecting future trends. Can you assume the greater improvement in the last 5 or 6 years will now continue? Would a time-series model with highly fluctuating rates of improvement be better at the older ages? Perhaps in both genders it would be appropriate to calculate trends under different assumptions then include all the scenarios, with selected weights, in the overall longevity improvement uncertainty model.

Pension Plan Application

Fitting models to pension plan data has both advantages and disadvantages compared to population data. There is typically much less data available, but the population is usually more homogeneous, and usually has access to health care. There is often more detail in the

data as well, such as actual dates of birth and death, which can allow fitting the force of mortality at each age instead of annual average rates as well as using practically continuous trend models. Also smoking status might be identified, which can give better predictions of mortality. The use of postal codes in creating geodemographic profiles is also possible. There are databases of demographic characteristics by postal code, and these have predictive power regarding mortality rates. Richards (2008) discusses some of these issues.

On the other hand, the time series might be shorter, and the population smaller, resulting in greater fluctuations in the data. This could require models simpler than LC for the base mortality, before the geodemographic effects are included. For instance a single generalized Makeham function could be fit with an assumed uniform time trend. Depending on the volume of data and the number of years available, the changing trends, variation of trend by age, cohort effects if any, and deviations of the data from the Makeham curve might not show up with any statistical significance. This could end up with a model that seems to fit well but with projections that understate the uncertainty in the trends.

Conclusion

Actual mortality data seems to follow patterns more complex than assumed by standard models, such as Lee-Carter, and cohort effects in these models might be misleading. Projections of longevity trends from any of the models available require additional assumptions about which parameter sets, etc. might apply in upcoming periods, and fairly different projections can result. All of these have some degree of validity and should be considered in the scenario set of possible outcomes. Additional detail is available in pension plan data, which can lead to more precision about individual annuitant probabilities, but limitations in the data can require cruder models of the overall mortality process, producing greater uncertainty in the projections, and this may not always be apparent from model diagnostics.

References

[1] Lee, R.D., and Carter, L.R. 1992 "Modeling and forecasting U.S. mortality," Journal of the American Statistical Association, 87: 659-675.

[2] Renshaw, A.E., and Haberman, S. 2006 "A cohort-based extension to the Lee-Carter model for mortality reduction factors," Insurance: Mathematics and Economics, 38: 556-570.

[3] Delwarde, A., M. Denuit, and C. Partrat 2007 "Negative binomial version of the Lee–Carter model for mortality forecasting," Applied Stochastic Models In Business And Industry, 23:385–401.

[4] Venter, G.G. 2010 "Mortality Trend Risk," ERM Symposium. (http://www.ermsymposium.org/2010/pdf/2010-erm-chicago-venter-mortality.pdf)

[5] Richards, S.J. 2008 "Applying Survival Models To Pensioner Mortality Data," Presented to the Institute of Actuaries, 25 February 2008, http://www.actuaries.org.uk/__data/assets/pdf_file/0018/103077/sm20080225.pdf

[6] Beard, R. E. 1959 "Note on some mathematical mortality models," In: The Lifespan of Animals, G. E. W. Wolstenholme and M. O'Connor (eds.), Little, Brown, Boston, 302-311.

[7] Perks, W. 1932 "On some experiments in the graduation of mortality statistics," Journal of the Institute of Actuaries, 63: 12-40.

Appendix: Count Distributions

The negative binomial distribution has two parameters r and β, with mean $r\beta$ and variance $r\beta(1+\beta)$. In the sample data there are 1260 cells, and when the negative binomial is used, each cell has values for r and β. The mean $\mu = r\beta$ is the value given by the RH model, but how r and β vary across cells depends on how the model is set up. In the NB1, it is assumed that every cell has the same value of β, so the ratio of variance to mean is $1+\beta$ for every cell. In the NB2 every cell is assumed to have the same value of r, with β set to μ/r, which gives variance to mean ratio $1+\mu/r$, which is higher for the cells with higher means. However there are many other ways the parameters can vary across cells. For instance, suppose there is a constant q for all cells, with r and β given by $r = q\mu^{1/2}$ and $\beta = \mu^{1/2}/q$. Then the mean is still $r\beta = \mu$, and the variance to mean ratio for a cell is $1+ \mu^{1/2}/q$. This can be called the NB3. Its variance/mean ratio is still higher for the larger cells, but not by as much as in the NB2.

This can be generalized to the NBp distribution, which adds a parameter p to control the variance/mean ratio. It sets $r = q\mu^{1-p}$ and $\beta = \mu^{p}/q$. The mean is again $r\beta = \mu$, but

now the variance to mean ratio for a cell is $1+ \mu^p/q$. The value of p can be found by MLE.

When fitting a single NB distribution to a data set, all of these forms are the same. The difference comes when fitting a number of distributions to a number of cells where a common relationship of variance and mean is desired. The NBp forms discussed here by no means exhaust the possible such relationships. In general if the variance/mean ratio desired is $1+G(\mu)$, just set $r = \mu/G(\mu)$ and $\beta = G(\mu)$. For instance, $G(\mu) = q \log(\mu)$ might work in some cases.

To simplify the fitting process, the NBp distribution was used in the fits in the paper. In almost all cases the result was very close to either the NB1 or NB2, so it would be possible to refit using the appropriate distribution in each case, but this approach lets each data set and model find its own best form of the NB with a single optimization.