Making the Most of Experience Data:
An Augmented Beta-Binomial Approach
Discussant: P.J. Sweeting

Presented at the Living to 100 Symposium

Orlando, Fla.

January 5-7, 2011

Abstract

The popular beta-binomial approach to credibility offers an attractive way of combining the results of experience and risk rating. However, when applied to a particular age, the information available from surrounding ages is ignored. In this paper, I propose an augmentation to the beta-binomial approach that not only allows for the information contained in ages near to the age under analysis but also for variation in risk types across the different ages.

Introduction

When considering the appropriate mortality rate to use for a group of lives, two sources of data are available. The first is the mortality experience of that group of lives. Calculating a mortality rate using this approach is known as experience rating. However, the smaller the number of lives, the greater the extent to which calculated mortality rates reflect random volatility rather than the underlying rate of mortality. It is therefore useful to consider the characteristics of the lives under review. The mortality rate for the subsection of a broader population with the same characteristics can then also be calculated and used instead. This is known as risk rating.

Risk rating is, though, imperfect. For any group of lives there may well be idiosyncrasies that have an impact on mortality rates that will not be picked up through risk rating. Consider, for example, the mortality experience of a group of pension plan members. This mortality might be influenced by hard-to-measure factors such as a sense of community. Such factors will not be allowed for in a risk-rating approach using, say, socio-economic group as the key variable.

In practice, therefore, the results of experience and risk rating will be combined using a measure of credibility. This gives an increasing weight to the result of experience rating the larger the number of observations is. However, importance of idiosyncratic factors means it is essential to make as much use of experience data as possible.

Credibility can be expressed as a factor by which the results of experience and risk rating are combined, as shown in Equation 1:

$$\hat{q}_x = Zq_x^e + (1 - Z)q_x^r \tag{1}$$

Here, the initial mortality rate for age x, $\hat{q}_x$, is estimated as a weighted average of the rate derived from mortality experience, q_x^e, and the rate derived from the risk factors, q_x^r. The term Z is the credibility factor, which ranges from zero (for no credibility) to one (for full credibility).

It is also possible to combine the results of experience and risk rating through a single expression, although here it is still possible to calculate the credibility factor implied by the resulting mortality rate. This can be useful not only for finding the most appropriate set of mortality assumptions but also for calculating margins for error as discussed by Hardy and Panjer (1998). One method that is particularly relevant for mortality rating is the beta-binomial approach. This Bayesian approach is discussed by Mayerson (1964) and others, with more detail being given in many standard textbooks on credibility, such as that by Herzog (1999). The mortality rate at age x derived from experience, $\hat{q}_x$, is calculated as the number of deaths at that age, d_x, divided by the initial number of lives, l_x, with l_x being adjusted to reflect any entrants to or exits from the population that might affect the level of exposure. However, it is assumed that underlying mortality rate q_x has a beta distribution with parameters $\beta_{1,x}$ and $\beta_{2,x}$. The expected underlying mortality rate is $q_x^r = E(q_x) = \beta_{1,x}/(\beta_{1,x} + \beta_{2,x})$, and the variance is $Var(q_x) = \beta_{1,x}\beta_{2,x}/(\beta_{1,x} + \beta_{2,x})^2(\beta_{1,x} + \beta_{2,x} + 1)$. This means that if the first two moments of the distribution of the underlying rate are known, $\beta_{1,x}$ and $\beta_{2,x}$ can be determined using the method of moments.

