

# MIS-ESTIMATION RISK: MEASUREMENT AND IMPACT

BY S. J. RICHARDS

UPDATED ON 3RD FEBRUARY 2016

© LONGEVITAS LTD. ALL RIGHTS RESERVED.

## ABSTRACT

When deriving a demographic basis from experience data it is useful to know (i) what uncertainty surrounds that basis, and (ii) the financial impact of that uncertainty. Under the Solvency II regime in the European Union, insurers must hold capital against a number of risks. One of these is mis-estimation risk, i.e. the uncertainty over the current rates of mortality and other biometric risks experienced by a portfolio. We propose a general method for assessing mis-estimation risk, and by way of illustration we look at how mis-estimation risk can be assessed for a portfolio of pensions in payment from a U.K. pension scheme. We find that the impact of mis-estimation risk varies according to the risk factors included in a model, and that the inclusion of some necessary risk factors increases the financial impact of mis-estimation risk. In particular, the inclusion of risk factors which improve the model's fit and financial applicability can lead to an increase in the mis-estimation risk. We also find that a full portfolio valuation is preferable to using model points when assessing mis-estimation risk.

## KEYWORDS

mis-estimation risk, parameter risk, Solvency II, mortality risk, longevity risk, survival model, annuities.

## CONTACT ADDRESS

Stephen Richards, 24a Ainslie Place, Edinburgh, EH3 6AJ, United Kingdom; Telephone: +44 (0)131 315 4470; Email: [stephen@longevitas.co.uk](mailto:stephen@longevitas.co.uk); Web: [www.longevitas.co.uk](http://www.longevitas.co.uk)

## ACKNOWLEDGEMENTS

The author thanks Gavin P. Ritchie, Dr. Shane F. Whelan, Dr. Iain D. Currie, Dr. Matthias Börger, Professor Andrew J. G. Cairns and Kai Kaufhold for helpful comments. Data validation and preparation for modelling were done using Longevitas [Longevitas Development Team (2014)], which was also used to fit all the models and assess mis-estimation risk. Graphs were done in R [R Core Team (2012)] and typesetting was done in pdfL<sup>A</sup>T<sub>E</sub>X. Any errors or omissions remain the sole responsibility of the author.

## 1 Introduction

**1.1** In this paper we will consider mis-estimation risk in setting a demographic basis, i.e. the risk that the current estimate of biometric risk is incorrect. We will leave to one side the question of projecting future rates and the uncertainty therein, as this is usually dealt with separately. The methodology in this paper applies to bases for any demographic risk which can be modelled statistically. However, for simplicity we will illustrate our points with reference to a single-decrement example, i.e. the mortality rates for a portfolio of pensions in payment.

**1.2** The methodology presented in this paper requires that the basis is derived from the portfolio's own experience. There are, of course, other techniques available where the portfolio has insufficient amounts of its own data. However, techniques which do not use a portfolio's own experience data introduce *basis risk*, i.e. the risk that the rating method fails to capture some portfolio-specific characteristics. As a result it is usually preferable to use a portfolio's own experience data wherever it is both available and credible. What counts as "credible" will be partly quantitative (number of lives and deaths, total exposure time) and partly qualitative (are the data free of obvious corruption?), so judgement will have to be on a case-by-case basis.

**1.3** When setting a basis using experience analysis, there are two natural questions to ask: “what other bases could be credibly supported by the data?” and “what would be the financial impact of using those other bases?”. In this paper we will demonstrate a method for answering both of these questions. Besides the obvious use of such answers in pricing and risk management, some regulatory regimes require that insurers itemise the capital held against specific risks, including mis-estimation risk. Examples include the former ICA regime in the U.K. and the Solvency II regime in the EU. Richards et al. (2014) give a stylised list of sub-risks for longevity, which is reproduced in Table 1. One of these is mis-estimation risk, i.e. the risk that an insurer has got its current estimate of risk rates wrong. In territories with this kind of itemised approach, it is implicit that the capital requirements be calibrated probabilistically, i.e. with reference to some kind of statistical model. The method demonstrated in this paper is based on a well-specified statistical model. In the U.K. insurers currently use a wide variety of ad hoc methods for assessing mis-estimation risk, but this paper is not a review or summary of these methods. Instead, this paper proposes an objective method for calculating an allowance for mis-estimation risk, based on robust statistical foundations.

**1.4** At a high level our approach is to use the variance-covariance matrix to explore consistent alternative parameter vectors. This makes mis-estimation risk synonymous with parameter risk. A full-portfolio valuation is performed with each new parameter vector, and this gives rise to a distribution of possible expected portfolio values. We use this distribution to set the allowance for mis-estimation risk, typically by using a particular quantile such as 99.5% or calculating a contingent tail expectation (CTE). In this paper we will illustrate our approach to setting the capital for mis-estimation risk with reference to a portfolio of annuities or pensions in payment. The basic principles also apply to other types of insurance risk.

**1.5** Other authors have looked at the subject of mis-estimation risk. For example, Hardy and Panjer (1998) used a credibility approach based around A/E ratios against a standard table. However, in this paper we will use a parametric model with risk factors. The conversion of the results into a percentage of a standard table is covered in Appendix 4.

**1.6** The plan of the paper is to first define the components of a longevity-risk module in Section 2. Having defined what is and is not covered by mis-estimation risk, we give a short introduction to one-parameter mortality modelling in Section 3, which includes a demonstration of the key properties of maximum-likelihood estimates. This is followed in Section 4 by an illustration of how these maximum-likelihood properties can be used to assess the impact of mis-estimation risk on an insurance liability. Section 4 also demonstrates how simulation can be used to assess mis-estimation risk in place of an analytical, stress-test approach. In Section 5 we extend the mortality model to include an arbitrary number of parameters, and in Section 6 we look at assessing mis-estimation risk for a basic two-parameter model. Section 7 looks at the benefits of using data which span multiple years, and how these benefits can be more modest than might be expected where there is a time trend in the basic mortality level. Section 8 considers the minimum requirements of a model for financial applications, and the resulting impact on mis-estimation risk. Section 9 considers the impact of portfolio size and some pitfalls to guard against, while Section 10 concludes the paper.

**1.7** In this paper we will denote a single parameter by  $\theta$  and a vector of multiple parameters by  $\underline{\theta}$ . The unknown underlying value of a parameter will be marked with an asterisk (\*), while an estimate of that parameter will be marked with a circumflex ( $\hat{\cdot}$ ). A stressed estimate of a parameter will be marked with  $\prime$ . All numerical examples are based on the actual data for a pension scheme in England & Wales, details of which are given in Appendix 1.

## 2 Components of longevity risk

**2.1** In modern insurance work it is often necessary to quote a single capital amount or percentage of reserve held in respect of a risk such as longevity. This single figure is usually made up of sub-risks, such as those itemised in Richards et al. (2014) and reproduced in Table 1.

**2.2** Table 1 is not intended to be exhaustive and, depending on the nature of the liabilities, other longevity-related elements might appear. In a defined-benefit pension scheme, or in a portfolio of bulk-purchase annuities, there would be uncertainty over the proportion of pensioners who were married, and whose death might therefore lead to the payment of a spouse’s pension. Similarly, there would be

Table 1: Sample itemization of the components of longevity risk. A diversifiable risk can be reduced by growing the size of the portfolio and benefiting from the law of large numbers. Reproduced from Richards et al. (2014).

| Component           | Diversifiable? | Comment   |
|---------------------|----------------|---|
| Model risk          | No             | It is impossible to know if the selected model is correct. Capital must be held in respect of the risk that one's chosen model is wrong. Model risk applies not only to the projection model, but also to the risk factors included in a model of current differentials.  |
| Basis risk          | No             | Models are sometimes calibrated to population or industry data, not the data of the portfolio in question. This is particularly the case for some projection models. However, there are cases where pension schemes in particular have mortality bases set without reference to the portfolio's own experience. Capital must be held for the risk that the lives in a portfolio are different from those of the population used to calibrate the model. |
| Trend risk          | No             | Even if the projection model is correct and there is no basis risk, an adverse trend may result by chance which is nevertheless fully consistent with the chosen model.   |
| Volatility          | Yes?           | Over a one-year time horizon, capital must be held against the case of unusually light mortality experience from seasonal or environmental variation, such as an unusually mild winter and lower-than-normal deaths due to influenza and other infectious diseases. Note that this risk may not be wholly diversifiable, as one year's light mortality experience may equally be the start of an adverse trend.   |
| Idiosyncratic risk  | Yes            | Over a one-year time horizon, capital must be held against the case of unusually light mortality experience from random individual variation. See Plat (2011) and Richards & Currie (2009) for examples.  |
| Mis-estimation risk | Yes            | Uncertainty exists over the portfolio's actual underlying mortality rates, since these can only be estimated to a degree of confidence linked to the scale and richness of the data. This is the subject of this paper. As our approach involves using a parametric model, mis-estimation risk here is synonymous with parameter risk for current mortality rates.  |

uncertainty over the age of that spouse. Within an active pension scheme there might be risk related to early retirements, commutation options or death-in-service benefits. These risks might be less important to a portfolio of individual annuities, but such portfolios would be exposed to additional risk in the form of anti-selection from policyholder behaviour. An example of this would be the existence of the enhanced-annuity market in the U.K.

**2.3** This paper will only address the mis-estimation component of Table 1, so the figures in Tables 3, 4 and 5 can only be minimum values for the total capital requirement for longevity risk. Other components will have to be estimated in very different ways: reserving for model risk requires a degree of judgement, while idiosyncratic risk can best be assessed using simulations of the actual portfolio. For large portfolios the idiosyncratic risk will often be diversified away almost to zero in the presence of the other components. In contrast, trend risk and model risk will always remain, regardless of how large the portfolio is.

### 3 A one-parameter primer in mortality analysis

**3.1** Before we illustrate the full multivariate approach in Section 5, we begin with a one-parameter primer to establish the basics. This section looks at the model behind the mortality analysis, while Section 4 looks at how it can be applied to the task of assessing mis-estimation risk. We assume for simplicity that we have a risk which can be modelled with a single parameter,  $\theta$ . We further assume that we have a log-likelihood function,  $\ell(\theta)$ , which can be differentiated at least twice.  $\ell$  is maximised at  $\hat{\theta}$ , i.e. where Equations 1 and 2 are satisfied:

$$\frac{\partial}{\partial \theta} \ell(\hat{\theta}) = 0 \quad (1)$$

$$\frac{\partial^2}{\partial \theta^2} \ell(\hat{\theta}) < 0 \quad (2)$$

**3.2**  $\hat{\theta}$  is then the maximum-likelihood estimate (MLE) of the unknown true parameter,  $\theta^*$ . The maximum-likelihood theorem states that  $\hat{\theta}$  has a normal distribution with unknown mean  $\theta^*$  and unknown variance  $\sigma^2$  (Cox and Hinkley, 1996, p296). Since  $\hat{\theta}$  is an estimate for  $\theta^*$ , and since the curvature in Equation 2 is inversely related to the variance of  $\hat{\theta}$ , we can use the approximation  $\hat{\theta} \sim N(\hat{\theta}, \hat{\sigma}^2)$ , where  $\hat{\sigma}^2 = \left[ -\frac{\partial^2}{\partial \theta^2} \ell(\hat{\theta}) \right]^{-1}$ .

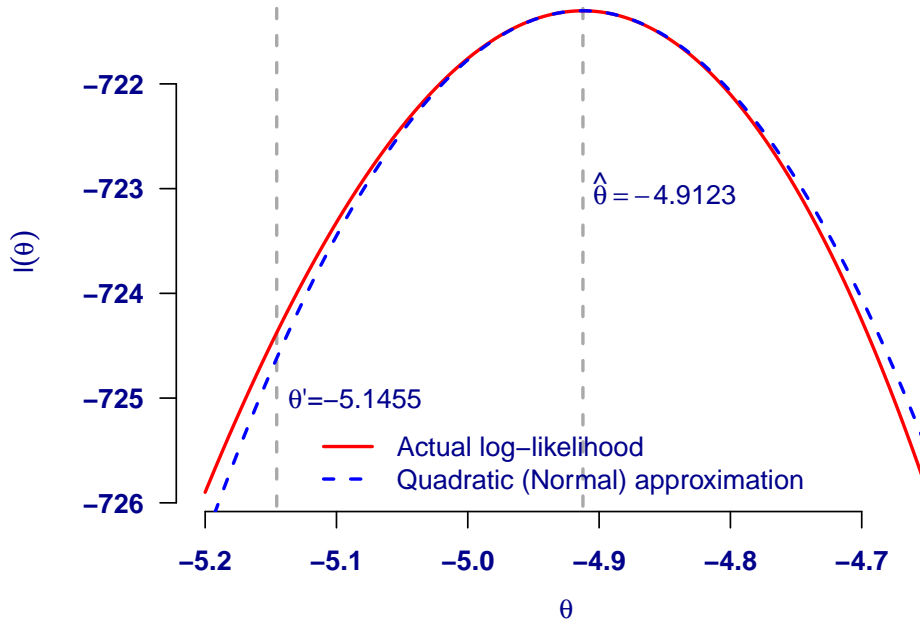
**3.3** To illustrate this, consider a simplistic example assuming constant mortality between the ages of 60 and 65. We assume that  $\theta$  represents the logarithm of the constant force of mortality, i.e.  $\mu_x = e^\theta$  for all ages  $x \in [60, 65]$ . Past experience of the scheme in Appendix 1 observes  $d = 122$  deaths out of  $E^c = 16,586.3$  life-years lived by 6,439 pensioners between 2007 and end-2012. Assuming a Poisson-distributed number of deaths, the likelihood function,  $L$ , is given in Equation 3:

$$L(\theta | d, E^c) \propto \frac{e^{-E^c e^\theta} (E^c e^\theta)^d}{d!} \quad (3)$$

and so the log-likelihood simplifies to Equation 4:

$$\begin{aligned} \ell(\theta | d, E^c) &= \log_e L \\ &= -E^c e^\theta + d\theta + \text{constant} \end{aligned} \quad (4)$$

**3.4** Equations 3 and 4 are maximised at  $\hat{\theta} = \log_e \frac{d}{E^c}$ . A plot of the log-likelihood function is shown as the solid line in Figure 1, where the function is seen to be maximised at  $\hat{\theta} = \log_e \frac{122}{16586.3} = -4.9123$ . The curvature around this value is  $\frac{\partial^2}{\partial \theta^2} \ell = -E^c e^\theta$ , and so the approximate standard error of  $\hat{\theta}$  is  $\hat{\sigma} = \sqrt{(E^c e^{\hat{\theta}})^{-1}} = 0.09054$ . The near-quadratic form of the log-likelihood function is consistent with  $\hat{\theta}$  having an approximately Normal distribution with mean  $\hat{\theta}$  and variance  $\hat{\sigma}^2$ , the log-likelihood for which is shown as the dashed line in Figure 1. Thus, the log-likelihood gives us an estimate for  $\theta$  and an idea of what other estimates would be consistent with the data. We also have a close approximation for the log-likelihood, which can simplify the generation of likely alternative values that are consistent with the data. We can now assess the impact of these plausible alternative estimates on the value of a liability, i.e. to assess mis-estimation risk, which we do in Section 4.

Figure 1: Log-likelihood in Equation 4 with  $d=122$  and  $E=16,586.3$ .

## 4 A one-parameter primer in mis-estimation risk

**4.1** Whereas  $\theta$  is the parameter describing the mortality risk in Section 3, actuaries have to value a monetary liability based on that risk. This can be described as a function of the risk parameter, say  $a(\theta)$ .  $a(\theta)$  could be a valuation function for a single policy or, more usefully, the valuation of the liability for an entire portfolio. The best-estimate of the liability in our simplistic mortality model would therefore be  $a(\hat{\theta})$ . However, to allow for mis-estimation risk we would calculate something like  $a(\theta')$ , where  $\theta'$  was a stressed alternative value to  $\hat{\theta}$  which was less likely but nevertheless consistent with the observed data. For example, in Figure 1 an approximate 99.5% stress value for low mortality would be around  $\theta' = -5.1455$  (see ¶4.4 for derivation of  $\theta'$ ).

**4.2** We can see from Figure 1 that the log-likelihood for  $\theta$  is nearly quadratic and thus that  $\hat{\theta}$  has an approximate Normal distribution. This means that we can generate alternative values to  $\hat{\theta}$  which are consistent with the data by drawing values from a Normal distribution with mean  $\hat{\theta}$  and variance  $\hat{\sigma}^2$ . In other words, we can generate  $\theta' = \hat{\theta} + \hat{\sigma}Z$ , where  $Z$  represents a value from the cumulative distribution function for a  $N(0,1)$  variable,  $\Phi(\cdot)$ . For finding a specific, consistent-but-stressed value for  $\theta'$  at a given p-value, we would calculate either  $Z = \Phi^{-1}(p)$  or  $Z = \Phi^{-1}(1-p)$ , depending on whether increasing or decreasing  $\theta$  raises or lowers the liability function,  $a(\theta)$ .

**4.3** To illustrate this, assume that we want to value a temporary pension from age 60 to age 65. The survival curve for our simple model is given by  ${}_t p_{60} = e^{-te^\theta}$ . Ignoring discounting for simplicity, the valuation function for a continuously paid five-year temporary annuity is given in Equation 5:

$$a(\theta) = \frac{1 - e^{-5e^\theta}}{e^\theta} \quad (5)$$

**4.4** In our example from ¶3.4, the best-estimate liability is  $a(\hat{\theta}) = a(-4.9123) = 4.9092$ . If we then wanted to find the 99.5<sup>th</sup> percentile for mis-estimation risk, we would use  $Z = \Phi^{-1}(0.005) = -2.5758$  (Lindley and Scott, 1984, p35), as lowering  $\theta$  increases the liability. The value of the liability allowing for mis-estimation risk at the 99.5% level is then  $a(\hat{\theta} + \hat{\sigma}Z) = a(-5.1455) = 4.9279$ , i.e. 0.38% higher

than the central estimate. Our 99.5% allowance for mis-estimation risk would then be 0.38% of the best-estimate reserve.

**4.5** An implicit assumption in ¶4.1 is that  $a(\theta)$  is a monotonically increasing or decreasing function of  $\theta$  (which is the case in Equation 5). If the liability function  $a(\theta)$  is not simple or neatly behaved, then we can use simulation to generate values of  $Z$ . Specifically, we can repeatedly draw  $Z$  from the  $N(0,1)$  distribution, use  $Z$  to calculate  $\theta'$  and thus  $a(\theta')$ , then add the value of  $a(\theta')$  to a set,  $S$ . We can then calculate the appropriate percentile of  $S$  as an estimate for the liability allowing for mis-estimation risk. A short R script for doing this with our simplistic temporary-annuity example is given below:

```
set.seed(-1) # Set random-number seed
Z = rnorm(10000) # Generate 10,000 N(0,1) values
thetaprime = -4.9123 + 0.09054*Z # Calculate parameter values
S = (1-exp(-5*exp(thetaprime)))/exp(thetaprime) # Evaluate equation (5)
Q = quantile(S,0.995) # Calculate 99.5th percentile
round(Q, 4) # Display rounded percentile
```

which gives a value of 4.9275 for the 99.5<sup>th</sup> percentile, i.e. close to the analytical value in ¶4.3. A more efficient estimate of the 99.5<sup>th</sup> percentile can be obtained by using the estimator from Harrell and Davis (1982), as shown in the R script below:

```
library(Hmisc) # Library for Harrell-Davis
Q = hdquantile(S, 0.995, se=TRUE, names=FALSE) # Calculate 99.5th percentile
round(Q, 4) # Display rounded percentile
```

which produces a value of 4.9278 for the 99.5<sup>th</sup> percentile, i.e. even closer agreement with the analytical approximation in ¶4.3. In subsequent sections we will see why correlations between parameters mean that, in most practical situations, we will invariably have to use the simulation approach of ¶4.5. Appendix 3 considers methods for assessing mis-estimation risk which do not require large numbers of simulations.

**4.6** In Section 5 we will describe the general, multi-parameter case involving simulation, then in Section 6 we will return to a simple two-parameter example for illustrative purposes.

## 5 The multi-parameter case

**5.1** We now assume a more realistic case where there are multiple parameters in a vector,  $\underline{\theta}$ . As in Section 3, we have a log-likelihood function,  $\ell(\underline{\theta})$ , which we assume we can differentiate at least twice. The vector  $\underline{\theta}$  is the analogue of the scalar  $\theta$  in Section 3, while the analogue of  $\sigma^2$  is the variance-covariance matrix of  $\underline{\theta}$ , say  $\mathbf{V}$ . As in ¶3.2, we have an unknown true parameter vector,  $\underline{\theta}^*$ , and an unknown true variance-covariance matrix,  $\mathbf{V}^*$ .

**5.2** Unlike in the one-parameter case in ¶4.3, there is no easy analytical option to calculate the liability value allowing for mis-estimation. This is because the analogue of the scalar  $\sigma^2$  in ¶4.3 is a matrix,  $\mathbf{V}$ , i.e. the parameters in  $\underline{\theta}$  will have various positive and negative correlations with each other. A visual example of this is given in Figure 4, where changing the value of the intercept ( $\alpha_0$ ) leads to a changed value for the slope ( $\beta_0$ ). Thus, to assess the impact of mis-estimation risk we have to perform repeated valuations of the entire portfolio using a series of consistent alternative parameter vectors, analogous to the procedure in ¶4.5. We repeat this process of valuation, which gives rise to a set,  $S$ , of portfolio valuations.  $S$  can then be used to set a capital requirement to cover mis-estimation risk. For example, Solvency II regime in the European Union demands that this calculation takes place at the 99.5th percentile, and so the capital requirement would be given by Formula 6:

$$\left( \frac{\text{99.5th percentile of } S}{\text{mean of } S} - 1 \right) \times 100\% \quad (6)$$

**5.3** The question then is how we generate those “consistent alternative parameter vectors”. To do this we use the same result from the theory of maximum likelihood as in ¶3.2, which states that the joint estimate at the MLE is distributed as a multivariate Normal random variable with mean  $\underline{\theta}^*$  and a variance-covariance matrix,  $\mathbf{V}^*$  (Cox and Hinkley, 1996, p296). As in the one-parameter case, the true value of  $\underline{\theta}^*$  is unknown, and so we substitute  $\hat{\underline{\theta}}$ , the vector of maximum-likelihood estimates. We now seek a similar substitute for the unknown  $\mathbf{V}^*$ .

**5.4** In some software packages an estimated variance-covariance matrix is available directly. For example, in R (R Core Team, 2012) an estimate of  $\mathbf{V}^*$  is returned from using the `vcov()` function on a model object. However, actuaries often need to fit models which are unavailable in such software (Richards, 2012), so it is useful to outline the general principle for estimating  $\mathbf{V}^*$  from knowledge of  $\ell(\underline{\theta})$  alone.

**5.5** Let  $\mathcal{H}(\underline{\theta})$  be the matrix of second-order partial derivatives of the log-likelihood function, known to mathematicians as the Hessian (McCullagh and Nelder, 1989, p6). Let  $\mathcal{I} = -\mathcal{H}(\hat{\underline{\theta}})$ , i.e. the negative Hessian evaluated at the MLE  $\hat{\underline{\theta}}$ .  $\mathcal{I}$  is known to statisticians as the observed information matrix (sometimes also called the observed Fisher information). The diagonal elements of  $\mathcal{I}^{-1}$  are the Cramer-Rao lower bounds for the diagonal elements of  $\mathbf{V}^*$  (Cox and Hinkley, 1996), and for practical purposes we substitute  $\mathcal{I}^{-1}$  for  $\mathbf{V}^*$ .

**5.6** We thus have a multivariate Normal distribution for the MLE vector with mean  $\hat{\underline{\theta}}$  and variance-covariance matrix  $\mathcal{I}^{-1}$ , i.e.  $\text{MVN}(\hat{\underline{\theta}}, \mathcal{I}^{-1})$  in place of  $\text{MVN}(\underline{\theta}^*, \mathbf{V}^*)$ . We can use this to simulate consistent vectors of alternative parameters as a means of investigating parameter risk and thus mis-estimation risk. To do this we calculate the expression in Formula 7:

$$\hat{\underline{\theta}} + A\mathbf{z} \tag{7}$$

where  $\mathbf{z}$  is a vector of independent, identically distributed  $N(0,1)$  variates with the same length as  $\hat{\underline{\theta}}$ . The matrix  $A$  represents the “square root” of  $\mathcal{I}^{-1}$ , of which there are generally several non-unique possibilities. However, the variance-covariance matrix  $\mathcal{I}^{-1}$  is a non-negative, definite matrix, and it is positive-definite apart from some trivial cases — see Lindgren (1976, p464) for more details. We therefore set  $A$  to be the Cholesky decomposition of  $\mathcal{I}^{-1}$ , i.e.  $A$  is a lower-triangular matrix such that  $AA^T = \mathcal{I}^{-1}$  — see Venables and Ripley (2002, pages 62 and 422). From this we can use Equation 7 to simulate parameter error consistent with the data, and thus use these perturbed parameter vectors to value the portfolio and explore mis-estimation risk.

**5.7** We then only require to calculate (i) the first derivatives of  $\ell$  for finding the joint maximum-likelihood estimates, and (ii) the second partial derivatives to calculate  $\mathcal{I}$ . These derivatives can be either worked out analytically or else approximated using finite differences. Appendix 2 considers these two approaches, and finds that analytical derivatives are strongly preferred. We therefore use analytical derivatives throughout the main body of this paper.