Because the number of deaths occurring at each age has a binomial distribution, the Bayesian probability of there being d_x deaths from a group of l_x lives can be found by using $E(q_x) = \beta_{1,x}/(\beta_{1,x} + \beta_{2,x})$ in the calculation of the binomial probability formula. The result is that the posterior distribution of the expected number of deaths also has a beta distribution with parameters $d_x + \beta_{1,x}$ and $l_x + \beta_{1,x} + \beta_{2,x}$. This means the expected number of deaths is:

$$E(q_x|d_x) = \frac{d_x + \beta_{1,x}}{l_x + \beta_{1,x} + \beta_{2,x}} \tag{2}$$

If $\hat{q}_x$ is substituted for $E(q_x|d_x)$, then Bailey (1950) showed this can be rewritten as:

$$\begin{aligned}\hat{q}_x &= \frac{l_x}{l_x + \beta_{1,x} + \beta_{2,x}} q_x^e + \left(1 - \frac{l_x}{l_x + \beta_{1,x} + \beta_{2,x}}\right) q_x^r \\ &= Z q_x^e + (1 - Z) q_x^r\end{aligned} \tag{3}$$

where the credibility factor $Z = l_x/(l_x + \beta_{1,x} + \beta_{2,x})$.

While this technique gives an intuitively attractive way of combining the results of experience and risk rating, it has a serious drawback. This is that the information on the underlying mortality rate at age x that can be gleaned from the mortality experience from ages either side of age x is ignored. This is important, as even the most robust risk-rating approach can overlook underlying factors important to a particular group of lives, as discussed above. These factors can only be detected through experience rating. The solution is to augment the credibility approach such that information on mortality for ages $x \pm 1$, $x \pm 2$ and so on can be used to supplement the information available on mortality at age x, thus reducing the reliance on risk rating. In this paper, I therefore propose an approach to make better use of the information available from the full mortality experience.

Assumptions and Notation

The objective is to arrive at a single initial mortality rate that can be applied to a group of lives age x that adequately reflects their average level of mortality.

Each individual is assumed to be of some homogeneous mortality group g where $g = 1, 2, \ldots, G$, as described by Richards (2008). Each group refers to a combination of markers, such as whether the individual is a smoker or not, or the socio-economic group to which the individual belongs. The number of deaths at age x for group g is therefore denoted $d_{x,g}$, with the corresponding number of lives being $l_{x,g}$ and the initial mortality rate derived from experience therefore being:

$$q^{e}_{x,g} = \frac{d_{x,g}}{l_{x,g}} \tag{4}$$

The above information relates to the estimate derived from experience rating. For the risk-rating estimate, the mortality rate for individuals of group g and age x in the more general population—or as calculated from some other external source—is $q^{r}_{x,g}$.

As implied above, it is assumed the mortality experience is calculated over a period of one calendar year to give an annual rate. For a longer period of investigation, rates would need to be scaled accordingly.

Approach

The philosophy behind the approach proposed here is that the underlying pattern of mortality rates combined with the rates of mortality at a range of ages can provide information on the rate of mortality at a particular age.

The simplest approach is to assume that, for older ages, the natural logarithm of mortality rates is approximately linear so, ignoring all subscripts apart from x:

$$\ln q_x = a + bx \tag{5}$$

where a and b are constants. It is not essential that this relationship holds for the full range of ages considered, only that it is approximately true for the range of a few years around the age being analyzed. This means, therefore, that to remain valid, this approach can be used only for a range of a few years either side of the age for which mortality rates are being calculated.

The relationship in Equation (5) can then be used, with some adjustment, to artificially increase the number of deaths observed in each group, g, as follows. First, choose the number of years either side of the age for which the mortality rate is being estimated that will be used to give additional information. Let this bandwidth be h years. Then, fit Equation (5) to the ages $x-h$ to $x+h$.

There are a number of ways in which Equation (5) can be fitted. The most straightforward is to use ordinary least squares regression to find a and b, so that the estimated mortality rate is given directly as e^{a+bx}. However, such an approach ignores important information, in particular the size of the population at each age. For consistency with the binomial model used in the calculation of credibility, it is possible to use a binomial maximum likelihood approach to fit Equation (5). This involves first defining the probability that the number of deaths will be as observed for a particular age x as:

$$\Pr(D_{x,g} = d_{x,g}) = \frac{l_{x,g}!}{d_{x,g}!\,(l_{x,g} - d_{x,g})!} (\hat{q}^{e}_{x,g})^{d_{x,g}} (1 - \hat{q}^{e}_{x,g})^{l_{x,g} - d_{x,g}} \tag{6}$$