## 6 A two-parameter case — the importance of acknowledging correlations

**6.1** In Section 3 we had a simple one-parameter model behind the log-likelihood function, while the liability function in Section 4 was not particularly realistic. In this section we illustrate the calculation of mis-estimation risk for a pension paid throughout life and calibrate the mortality model using a wider age range from the pension-scheme data in Appendix 1. Although the mortality model only contains two parameters, the example is sufficient to demonstrate the critical importance of acknowledging correlations between parameters when assessing mis-estimation risk.

**6.2** We start by building  $\ell(\underline{\theta})$  using a survival model for the force of mortality,  $\mu_x$ , which is defined in Equation 8:

$$\begin{aligned}\mu_x &= \lim_{h \rightarrow 0^+} \frac{1}{h} \Pr(\text{death before age } x+h | \text{alive at age } x) \\ &= \lim_{h \rightarrow 0^+} \frac{h q_x}{h}\end{aligned}\tag{8}$$

and where the survival probability from age  $x$  to age  $x+t$ ,  ${}_t p_x$ , is given by Equation 9 for any form of  $\mu_x$ :

$${}_t p_x = \exp\left(-\int_0^t \mu_{x+s} ds\right)\tag{9}$$

**6.3** By using survival models we will therefore be modelling mortality at the level of the individual, rather than the group-level modelling of Section 3. For further details of actuarial applications of survival models to pensioner and annuitant mortality the reader can consult Richards et al. (2013).

**6.4** In this section we use a simple, two-parameter Gompertz (1825) model for each life in Equation 10:

$$\mu_{x_i} = e^{\alpha_0 + \beta_0 x_i}\tag{10}$$

**6.5** The model in Equation 10 is a simple model in age only, i.e. ignoring gender or any of the other known relevant risk factors (inclusion of other risk factors is considered in Section 8). The results of fitting the model to the pension-scheme experience data described in Appendix 1 are shown in Table 2, while Figure 2 shows the essentially quadratic profile of the log-likelihood for the two parameters. Note that we are applying a deliberately over-simple model to a fraction of the available data in order to reveal some important basic features. In Section 8 we will make the model more realistic and practical.

Table 2: Summary of simple Gompertz (1825) model fitted to 2012 mortality experience of the pension scheme described in Appendix 1. Data for ages 60 and over, males and females combined.

| Parameter                | Estimate | Standard error | Z-value | Pr(> z ) | Sig. | Lives  | Deaths |
|--------------------------|----------|----------------|---------|----------|------|--------|--------|
| Age ( $\beta_0$ )        | 0.122872 | 0.00564083     | 21.78   | 0        | ***  | 13,085 | 365    |
| Intercept ( $\alpha_0$ ) | -12.972  | 0.466992       | -27.78  | 0        | ***  | 13,085 | 365    |

Note: Parameter significance is labelled according to the same scheme used in R [R Core Team (2012)], i.e. . for 10%, \* for 5%, \*\* for 1% and \*\*\* for 0.1%.

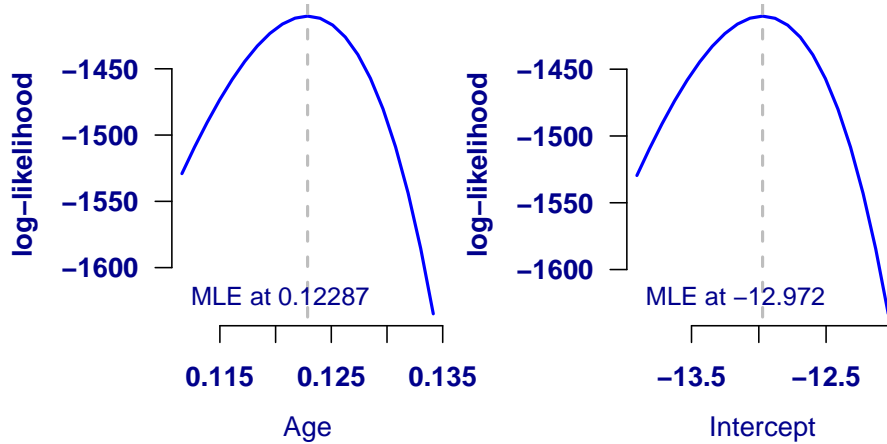
**6.6** Figure 3 shows the fitted mortality hazard from Table 2 against the crude mortality hazard for the pension scheme. We will not get side-tracked with questions of quality of fit or adequacy at this point, as our aim is to demonstrate mis-estimation risk. However, we note that Section 8 introduces the concept of a *financially suitable* model as a critical foundation for assessing mis-estimation risk, together with a test for determining financial suitability.

**6.7** In our model in Table 2 we have the maximum-likelihood estimates  $\hat{\alpha}_0 = -12.972$  and  $\hat{\beta}_0 = 0.122872$ . If we define our MLE vector as  $\hat{\theta} = (\hat{\beta}_0, \hat{\alpha}_0)'$ , then the estimated variance-covariance matrix using the approach defined in ¶5.4 is as follows:

$$\begin{array}{c} \hat{\beta}_0 \quad \hat{\alpha}_0 \\ \hat{\beta}_0 \begin{pmatrix} 3.18189 \times 10^{-5} & -0.00261762 \\ -0.00261762 & 0.218081 \end{pmatrix} \\ \hat{\alpha}_0 \end{array}$$



Figure 2: Log-likelihood profiles for model in Table 2. These profiles demonstrate the quadratic shape around the MLEs, which is consistent with the multivariate Normal distribution for the estimates used in Equation 7.



**6.8** The correlation between  $\hat{\alpha}_0$  and  $\hat{\beta}_0$  is  $-99.4\%$  ( $-99.4\% = \frac{-0.00261762}{\sqrt{3.18189 \times 10^{-5} \times 0.218081}} \times 100\%$  (see Table 9 for further illustration of this)). In other words, if  $\hat{\alpha}_0$  is mis-estimated then  $\hat{\beta}_0$  changes in the opposite direction by an almost perfectly known amount (and vice versa). This is illustrated in Figure 4, where we stress the intercept by 1.96 standard errors and re-estimate the age slope. Due to the strong correlation, a change in one parameter is accompanied by an important offsetting change in another. This is why a simple parallel shift in mortality table is not in general a correct statement of mis-estimation risk: the level of mortality (as represented by  $\alpha_0$ ) and the increase with age (as represented by  $\beta_0$ ) are highly negatively correlated. A downward shift in level would result in an upward shift in the rate of increase by age, as shown in Figure 4. Mortality levels and rates of change by age are generally negatively correlated, as demonstrated later in Table 9, and this topic is explored in some detail for various risk factors in Richards et al. (2013). Appendix 5 considers how to restructure a model to reduce parameter correlation, but this only works for the very simplest models. It is therefore important that any assessment of mis-estimation risk should acknowledge these correlations.

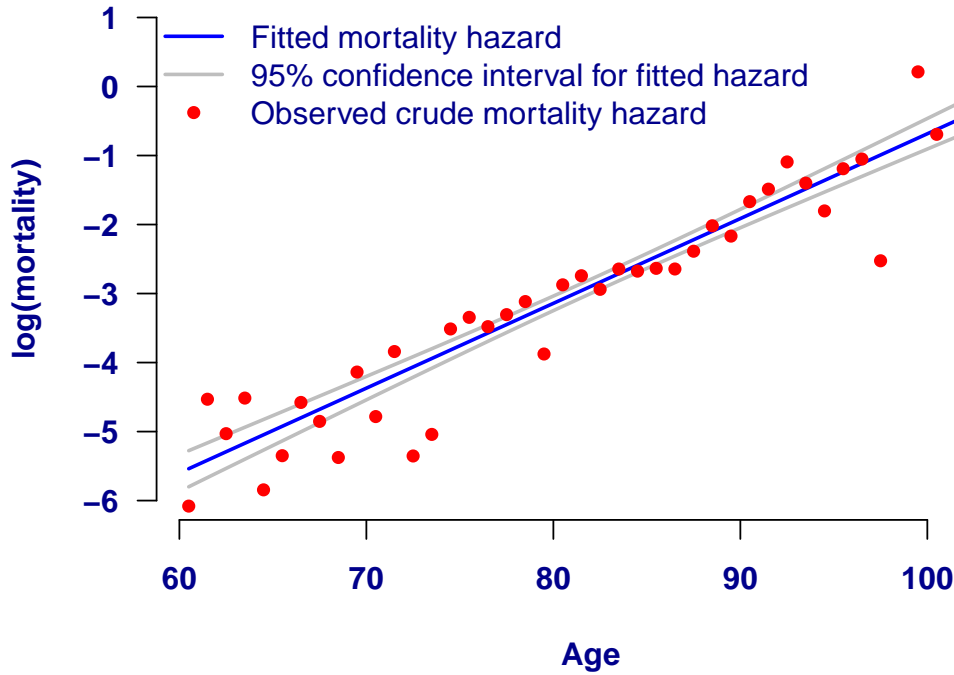
**6.9** However, there are further consequences of the variance-covariance matrix in ¶6.7, namely that the impact of mis-estimation risk varies by age. Figure 3 shows the 95% confidence interval for the fitted mortality hazard, which forms a bowed shape courtesy of the correlation in ¶6.8. Within the model structure, the relative uncertainty is greatest at the youngest and oldest ends of the age range. The reason for this is that we are fitting a straight line, which must go through the data in the central age range. As a consequence, any mis-estimation of  $\alpha_0$  will cause a change in the estimation of  $\beta_0$ . As with a child's see-saw, the change in fit near the centre will be much smaller than at either end. This same see-saw phenomenon will have major consequences for using multi-year data, as covered in Section 7.

**6.10** It is therefore important that any assessment of mis-estimation risk should acknowledge the age-related distribution of liabilities as well as the parameter correlations. To demonstrate this, assume that the liability function is for a single annuity at outset age  $x$ , i.e.  $a(\underline{\theta})$  is defined as in Equation 11:

$$\begin{aligned} a(\underline{\theta}) &= \bar{a}_x \\ &= \int_0^{\infty} {}_tP_x v(t) dt \end{aligned} \tag{11}$$

where  $v(t)$  is the continuous-time discount function. We use a constant net annual discount rate of 1% in this paper — U.K. government gilts with a maturity of 14–18 years yield around 3% at the time of writing and the Bank of England has a 2% inflation target. However, it could also be argued that

Figure 3:  $\log_e(\text{crude mortality hazard})$  and 95% confidence intervals for fitted model for ages 60 and over. Data and model from Table 2.



Note: The crude mortality hazard is the actual number of deaths in the age interval  $[x, x+1)$  divided by the time lived in that interval.

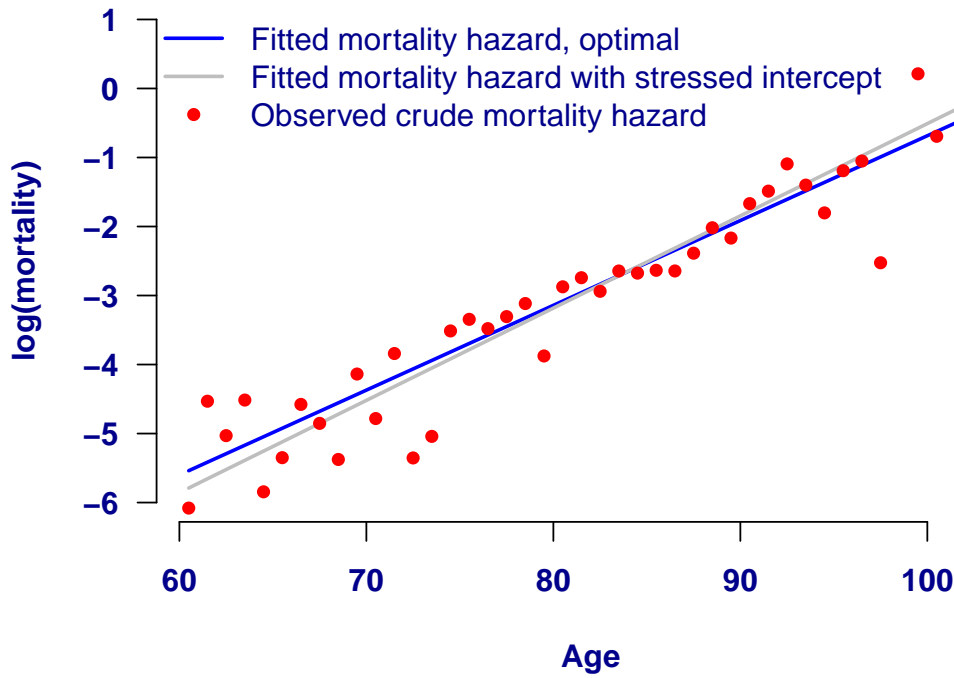
index-linked gilt yields are relevant for valuing pension-scheme cashflows, and the yields for these are all negative at the time of writing. Using the model in Table 2 and a 1% net discount rate in the liability function in Equation 11, the procedure described in ¶4.5 yields the mis-estimation capital requirements shown in Figure 5. The mis-estimation capital required increases rapidly with age, which emphasises the importance of allowing for the age-related distribution of liabilities. The choice of discount rate (or yield curve) is obviously also important, as demonstrated in Figure 6 where a lower effective discount rate leads to a higher mis-estimation capital requirement.

**6.11** A selection of the mis-estimation capital requirements from Figure 5 is listed in Table 3. The average age of the 12,720 survivors at 1<sup>st</sup> January 2013 was 71.9 years (71.2 years weighted by annualised pension). Using a single model point would therefore suggest a mis-estimation capital requirement somewhere in the interval (4.90%, 5.22%).

Table 3: Specimen mis-estimation capital requirements from Figure 5 (95% confidence intervals using the method of Harrell and Davis (1982), calculated from 10,000 valuations of a single annuity discounted at a net rate of 1% p.a.).

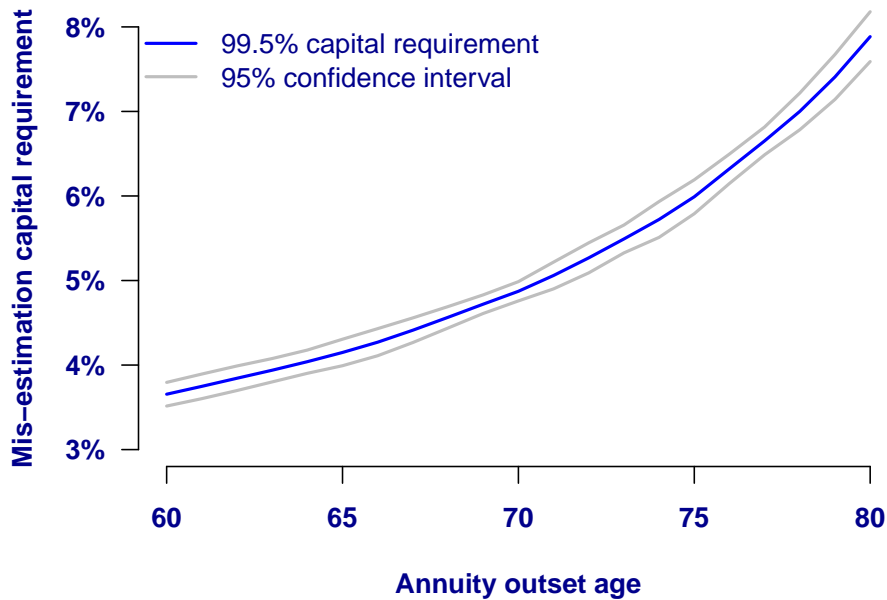
| Annuity<br>outset<br>age | Mis-estimation<br>capital<br>requirement |
|--------------------------|--|
| 69                       | 4.61–4.83%                               |
| 70                       | 4.76–4.99%                               |
| 71                       | 4.90–5.22%                               |
| 72                       | 5.09–5.45%                               |
| 73                       | 5.33–5.66%                               |

Figure 4:  $\log_e(\text{crude mortality hazard})$  with best-estimate fit and alternative line with stressed intercept ( $\alpha_0$ ) and re-estimated age slope ( $\beta_0$ ). Ages 60 and over, data and model from Table 2



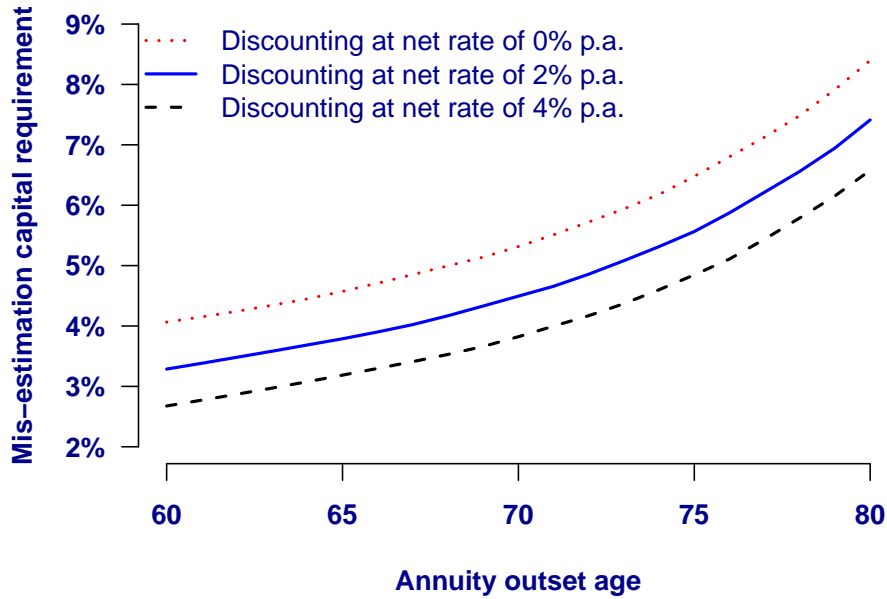
Note: The crude mortality hazard is the actual number of deaths in the age interval  $[x, x+1)$  divided by the time lived in that interval.

Figure 5: Mis-estimation risk capital requirement at 99.5% level as percentage of best-estimate reserve (with 95% confidence interval using Harrell and Davis (1982) estimate).



Note: The liability is a single level lifetime annuity from each stated age at outset, valued as in Equation 11. Estimated from 10,000 simulations using model from Table 2 and discounting a level annuity at 1% per annum.

Figure 6: Mis-estimation risk capital requirement at 99.5% level as percentage of best-estimate reserve under various discount rates.



Note: The liability is a single level lifetime annuity from each stated age at outset, valued as in Equation 11. Estimated from 10,000 simulations using model from Table 2 with varying discount rates.

**6.12** However, it is doubtful whether a single model point, or even a handful of model points, can capture the impact of mis-estimation risk on an entire portfolio. It would obviously be more accurate to perform an entire portfolio valuation instead. This is relatively straightforward to do, and we therefore redefine the valuation function as in Equation 12:

$$a(\boldsymbol{\theta}) = \sum_{i=1}^n w_i \bar{a}_{x_i} \quad (12)$$

for all  $n$  survivors in the pension scheme. This automatically allows for not only the age distribution via the individual ages,  $x_i$ , but also the liability distribution via the individual pensions,  $w_i$ . Using the full-portfolio valuation function in Equation 12 and the procedure in ¶4.5 we get a 95% confidence interval for the mis-estimation capital of (4.57%, 4.90%). We can see that the range of mis-estimation capital requirements using the full-portfolio valuation is lower than would be implied by using a specimen model point at the average age and looking up Table 3. To show the impact of the distribution of the  $w_i$ , setting all the pensions to the same value and redoing the full-portfolio valuation with ¶4.5 yields a 95% confidence interval for the mis-estimation capital range of (4.64%, 5.00%). A full-portfolio valuation of individual liabilities yields a more accurate value for the mis-estimation capital requirement, which in this case also meant a lower requirement.

**6.13** However, the model in Table 2 is implausibly simple: it does not contain enough relevant risk factors and only uses one year's experience when more data are available. In Section 8 we will fit models with more risk factors and consider other mortality laws, but we first need to consider the impact of using multi-year data.

## 7 The impact of using multi-year data

**7.1** The mis-estimation capital requirements in Figure 5 and Table 3 are based on only a single year's mortality experience. In practice, most portfolios have experience data spanning several years, and this extra data contributes to improved estimation of parameters. This extra data also contributes to significantly reduced mis-estimation risk, as illustrated by comparing the first two rows in Table 4.

Table 4: 99.5% mis-estimation capital requirements as percentage of best-estimate reserve.

| Data range | Model    | Mis-estimation capital: |                  |
|------------|----------|-------------------------|------------------|
|            |          | (a) Central estimate    | (b) 95% interval |
| 2012 only  | Age      | 4.73%                   | 4.57–4.90%       |
| 2007–2012  | Age      | 2.12%                   | 2.06–2.19%       |
| 2007–2012  | Age+Time | 3.97%                   | 3.83–4.10%       |

Note: Results for various models and time periods calculated from 10,000 simulations. The 95% confidence intervals were calculated using the method of Harrell and Davis (1982). Models calibrated using data for ages 60 and over. For models including a time trend, the rates fitted are for 1<sup>st</sup> January 2010, i.e.  $\mu_{x,2010}$ .

**7.2** However, an implicit assumption in the simple age-only model is that mortality is a stationary process in time. As many studies have shown, such as Willets (1999), mortality rates of pensioners and annuitants have been continuously falling for decades. We can allow for the fact that mortality is a moving target by including a time-trend parameter,  $\delta$ , as in Equation 13:

$$\mu_{x,y} = e^{\alpha_0 + \beta_0 x_i + \delta(y-2000)} \quad (13)$$

where  $\mu_{x,y}$  is the instantaneous force of mortality at exact age  $x$  and calendar time  $y$ . The offset of -2000 keeps the parameters well-scaled. The  $\delta$  parameter could obviously be used to project mortality rates into the future as well as measuring the recent changes in mortality levels. However, we are concerned in this paper with mis-estimation risk of *current* rates, so to keep things comparable in Table 4 we generate static mortality rates as at 1<sup>st</sup> January 2010, i.e. using Equation 13 the fitted rates used in the mis-estimation assessment will be  $\mu_{x,2010}$ .

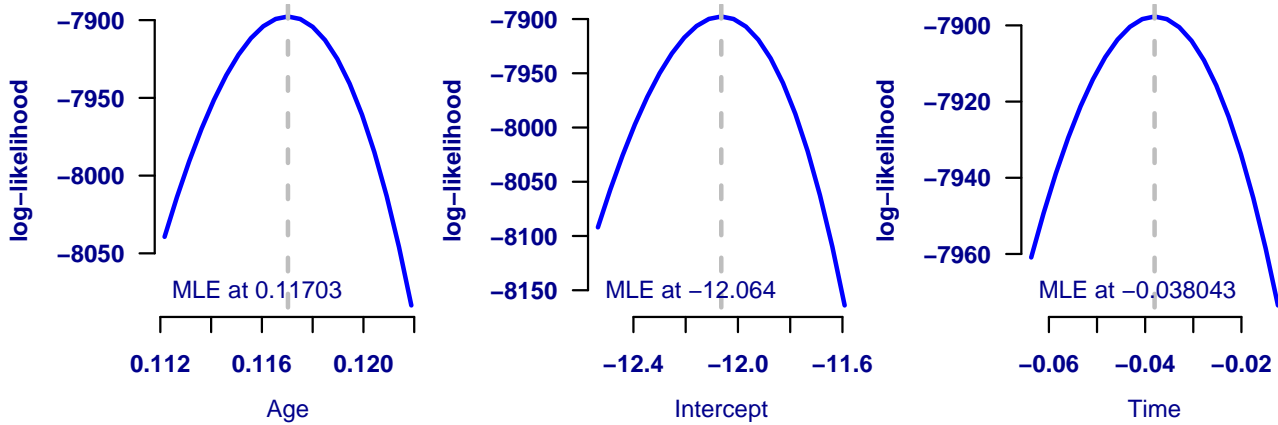
**7.3** One consequence of including this time-trend parameter is that it offsets a large part of the benefit of having the extra data, as shown in the third row of Table 4. This is an example of a phenomenon which occurs repeatedly with mis-estimation risk: enhancing the model’s fit can lead to an increase in mis-estimation risk. In the case of Table 4 there is little doubt that the model including time trend is a better fit: the AIC (Akaike, 1987) is 15801.3 with the time trend and 15808.1 without it, while the log-likelihood profiles in Figure 7 confirm the validity of including each parameter as none of their plausible ranges includes zero. However, the better-fitting of the two models results in a higher mis-estimation capital requirement. Indeed, the inclusion of the time-trend parameter has added back over two thirds of the reduction in mis-estimation capital requirements from the extra data.

**7.4** Since this phenomenon might strike some readers as counter-intuitive, it is worth explaining it in more detail. At its heart we have the same see-saw phenomenon as in Figure 4, but here operating in calendar time. The time-trend parameter,  $\delta$ , will be forced to pass through the “average” level of mortality across the period of observation, but it will be much less constrained at either end. This was the case in Figure 4 when looking at the behaviour of  $\beta_0$  in response to changes in  $\alpha_0$ . Thus, there is more uncertainty over the level of mortality at either end of the exposure period than there is over the mid-point of the period. In the presence of a time-varying mortality process, it is this see-saw effect which is offsetting much of the extra estimating power from using multi-year data.

**7.5** This can be thought of more figuratively. Leaving aside the fact that mortality increases with age, if mortality were a static process in time then more experience data over a longer period of time will validly reduce mis-estimation risk. If there is no time trend, then the experience data of five years ago (say) will help inform you about mortality levels now: there is more relevant data on a process which is not changing in time. This is illustrated by comparing the first two rows of Table 4. However, if there is a time trend, then the experience data of five years ago is of less use in estimating current levels of mortality. While the time trend can be estimated, there is still uncertainty over its precise value, and it is this uncertainty which is

undoing some of the benefit of the multi-year data. This is illustrated by comparing the last two rows of Table 4. Thus, where risk levels vary in time, as with pensioner and annuitant mortality, experience data spanning multiple years will only yield modest reductions in mis-estimation risk. This is illustrated by comparing the first and last rows of Table 4. Other necessary improvements to the model will also increase the mis-estimation capital requirements, but for quite different reasons, and we will see why this is so in Section 8.