where $D_{x,g}$ is the random number of deaths for age x and group g, and $\hat{q}^{e}_{x,g}$ is the regression estimate of $q^{e}_{x,g}$. In the above example, the calculation of $\hat{q}^{e}_{x,g}$ is trivial. However, it becomes more interesting if a likelihood function is constructed from probabilities spanning ages from $x-h$ to $x+h$, and if $\ln \hat{q}^{e}_{x,g}$ is a linear function of age. Adding the sub- and superscripts back to Equation (5), the linear relationship estimated can be defined as:

$$\ln \hat{q}^{e}_{x+h,g} = a + b(x+h) \tag{7}$$

This can be rearranged to give $\hat{q}^{e}_{x+h,g}$ in terms of the other variables as:

$$\hat{q}^{e}_{x+h,g} = e^{a+b(x+h)} \tag{8}$$

The parameters a and b are then chosen to maximize the likelihood function, L:

$$L = \prod_{t=-h}^{h} \frac{l_{x+t,g}!}{d_{x+t,g}!\,(l_{x+t,g} - d_{x+t,g})!} (\hat{q}^{g}_{x+t,g})^{d_{x+t,g}} (1 - \hat{q}^{g}_{x+t,g})^{l_{x+t,g} - d_{x+t,g}} \tag{9}$$

A more common alternative is Poisson maximum likelihood estimation based. The standard approach is described by Brouhns et al. (2002) and others. Under this approach, the probability that the number of deaths will be as observed for a particular age x is:

$$\Pr(D_{x,g} = d_{x,g}) = \frac{e^{-\lambda_{x,g}} (\lambda_{x,g})^{d_{x,g}}}{d_{x,g}!} \tag{10}$$

where $\lambda_{x,g} = \hat{q}^{g}_{x,g} l_{x,g}$. This means that the likelihood function to be maximized is:

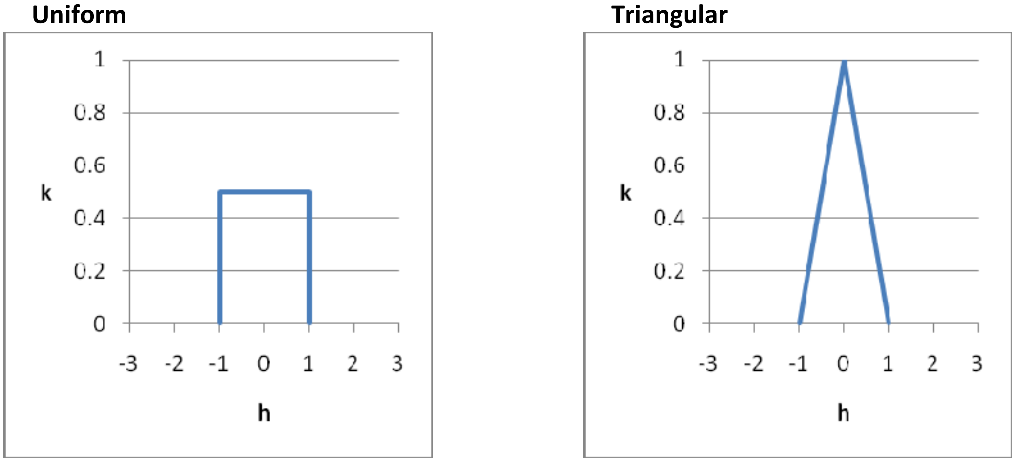
$$L = \prod_{t=-h}^{h} \frac{e^{-\lambda_{x+t,g}} (\lambda_{x+t,g})^{d_{x+t,g}}}{d_{x+t,g}!} \tag{11}$$

Whatever approach is used, the result is a series of regression estimates, $\hat{q}^{g}_{x,g}$. This means the number of deaths that would have been observed in group g at age x from lives actually age $x + h$ had those lives actually been age x can be calculated as:

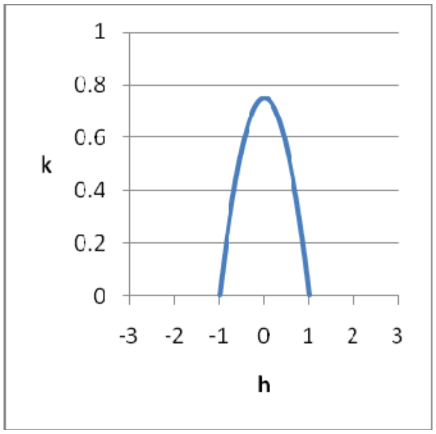
$$\frac{d_{x+h,g}}{l_{x+h,g}\hat{q}^{g}_{x+h,g}} l_{x+h,g} \hat{q}^{g}_{x,g} = \frac{\hat{q}^{g}_{x,g}}{\hat{q}^{g}_{x+h,g}} d_{x+h,g} \tag{12}$$

Figure 1
Various Kernel Functions – Unit Bandwidth

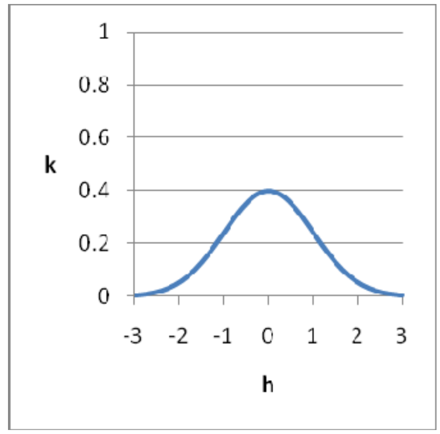
Uniform **Triangular**



Epanechnikov



Normal



At this stage, it is possible simply to aggregate the number of lives and the number of deaths. However, this gives an equal weight to the experience at each age $x+h$. Because a stable population will have experience data that decreases with age, and because the relevance of mortality data will decrease the further away the analysis moves from the age of interest, it makes sense to weight the information somehow. One approach is to use a weighting system based on a kernel function.

A kernel function is a function that weights observations in relation to their distance from the point of interest. These functions are typically used to smooth data. The area under a kernel function is equal to one, but an adjustment is needed when a kernel is applied to discrete data. In particular, this means that if a bandwidth of h is chosen and the kernel weight is given by $k_{|h|}$, the kernel is scaled such that $\sum_{t=-h}^{h} k_{|t|} = 1$.

However, following this approach would do nothing to increase the effective number of lives used. Instead, the kernel weights should be scaled such that $k_0 = 1$. This means that the deaths—and lives— at age x for group g are included at their full weight, while the contribution of the deaths—and lives— for ages $x \pm h$ are given by $k_{|h|}$, which can be anything from zero to one depending on the form of kernel used. The total scaled number of deaths for age x and group g given a bandwidth of h, $d_{x(h),g}$ can be defined as:

$$d_{x(h),g} = \sum_{t=-h}^{h} \frac{q^{s}_{x,g}}{q^{s}_{x+t,g}} d_{x+t,g} k_{|t|} \tag{13}$$

Similarly, the number of lives can be defined as:

$$l_{x(h),g} = \sum_{t=-h}^{h} l_{x+t,g} k_{|t|} \tag{14}$$

This means that for each group, g, the expected initial mortality rate conditional on the number of deaths experienced and a bandwidth of h can be derived from the beta-binomial credibility formula in Equation (2):

$$E(q_x|d_{x(h),g}) = \frac{d_{x(h),g} + \beta_{1,x,g}}{l_{x(h),g} + \beta_{1,x,g} + \beta_{2,x,g}} \tag{15}$$

where $\beta_{1,x,g}$ and $\beta_{2,x,g}$ are the parameters from the beta distribution for group g.