Figure 7: Log-likelihood profiles for Age+Time model in Table 4. These profiles demonstrate the quadratic shape around the MLEs, which is consistent with the multivariate Normal distribution for the estimates in Equation 7.



## 8 A minimally acceptable model for financial purposes

**8.1** So far we have considered a simple two-parameter Gompertz model, which works well enough for ages above age 60. However, Figure 9 shows an important amount of unused experience data in the age range 50–60, while Figure 10 shows that the data in this range will not fit the straight-line Gompertz assumption on a logarithmic scale. We solve this problem by using the mortality law in Equation 14, which includes a constant component as per Makeham (1860):

$$\mu_{x_i} = \frac{e^\epsilon + e^{\alpha_0 + \beta_0 x_i}}{1 + e^{\alpha_0 + \beta_0 x_i}} \quad (14)$$

As before, we can also define a time-varying version of Equation 14:

$$\mu_{x_i, y} = \frac{e^\epsilon + e^{\alpha_0 + \beta_0 x_i + \delta(y-2000)}}{1 + e^{\alpha_0 + \beta_0 x_i + \delta(y-2000)}} \quad (15)$$

**8.2** The models in Equations 14 and 15 do not allow for many risk factors, however. We can extend our model's capabilities as follows:

$$\mu_{x_i, y} = \frac{e^\epsilon + e^{\alpha_0 + \sum_j \alpha_j I_{i,j} + \beta_0 x_i + \delta(y-2000)}}{1 + e^{\alpha_0 + \sum_j \alpha_j I_{i,j} + \beta_0 x_i + \delta(y-2000)}} \quad (16)$$

where  $\alpha_j$  represents the effect of risk factor  $j$  and  $I_{i,j}$  is an indicator function taking the value 1 when life  $i$  possesses risk factor  $j$  and 0 otherwise. In this section we will consider gender and pension size-band as additional risk factors, but there is no practical limit to the number of risk factors which can be considered when modelling individual-level mortality. The benefits of individual- over group-level modelling are discussed in Richards et al. (2013).

**8.3** For a model to be useful for actuarial purposes we additionally require that all financially significant risk factors are included. To test whether a model achieves this we use the process of bootstrapping described in Richards et al. (2013), i.e. we sample from the data and calculate the ratio of the actual number of deaths in the sample to the predicted number according to the model. The results of this are shown for two models in Table 5, where the ratios for the lives column are close to 100%. This means that both models have a good record in predicting the number of deaths, as one would expect for models fitted by the method of maximum likelihood. However, we can see that the model in the first row has a poor record when the ratio is weighted by pension size. This is because those with larger pensions have a lower mortality rate, thus leading the first model to over-state pension-weighted mortality. The first model in Table 5 is therefore unacceptable for financial purposes.

Table 5: Bootstrapping results for two alternative models.

| Model                        | Median bootstrap ratio weighted by: |              | 95% interval for mis-estimation risk capital |
|------------------------------|-------------------------------------|--------------|--|
|                              | (i) Lives                           | (ii) Amounts |  |
| Age+Gender+Makeham+Time      | 99.8%                               | 91.7%        | 3.75–3.96%                                   |
| Age+Gender+Makeham+Time+Size | 99.8%                               | 98.8%        | 4.40–4.68%                                   |

Note: Calibrated to 2007–2012 mortality experience of pension scheme described in Appendix 1 (ages 50 and over). Bootstrap sample size is 1,000 (sampling with replacement) and 1,000 samples are taken. Mortality rates are as at 1<sup>st</sup> January 2013, i.e. we are using  $\mu_{x,2013}$  without projection. Model notation follows that of Baker and Nelder (1978).

**8.4** The second row in Table 5 shows that the addition of a three-level risk factor for pension size makes a material improvement in the amounts-weighted bootstrap ratio. Although the median bootstrap ratio is not quite as close to 100% as we might like, it is a clear improvement on the first model and could therefore be regarded as a minimally acceptable model for financial purposes. We can also see that the inclusion of a necessary risk factor (pension size) has increased the capital requirement for mis-estimation risk. The reason for this is the concentration of risk demonstrated in Table 11 — the bulk of the total pension is paid to a small proportion of the overall lives, yet these same lives experience lower mortality rates, as shown in Table 6. In effect, the liability of the scheme is driven by a much smaller effective number of lives than a simple headcount would imply, and the mortality level of these lives is lower. This concentration of risk makes it very important that (i) the model reflects these dynamics, and (ii) that the mis-estimation valuation function  $a(\boldsymbol{\theta})$  reflects the impact of these lives. Using a valuation function which performs a valuation of the whole portfolio will do this.

Table 6: Summary of Age+Gender+Makeham+Time+Size model using Equation 16.

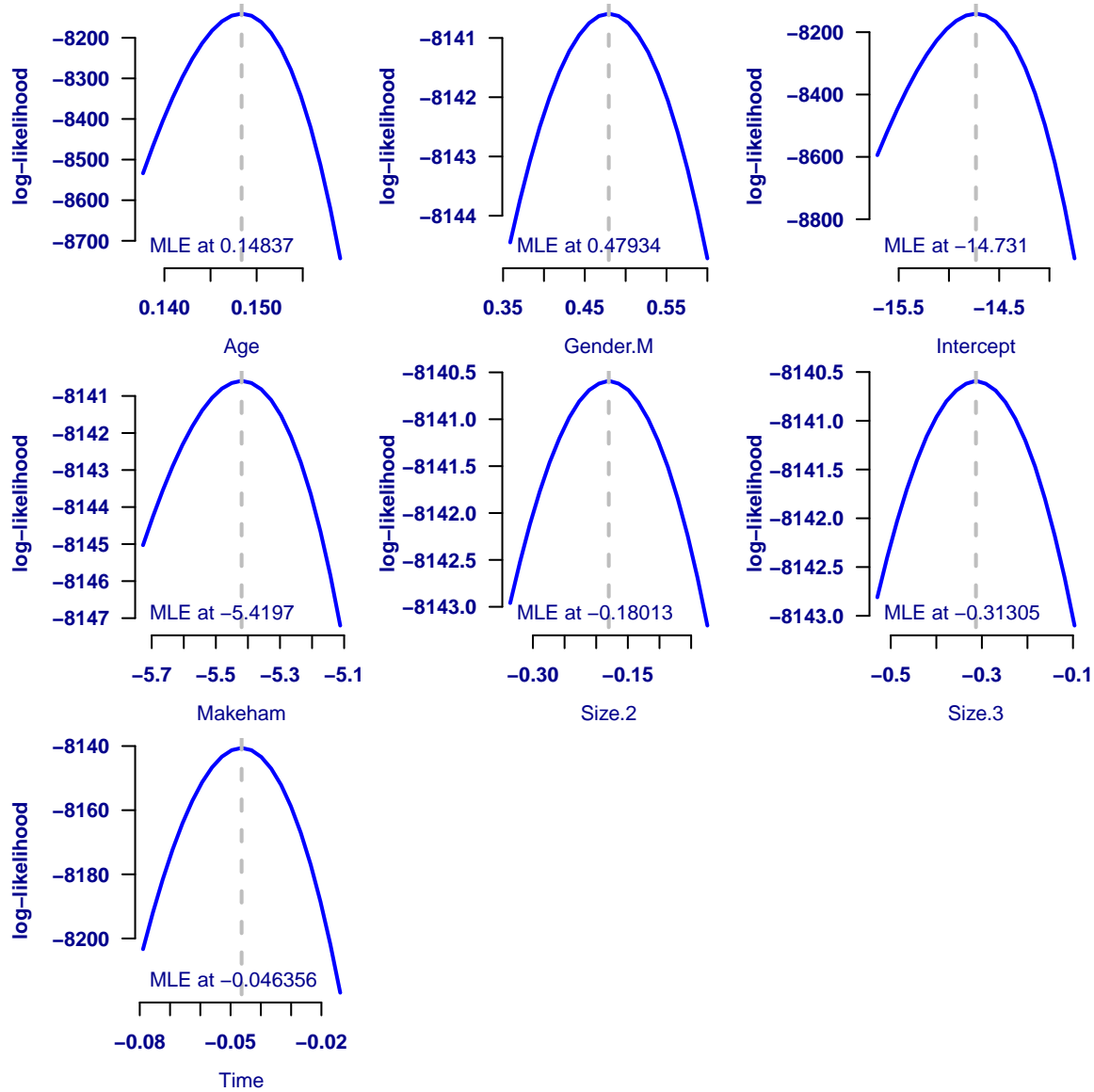
| Parameter                           | Estimate   | Std. error | Z-value | Pr(> z )     | Sig. | Lives  | Deaths |
|-------------------------------------|------------|------------|---------|--------------|------|--------|--------|
| Age ( $\beta_0$ )                   | 0.148372   | 0.00534602 | 27.75   | 0            | ***  | 15,698 | 2,076  |
| Gender.M ( $\alpha_{\text{male}}$ ) | 0.479342   | 0.0603327  | 7.94    | 1.94235e-015 | ***  | 5,956  | 901    |
| Intercept ( $\alpha_0$ )            | -14.7313   | 0.490689   | -30.02  | 0            | ***  | 15,698 | 2,076  |
| Makeham ( $\epsilon$ )              | -5.41968   | 0.15372    | -35.26  | 0            | ***  | 15,698 | 2,076  |
| Size.2 ( $\alpha_{\text{Size.2}}$ ) | -0.180133  | 0.0778913  | -2.31   | 0.0207435    | *    | 3,140  | 362    |
| Size.3 ( $\alpha_{\text{Size.3}}$ ) | -0.313053  | 0.108229   | -2.89   | 0.00382173   | **   | 1,567  | 164    |
| Time ( $\delta$ )                   | -0.0463558 | 0.0162836  | -2.85   | 0.00441649   | **   | 15,698 | 2,076  |

Note: Calibrated using experience data for ages 50 and over. Parameter significance is labelled according to the same scheme used in R [R Core Team (2012)], i.e. . for 10%, \* for 5%, \*\* for 1% and \*\*\* for 0.1%.

**8.5** In practice an actuary would use a more sophisticated model than the one in Table 6. Risk factors like gender and pension size would normally be allowed to vary by age, for example, and most

portfolios have scope for additional rating factors such as early-retirement indicators or postcode-based profiling — see Richards (2008) for further details.

Figure 8: Log-likelihood function in profile for each of the seven parameters in Table 6, showing the essentially quadratic nature of the function, and thus the validity of the multivariate normal assumption in ¶5.3.



## 9 Discussion

**9.1** Before using the mis-estimation method described in this paper, it is important that the analyst check three items. The first requirement is that the independence assumption must hold: data must be deduplicated and the model must be well specified. Data with duplicates will give a falsely low estimate of mis-estimation risk, but a bigger risk comes from badly specified models. One example is the still-encountered practice of chopping up the experience of individuals into non-overlapping annual pieces and fitting a  $q_x$  model with a GLM. Such models will understate misestimation risk capital if applied in the manner described in this paper.

**9.2** The second requirement is that the model allows for all financially significant risk factors. A model which does not contain these will give an erroneously low estimate of mis-estimation risk. This is because the financial impact of each life is not the same — those with larger pensions have a much



larger influence than those with smaller pensions. This is compounded by the fact that those with larger pensions usually live longer, as shown by the lower mortality of the Size.3 group in Table 6. Thus, both the overall reserves and the associated mis-estimation risk are largely driven by a small proportion of the portfolio membership. The bootstrapping procedure of ¶8.3 can test whether a model has included all financially relevant risk factors.

**9.3** The third requirement for multi-year data is to check that the model includes a time-trend parameter, if one is needed. An additional useful check is that the log-likelihood function is approximately quadratic around each maximum-likelihood estimate, i.e. that the multivariate Normal assumption behind Formula 7 holds.

**9.4** Mis-estimation risk is portfolio-dependent. Different portfolios, and models with different risk factors, will produce different mis-estimation capital requirements. By and large, the more experience data there is, the lower the mis-estimation risk will be. We can illustrate this by using a large population of German pensioners. A comparison of the scale of the data sets is given in Table 7, where the German data set is a factor of around fifteen times larger.