It is worth discussing further the derivation of $\beta_{1,x,g}$ and $\beta_{2,x,g}$. These parameters define the distribution of the underlying mortality rate for each group g at each age x. This mortality rate might be calculated from, say, the population mortality rate for a particular socio-economic group at each age. The mean population mortality rate for each group g and age x has already been defined as $q^{r}_{x,g}$. If the lives at each age in each group are assumed to be homogeneous, then the variance the underlying mortality rate can be calculated as $q^{r}_{x,g}(1 - q^{r}_{x,g})/L_{x,g}$, where $L_{x,g}$ is the number of lives age x from group g in the whole population, adjusted for entrants and exits (probably immigration and emigration in this case) as appropriate.

If the underlying rate of mortality is assumed to have beta distribution, then mean and variance of these rates can be used to calculate $\beta_{1,x,g}$ and $\beta_{2,x,g}$ at each age and for each group using the method of moments.

One drawback of the by-group approach is that it requires separate mortality rates to be calculated for each group g. This means the overall level of credibility is reduced, particularly at older ages where data is limited. However, a review by Guilley et al. (2010) concludes that for extreme old ages, where the data is most sparse, the difference in mortality between socio-economic groups is significantly reduced as genetic factors become more important. This means a single group can be used.

Another potential issue is the choice of bandwidth. As alluded to above, a large bandwidth would stretch the assumption that mortality rates are log-linear with respect to age. However, using this approach with a bandwidth of only three years either side of the age of interest can give result in almost four times the information on deaths if an Epanechnikov kernel form is used.

The fact that a bandwidth must be chosen is also an issue at the upper age range. However, this can be dealt with by reducing the bandwidth as the upper age limit is approached, so that at the highest age only information from that age is used. The practical impact of this approach is that an increasing weight is given to the risk-rating mortality estimate, which, in the absence of any better data, seems appropriate.

Conclusion

Credibility is an important way of combining the results of experience and risk rating. However, it is important that as much information as possible is used in the calculation of the mortality rates, including information from rates either side of the rate under analysis.

This technique can be particularly helpful at older ages where the number of lives may be small. Indeed, where the number of lives varies significantly from age to age, as is often the case for very high ages, this approach can allow experience data to provide a useful contribution at an age where the actual experience is limited, providing there is greater experience at surrounding ages.

The proposed approach does have limitations. In particular, the division of the mortality experience into homogeneous groups reduces the degree of credibility, although the absence of grouping at older ages limits the impact of this problem. The choice of bandwidth must also be made carefully to ensure that the increase in useful data does not mean that log-linearity ceases to be a realistic approximation. However, this approach does offer the prospect that experience data might be able to play a bigger role in the mortality rates assumed.

References

Bailey, A.L. 1950. "Credibility Procedures: LaPlace's Generalization of Bayes' Rule and the Combination of Collateral Knowledge with Observed Data." *Proceedings of the Casualty Actuarial Society* 37: 7-23.

Brouhns, N., M. Denuit, and J.K. Vermunt. 2002. "A Poisson Log-bilinear Regression Approach to the Construction of Projected Life Tables." *Insurance: Mathematics and Economics* 31, no. 3: 373-93.

Guilley, E., M. Bopp, D. Faeh, and F. Paccaud. 2010. "Socioeconomic Gradients in Mortality in the Oldest Old: A Review." *Archives of Gerontology and Geriatrics* 51: e37-e40.

Hardy, M.R., and H.H. Panjer. 1998. "A Credibility Approach to Mortality Risk." *ASTIN Bulletin* 28, no. 2: 269-83.

Herzog, T.N. 1999. *Introduction to Credibility Theory* 3rd ed., ACTEX Publications: Winsted.

Mayerson, A.L. 1964. "A Bayesian View of Credibility." *Proceedings of the Casualty Actuarial Society* 51: 85-104.

Nielson, J.P., and B.L. Sandqvist. 2000. "Credibility Weighted Hazard Estimation." *ASTIN Bulletin* 30, no. 2: 405-17.

Richards, S.C. 2008. "Applying Survival Models to Pensioner Mortality Data." *British Actuarial Journal* 14, no. 2: 257-303.