Table 7: Comparison of U.K. pension scheme from Appendix 1 and a German pensioner population.

| Data set                                      | Years     | Lives   | Deaths | Time lived (years) |
|---|-----------|---------|--------|--------------------|
| U.K. pension scheme from Appendix 1           | 2007–2012 | 15,698  | 2,076  | 71,162.9           |
| German pensioners from Richards et al. (2013) | 2007–2011 | 244,908 | 31,546 | 1,008,739.0        |
| Ratio of German : U.K. metric                 | n/a       | 15.6    | 15.2   | 14.2               |

**9.5** A comparison of Tables 5 and 8 shows that the same minimally acceptable model does a reasonable — but not perfect — job of explaining financially weighted mortality variation in the U.K. scheme and amongst the German pensioners. However, although the German data set is around fifteen times larger, the mis-estimation capital requirement is only reduced by three quarters.

Table 8: Bootstrapping results for German pensioner data.

| Model                        | Median bootstrap ratio weighted by:<br>(i) Lives (ii) Amounts | 95% interval for mis-estimation risk capital |
|------------------------------|---|--|
| Age+Gender+Makeham+Time+Size | 99.4%   | 98.3%<br>1.10–1.19%                          |

Note: Calibrated to 2007–2011 mortality experience of pension scheme described in Richards et al. (2013) (ages 50 and over). Bootstrap sample size is 1,000 (sampling with replacement) and 1,000 samples are taken. Mortality rates are as at 1<sup>st</sup> January 2012, i.e. we are using  $\mu_{x,2012}$  without projection. Model notation follows that of Baker and Nelder (1978).

**9.6** More data can come from the portfolio being larger or from a longer exposure period, although the benefit of multi-year data is tempered when a time trend exists. The mis-estimation capital requirements also vary according to the discount rate or yield curve, as shown in Figure 6, so such calculations will need to be regularly updated as the interest-rate environment changes.

**9.7** The capital requirements produced by this method should also be taken as a lower bound, and actuarial judgement will need to be applied as to the extent of any additional allowances which might be necessary. For example, if a portfolio were growing quickly, a larger addition would be needed than if the portfolio were growing slowly. A still larger adjustment might be required if the portfolio were growing due to a new business source. For example, consider a life-office that historically wrote internal vesting annuities, but then made a decision to write open-market annuities or bulk-purchase annuities. This would create extra mis-estimation risk which could not be captured in a procedure calibrated using

only the internal-vesting mortality experience. Such circumstances would require additional capital to be held for mis-estimation risk, and this can only be determined by actuarial judgment.

**9.8** We note also that a matrix of correlations between parameters,  $C$ , can be derived from the variance-covariance matrix,  $V$ , by setting  $c_{i,j} = v_{i,j} / \sqrt{v_{i,i}v_{j,j}}$ , where  $c_{i,j}$  and  $v_{i,j}$  are the values in row  $i$  and column  $j$  of the matrices  $C$  and  $V$ , respectively.  $C$  is symmetric around the leading diagonal, as is  $V$ , and the leading diagonal of  $C$  is composed of 1s, as each parameter is perfectly correlated with itself. All entries in  $C$  lie between -1 and +1 (which represent perfect negative and perfect positive correlation, respectively). An example correlation matrix is given in Table 9, calculated according to the steps outlined above.

Table 9: Percentage correlations between coefficients in the model from Table 6, i.e.  $c_{i,j} \times 100\%$ .

|           | Age | Gender.M | Intercept | Makeham | Size.2 | Size.3 | Time |
|-----------|-----|----------|-----------|---------|--------|--------|------|
| Age       | 100 | 23       | -94       | 72      | -7     | -2     | -2   |
| Gender.M  | 23  | 100      | -26       | 17      | -17    | -19    | 0    |
| Intercept | -94 | -26      | 100       | -70     | 5      | 2      | -32  |
| Makeham   | 72  | 17       | -70       | 100     | -10    | -2     | -1   |
| Size.2    | -7  | -17      | 5         | -10     | 100    | 13     | -1   |
| Size.3    | -2  | -19      | 2         | -2      | 13     | 100    | 0    |
| Time      | -2  | 0        | -32       | -1      | -1     | 0      | 100  |

Note: As with the variance-covariance matrix, the correlation matrix is symmetric about the leading diagonal. The leading diagonal is 100% because a parameter value is perfectly correlated with itself.

**9.9** One item worth noting in Table 9 is that the time-trend parameter has little correlation with any parameter apart from the Intercept, with which it is negatively correlated. This is different from the model for the German pensioner data set in Richards et al. (2013), where the time-trend parameter was not strongly correlated with any of the other parameters. This has potential application to the correlation matrix used by insurers when allowing for diversification of risks under the Solvency II regime. The correlation between the time-trend parameter and the others could be used to support the assumed correlation between mis-estimation risk and mortality improvements. This would have to be assessed on a portfolio-by-portfolio basis, however, and would require actuarial judgement. In the case of this portfolio, however, the negative correlation between the time trend and the intercept could perhaps be used to justify a negative correlation between mis-estimation risk and trend risk. If so, this would bring an overall indirect capital benefit from using multi-year data, despite the modest reduction in mis-estimation capital in Section 7.

## 10 Conclusions

**10.1** Mis-estimation risk for a portfolio can be straightforwardly assessed using the portfolio's own experience data and some basic results for maximum-likelihood estimates. An approximation for the variance-covariance matrix is required, but this can be quickly derived from the log-likelihood function for any statistical model. Increased portfolio size leads to better estimation and thus lower mis-estimation capital requirements, but Section 9 shows that there is a diminishing return. Similarly, experience data spanning multiple years provide only a modest reduction in mis-estimation capital requirements where risk rates are changing in time, as shown in Section 7. However, larger portfolio size or a longer exposure period could bring indirect capital benefits from examining the matrix of parameter correlations, as in ¶9.9.

**10.2** Parameters estimated from any statistical model are correlated to some extent, and these correlations need to be acknowledged in assessing mis-estimation risk. Furthermore, uncertainty over fitted rates varies by age and some of the greatest parameter uncertainty applies to the lives with the largest concentration of liabilities. The concentration of liabilities in a small subset of lives is one reason why an improved model fit can lead to higher mis-estimation risk. A full-portfolio valuation using appropriately perturbed model parameters will therefore allow for all of these aspects when assessing mis-estimation risk.

## References

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika* 52, 317–333.
- Baker, R. J. and J. A. Nelder (1978). *The GLIM System Release 3 Manual*. Numerical Algorithms Group, Oxford.
- Cairns, A. J. G., D. Blake, and K. Dowd (2006). A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance* 73, 687–718.
- CMI Ltd (2014). *Graduations of the CMI SAPS 2004-2011 mortality experience based on data collected by 30 June 2012 — Final “S2” Series of Mortality Tables*. CMI Ltd.
- Cox, D. R. and D. V. Hinkley (1996). *Theoretical Statistics*. Chapman and Hall.
- Gompertz, B. (1825). The nature of the function expressive of the law of human mortality. *Philosophical Transactions of the Royal Society* 115, 513–585.
- Hardy, M. R. and H. H. Panjer (1998). A credibility approach to mortality risk. *ASTIN Bulletin* 28 (2), 269–283.
- Harrell, F. E. and C. E. Davis (1982). A new distribution-free quantile estimator. *Biometrika* 69, 635–640.
- Kaplan, E. L. and P. Meier (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53, 457–481.
- Lindgren, B. W. (1976). *Statistical Theory* (third ed.). Collier Macmillan.
- Lindley, D. V. and W. F. Scott (1984). *New Cambridge Elementary Statistical tables*. Cambridge University Press.
- Longevity Development Team (2014). *Longevity v2.7.1 User Guide*. Longevity Ltd, Edinburgh, United Kingdom.
- Makeham, W. M. (1860). On the law of mortality and the construction of annuity tables. *Journal of the Institute of Actuaries and Assurance Magazine* 8, 301–310.
- McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models* (second edition ed.), Volume 37 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London.
- R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Richards, S. J. (2008). Applying survival models to pensioner mortality data. *British Actuarial Journal* 14(II), 257–326 (with discussion).
- Richards, S. J. (2012). A handbook of parametric survival models for actuarial use. *Scandinavian Actuarial Journal* 2012 (4), 233–257.
- Richards, S. J., I. D. Currie, and G. P. Ritchie (2014). A value-at-risk framework for longevity trend risk. *British Actuarial Journal* 19 (1), 116–167.
- Richards, S. J., K. Kaufhold, and S. Rosenbusch (2013). Creating portfolio-specific mortality tables: a case study. *European Actuarial Journal* 3 (2), 295–319.
- Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S* (fourth ed.). Springer.
- Wetherill, G. B. (1982). *Elementary Statistical Methods* (third ed.). Chapman and Hall.
- Willets, R. C. (1999). Mortality in the next millennium. *Staple Inn Actuarial Society, London*.

## 11 Appendix 1: Details of scheme used to illustrate results

**11.1** The data used to illustrate results in this paper are for a medium-sized, local-authority pension scheme in England & Wales. The data fields available were as follows: date of birth, gender, commencement date, total annual pension, end date, postcode, National Insurance (NI) number and whether the pensioner was a child, retiree or widow(er) (C, R or W). The end date was determined differently for deaths, temporary pensions and survivors to the extract date. For deaths, the end date was the date of death. For children’s pensions and trivial commutations, the end date was the date the pension ceased or was commuted. For other survivors, the end date was the date of extract in early 2013. The experience data beyond 31<sup>st</sup> December 2012 were not used to avoid bias from delays in death reporting.

**11.2** There were 17,068 benefit records available before deduplication, of which one was rejected for having an end date inconsistent with the commencement date. Of the remaining 17,067 records, 2,265 were marked as deaths.

**11.3** Annuitants and pensioners often have multiple benefit records. It is particularly common in annuity portfolios for people to have multiple annuities, as demonstrated in Richards (2008). The phenomenon is less common for pension schemes, but multiple benefit records for the same individual can still arise. The first scenario is where an individual accrues two or more benefits from separate periods of service. The second scenario is where an individual receives a pension in respect of their own service and also a spouse’s pension if they were a widow(er) of a deceased pensioner in the same scheme. Administratively it is usually easier to handle these multiple benefits separately, even though they are paid to the same person, and so duplicate records arise.

**11.4** It is essential in any statistical model that the assumption of independence is valid, so we must perform deduplication, i.e. the identification of individuals with multiple benefit records. Following Richards (2008) we use two different deduplication schemes based on matching individual data items on each record. Each matching rule forms a deduplication key based on verified data items; if all items of the deduplication key match, then the two or more matching records are merged and the pension amounts added together. The details of this procedure are summarised in Table 10. The two records with conflicting life statuses were rejected, leaving 16,131 records ( $16,131 = 17,068 - 1 - 934 - 2$ ).

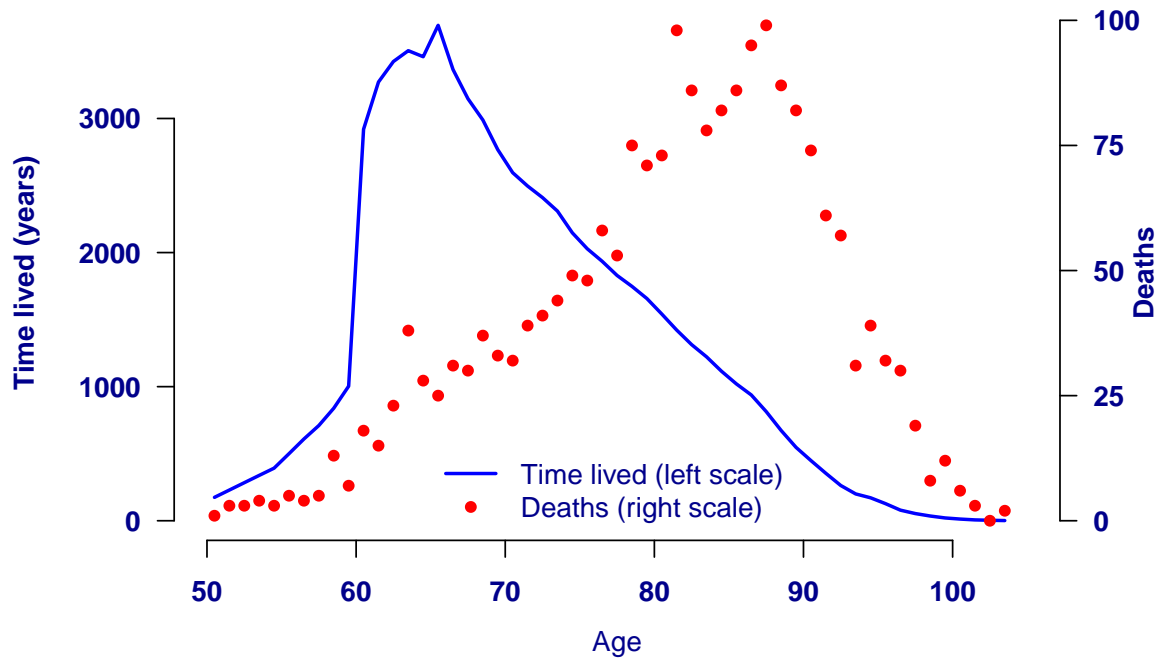
Table 10: Deduplication results for medium-sized pension scheme in England and Wales.

| Deduplication key                   | Duplicates eliminated | Conflicting statuses |
|-------------------------------------|-----------------------|----------------------|
| Date of birth, gender and postcode  | 891                   | 2                    |
| Date of birth, gender and NI number | 43                    | 0                    |
| Total                               | 934                   | 2                    |

**11.5** The number of duplicates may seem modest in relation to the overall number of records, but it is important to deduplicate for a number of reasons. For example, if an individual received two pensions of £4,000 each, this should be recognised as one individual in group S09 in Table 11 and not two records in group S07. Failure to deduplicate leads to bias in mortality models and to falsely comforting estimates of parameter variance, and thus under-statement of mis-estimation risk. The resulting data volumes are shown in Figure 9.

**11.6** One issue with pension schemes is that pensions are usually increased from year to year. This creates a bias problem for cases which terminate early, i.e. deaths and temporary pensions. To put all pension values on the same approximate financial footing, therefore, the annual pension amounts for early terminations were revalued by 2.5% per annum to the end of the period of observation (the Retail Prices Index RPIJ increased by a geometric average of 2.68% over this period, while the Consumer Price Index CPIH increased by 2.90%). A more accurate approach would have been to establish the actual

Figure 9: Deaths and time lived at ages 50 and over for 2007–2012, males and females combined.



Note: In the given age and date range there were 2,076 deaths among 15,698 lives, with a total exposure time of 71,162.93 years. Exposure time and deaths prior to 1<sup>st</sup> January 2007 were not used and exposure times and deaths after 31<sup>st</sup> December 2012 were discarded to reduce the influence of delays in death reporting.

scheme increases over the period, together with their timing and split between different types of benefits. However, this level of detail would not have made any material change to the results in this paper.

**11.7** Table 11 shows the breakdown of the deduplicated data by revalued annual pension. The pension scheme shows considerable concentration of risk, as the top 20% of lives account for 58.7% of the total pension. This phenomenon is common in the U.K., as demonstrated for other portfolios in Richards (2008).

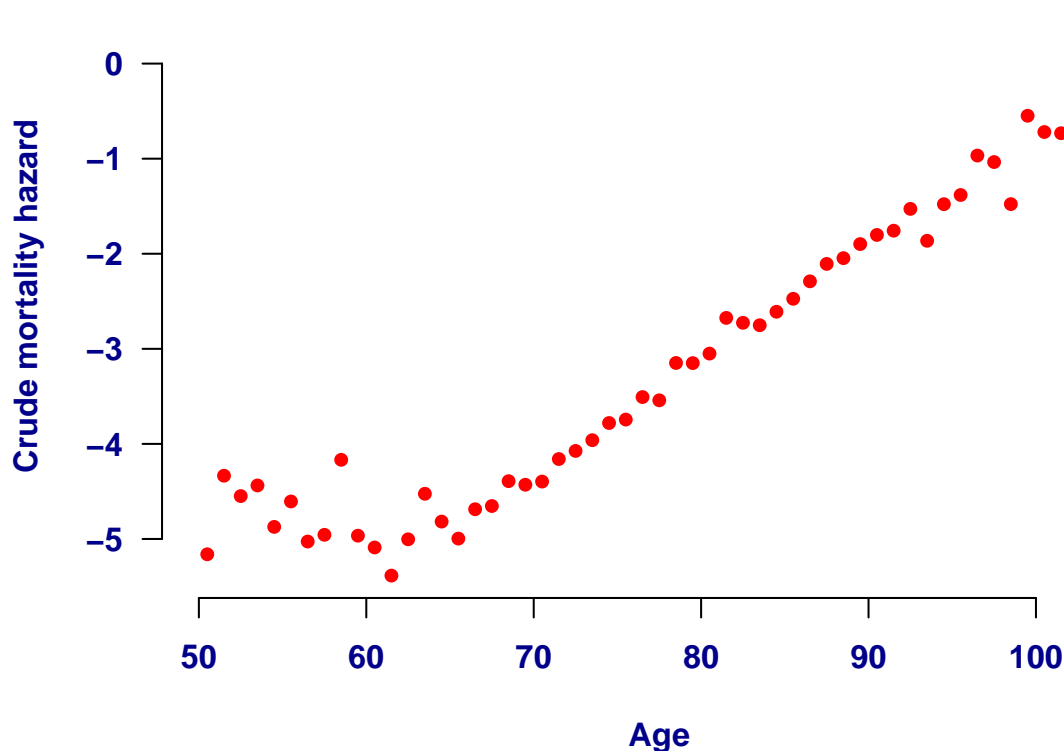
Table 11: Data by pension size-band.

| SizeBand | Revalued pension p.a. (£) |           | Lives  | Deaths | Exposure (years) | Pensions (£million) | Percentage of total scheme pension |
|----------|---------------------------|-----------|--------|--------|------------------|---------------------|------------------------------------|
|          | From...                   | ...to     |        |        |                  |                     |                                    |
| S01      | 0                         | 537.62    | 1,571  | 175    | 6,766.9          | 0.5                 | 0.7                                |
| S02      | 537.62                    | 954.22    | 1,570  | 235    | 6,989.4          | 1.2                 | 1.6                                |
| S03      | 954.22                    | 1,397.90  | 1,570  | 238    | 6,902.9          | 1.8                 | 2.5                                |
| S04      | 1,397.90                  | 1,962.39  | 1,570  | 259    | 7,110.7          | 2.6                 | 3.5                                |
| S05      | 1,962.39                  | 2,686.70  | 1,570  | 227    | 7,066.5          | 3.6                 | 4.9                                |
| S06      | 2,686.70                  | 3,616.08  | 1,570  | 213    | 7,083.6          | 4.9                 | 6.6                                |
| S07      | 3,616.08                  | 4,958.15  | 1,570  | 203    | 7,265.7          | 6.7                 | 9.0                                |
| S08      | 4,958.15                  | 7,025.51  | 1,570  | 178    | 7,291.6          | 9.3                 | 12.5                               |
| S09      | 7,025.51                  | 11,564.99 | 1,570  | 184    | 7,287.2          | 14.0                | 18.9                               |
| S10      | 11,564.99                 | 75,628.38 | 1,567  | 164    | 7,398.5          | 29.5                | 39.8                               |
| Total    |                           |           | 15,698 | 2,076  | 71,162.9         | 74.1                | 100.0                              |

Note: Pensions for deaths and early terminations are revalued at 2.5% per annum to the end of 2012.

**11.8** Figure 10 shows the crude mortality hazard on a logarithmic scale for males and females combined. There is no obvious evidence of a data problem, with the mortality hazard approximately log-linear above age 60. A further useful check on the validity of the data is to plot the Kaplan-Meier curves (Kaplan and Meier, 1958). We have encountered other portfolios where there is not a clear separation between survival curves for males and females. This is sometimes evidence of a data-corruption problem, which can be related to the processing of benefits for surviving spouses. However, Figure 11 shows that females have a consistently higher probability of reaching any given age, leading us to conclude that the data here do not suffer from any obvious corruption. Table 12 contains the mortality ratios for the deduplicated scheme data, which suggest falling mortality levels with a large degree of volatility in the annual experience.

Figure 10: Crude mortality hazard for males and females combined in medium-sized U.K. pension scheme.



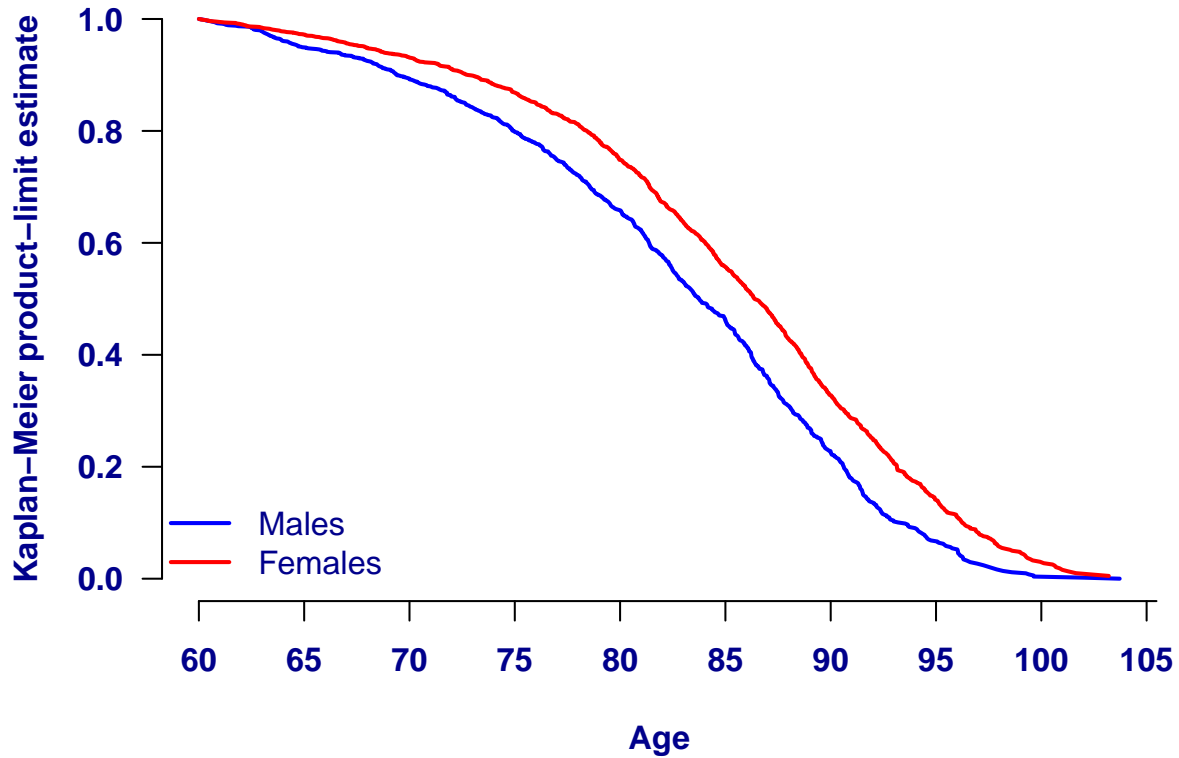
Note: Data cover ages 50 and over for the period 2007–2012. The crude mortality hazard is the actual number of deaths in the age interval  $[x, x+1)$  divided by the time lived in that interval. The falling mortality rate over the age range 50–60 is caused by excess mortality from ill-health retirements (the gender mix in this age range is broadly constant, with females accounting for variable 42–53% of exposure at individual ages).

Table 12: Actual deaths against S2PA without adjustment, weighted by lives and revalued pension.

| Weighting | Gender  | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|-----------|---------|------|------|------|------|------|------|
| Lives     | Males   | 112% | 119% | 120% | 102% | 108% | 104% |
|           | Females | 125% | 112% | 98%  | 101% | 88%  | 100% |
| Amounts   | Males   | 93%  | 92%  | 113% | 82%  | 92%  | 88%  |
|           | Females | 110% | 97%  | 83%  | 100% | 79%  | 89%  |

Note: The denominator of the mortality ratio is calculated by summing  $\mu_{x+\frac{1}{2}} E_{x+\frac{1}{2}}^c$ , where  $E_{x+\frac{1}{2}}^c$  is the central exposed-to-risk or time lived at average age  $x + \frac{1}{2}$ .  $\mu_{x+\frac{1}{2}}$  is approximated from the published table by  $-\log(1-q_x)$ . Pensions for deaths and early terminations are revalued at 2.5% per annum to the end of 2012.

Figure 11: Kaplan-Meier product-limit estimate from age 60 for males and females in medium-sized U.K. pension scheme.



Note: The data cover ages 60 and over in the period 2007–2012. The version of the Kaplan-Meier estimate plotted is that defined in Richards (2012).

## 12 Appendix 2: Analytical derivatives v. numerical approximations

**12.1** When fitting mortality models, the foundation of modern statistical inference is the log-likelihood function,  $\ell$ . The point at which the log-likelihood has its maximum value gives you the joint maximum-likelihood estimates of your parameters, while the curvature of the log-likelihood gives information about the uncertainty of those parameter estimates. The key to both is the calculation of derivatives: gradients (first derivatives) for maximizing the log-likelihood function and curvature (second partial derivatives) for estimating the variance-covariance matrix. In each case, we require either the analytical derivatives themselves, or else numerical approximations of them.

**12.2** Numerical approximations to derivatives can be obtained relatively straightforwardly by using difference quotients. This involves perturbing a parameter by a small value,  $h$  say, and then expressing the change in function value relative to  $h$ . For example, a central difference quotient of the single-parameter log-likelihood function,  $\ell(\theta)$ , will yield a numerical approximation of the first derivative, as in Equation 17:

$$\frac{\partial \ell}{\partial \theta} \approx \frac{\delta_h[\ell](\theta)}{h} = \frac{\ell(\theta+h/2) - \ell(\theta-h/2)}{h} \quad (17)$$

and a similar central difference quotient can be used to approximate the second derivative, as in Equation 18:

$$\frac{\partial^2 \ell}{\partial \theta^2} \approx \frac{\delta_h^2[\ell](\theta)}{h^2} = \frac{\ell(\theta+h) - 2\ell(\theta) + \ell(\theta-h)}{h^2} \quad (18)$$

**12.3** For the multi-parameter case, Equation 17 must be applied to each parameter in turn, with the other parameters unperturbed. A simple extension of Equation 18 is used for second partial derivatives. In theory, the smaller the value of  $h$ , the closer the numerical approximation will be to the actual derivative. Such numerical approximations have the advantage that they avoid the need to work out large numbers of analytical derivatives for each mortality law. However, there are drawbacks in terms of stability and accuracy, and this is most evident for estimating the curvature. By way of illustration, Table 13 shows the estimates and approximate standard errors for one of the parameters in a simple two-parameter Gompertz model using numerical approximations for the derivatives.

Table 13: Estimated value and approximate standard error for the Intercept ( $\hat{\alpha}_0$ ) parameter in the simple Gompertz model for mortality in Table 2. Numerical approximation is used with varying values of  $h$ .

| $h$       | Intercept<br>( $\hat{\alpha}_0$ ) | Standard<br>error |
|-----------|-----------------------------------|-------------------|
| $10^{-3}$ | -12.9491                          | 0.502297          |
| $10^{-4}$ | -12.9718                          | 0.467319          |
| $10^{-5}$ | -12.972                           | 0.459947          |
| $10^{-6}$ | -12.972                           | 0.193068          |
| $10^{-7}$ | -12.972                           | 0.020153          |

**12.4** The maximum-likelihood estimate of the Intercept parameter in Table 13 is consistent to five significant figures for values of  $h=10^{-4}$  and smaller, with a degree of convergence apparently achieved for values of  $h=10^{-5}$  and smaller. However, there is no such obvious convergence for the approximate standard error, with the estimate becoming unstable for very small values of  $h$ . This is not an isolated result — similar work with other schemes and other models produces the same finding. Furthermore, although the value of  $h=10^{-4}$  looks like the best compromise in Table 13, this is not the case for other models and other data sets. Our conclusion therefore is that while numerical approximations are often fine for finding the maximum-likelihood estimates, they are not reliable enough for estimating the variance-covariance matrix.

**12.5** For reliability one simply has to do the maths and find the expressions for the analytical derivatives. This has the additional advantage that points of possible arithmetic underflow can be identified and handled — see Richards (2012) for a list of some problematic expressions and how to cope with them. Another advantage is that using the analytical derivatives typically results in faster fitting times. Using analytical derivatives the values for the parameters in Table 13 are -12.972 for the Intercept and 0.467 for the standard error (see Table 2). From this we can deduce that  $h=10^{-4}$  was indeed the best-performing approximation step in Table 13. There is of course a degree of irony in having to calculate the analytical derivatives to find out which numerical approximation performs best — one might as well just use the analytical derivatives. However, it is useful to be able to do both as a means of checking one's mathematics.

**12.6** The methodology in this paper requires the accurate estimation of the variance-covariance matrix, which we estimate from the inverted matrix of second partial derivatives. The best way to accurately estimate the latter is to work out the analytical derivatives. This is admittedly more work than using difference quotients like Equations 17 and 18, but, as demonstrated in Table 13, using analytical derivatives is the most reliable way to fit the model parameters and to estimate their variance-covariance matrix.



## 13 Appendix 3: Number of simulations required

**13.1** In the main body of this paper we used 10,000 simulations of Equation 7, followed by full-portfolio valuations in Section 8. The simulations can be done quickly enough, but the portfolio valuations which accompany them can be computationally demanding. It is worth commenting on two practical methods to speed things up.

**13.2** The first method is to exploit the fact that each simulation in Equation 7 is independent of all the others, as are the associated portfolio valuations. We can therefore use parallel processing to perform the same work in a much shorter period of time. The mis-estimation algorithm in this paper is highly scaleable, i.e. spreading the work over  $u$  CPUs means the work can be done in  $\frac{1}{u}$  of the elapsed time. This was the method adopted for this paper, where calculations were spread over seven CPUs.

**13.3** The second method is to see whether there is an appropriate distributional assumption which can be made for the portfolio valuations in  $S$  (defined in Formula 6). The number of simulations could be curtailed if it were found that the values in  $S$  followed, say, a Normal distribution, as the mean and variance could be robustly estimated and all higher-order moments are a function of these. This approach was considered for Richards et al. (2014), where the much larger number of parameters in projection models — and the greater variance in the estimates — made an approach like Equation 7 trickier. However, the distribution of the values in the resulting set  $S$  in Richards et al. (2014) was only Normally distributed for a subset of projection models, so this second method could not be universally used. That said, the mis-estimation approach in Equation 7 is very different and the number of parameters is much smaller, so it is worth considering if this second method might apply here.

**13.4** The Normal distribution is fully characterised by its mean and variance, i.e. the first two moments of the distribution. Higher-order moments are merely a function of the first two. In general, if a distribution has mean  $\mu$ , then the coefficient of skewness (standardised third moment) is given by Formula 19:

$$\frac{E(s-\mu)^3}{[E(s-\mu)^2]^{\frac{3}{2}}} \quad (19)$$

where  $E$  denotes expectation — see Wetherill (1982, p88). Formula 19 is zero for the Normal distribution. Similarly, adjusting the coefficient of kurtosis (standardised fourth moment) given in Wetherill (1982, p88) we get Formula 20:

$$\frac{E(s-\mu)^4}{[E(s-\mu)^2]^2} \quad (20)$$

which is 3 for the Normal distribution. Thus, if the skewness coefficient of the data in  $S$  is close to zero and the kurtosis coefficient of the same data is close to 3, then it would be appropriate to use the Normal distribution as a means of finding the percentiles of  $S$ . The skewness and kurtosis of the valuations in  $S$  for the models used in the main body of the paper are given in Table 14. For comparison we also show the skewness and kurtosis of 10,000 simulations from the  $N(0,1)$  distribution using the R script below:

```
library(moments) # Load library of statistical functions
set.seed(-1)    # Initialise seed for random-number generator
Z = rnorm(10000) # Generate 10,000 N(0,1) variates
skewness(Z)     # Calculate and display coefficient of skewness
kurtosis(Z)     # Calculate and display coefficient of kurtosis
```

**13.5** The results in Table 14 show that the portfolio valuations in  $S$  have a skewness close to zero and a kurtosis close to 3. In each case the distribution of  $S$  can therefore be closely approximated by a Normal distribution. Thus, for these models and this particular portfolio at least, one could calculate the mis-estimation capital requirements from the mean and standard deviation of  $S$ , together with the appropriate value of  $\Phi^{-1}()$ . Indeed, one could arguably use far fewer than 10,000 simulations, as the mean and standard deviation of Normally distributed data can be reliably estimated with fewer data points.

Table 14: Skewness and kurtosis of portfolio valuations.

| Model   | Skewness | Kurtosis |
|---|----------|----------|
| Age model for 2012 data                               | 0.03     | 3.01     |
| Age model for 2007–2012 data                          | 0.04     | 2.96     |
| Age+Time model for 2007–2012 data                     | 0.01     | 2.98     |
| Age+Time+Gender+Makeham model for 2007–2012 data      | -0.01    | 3.04     |
| Age+Time+Gender+Makeham+Size model for 2007–2012 data | 0.01     | 2.96     |
| Simulated N(0,1) variates                             | 0.03     | 3.05     |

Note: There are 10,000 observations in each data set.

## 14 Appendix 4: Conversion to published tables

**14.1** It is one thing to be able to fit a bespoke model with multiple risk factors. However, actuaries also need to express their bespoke bases in terms of a published table for communication purposes. Examples include reserving bases, communication with third parties (such as regulators and auditors), and of course circumstances where one wants to keep one’s bespoke basis private (as in competitive pricing).

**14.2** A good way to convert mortality bases is by equating reserves, the so-called “equivalent-annuity method” — see Willets (1999) and Richards et al. (2013) for examples. This involves solving an equation where the only difference on each side is the mortality basis being used. For example, two mortality bases  $B$  and  $T$  would be deemed equivalent for a given annuity portfolio if Equation 21 held true:

$$\sum_{i=1}^n w_i \ddot{a}_{x_i}^T = \sum_{i=1}^n w_i \ddot{a}_{x_i}^B \quad (21)$$

where  $w_i$  is the annual pension paid to life  $i$  and  $\ddot{a}_{x_i}$  values an immediate level lifetime annuity paid to a life aged  $x_i$ . Without loss of generality we can think of  $B$  as the bespoke, multi-factor basis and  $T$  as the much simpler one based on the published table. In the case of gender-differentiated rates in  $T$ , which is usually the norm, the summation in Equation 21 is generally performed separately for males and females. For example, the equivalent percentages of S2PA (CMI Ltd, 2014) for the model in Table 6 are given in the first row of Table 15, where the values are consistent with the 2012 column of Table 12.

Table 15: Equivalent percentages of S2PA for the model in Table 6.

| Mortality basis                           | Percentage of S2PA: |              |
|---|---------------------|--------------|
|   | (i) Males           | (ii) Females |
| Best-estimate model in Table 6            | 88.50%              | 87.19%       |
| 95% confidence interval for best-estimate | 78.73–99.54%        | 79.31–96.11% |
| 99.5% stress                              | 75.98%              | 77.04%       |

Note: The basic S2PA mortality table is effective for 2007 and is unadjusted for improvements of any kind. Since the mortality model fits the hazard for 1<sup>st</sup> January 2013, the table percentages are below 100% because of the improvements between 2007 and 2013.

**14.3** Alternatively, the actuary can solve for a target reserve value, as in Equation 22:

$$\sum_{i=1}^n w_i \ddot{a}_{x_i}^T = S^p \quad (22)$$

where  $S^p$  would be the appropriate percentile of the set of mis-estimation valuations carried out with similar  $\ddot{a}$  functions. For example, an approximate 95% confidence interval for the best-estimate basis could be obtained by solving for  $S^{2.5\%}$  and  $S^{97.5\%}$ . This is done in the second row of Table 15, where we note that the confidence interval is not symmetric around the best-estimate percentages. One important point to note is that for gender-differentiated rates we would have to solve Equation 22 separately for males and females. Alternatively, we could solve for the aggregate male and female reserves combined and use the same percentage of table for both genders. Bearing in mind the similar percentages for males and females in Table 15, this would be an option here.

**14.4** Finally, we can also express our 99.5% mis-estimation stress in terms of a standard table by using  $S^{99.5\%}$  in Equation 22. This is given in the third row of Table 15. One reason for doing this might be if one needed a rough-and-ready approximation for daily stress-testing.

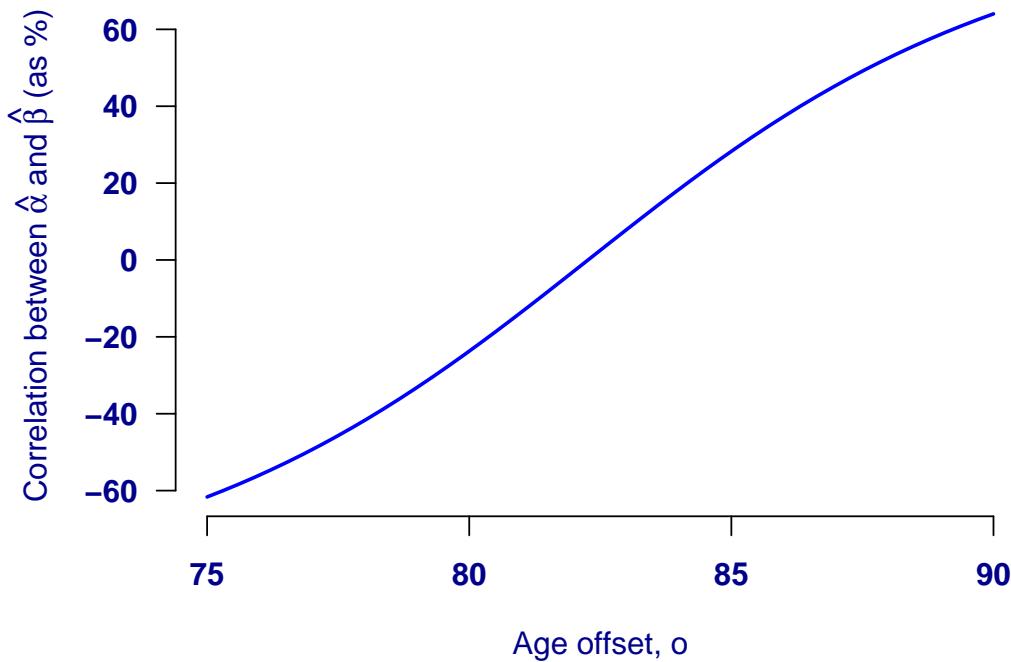
## 15 Appendix 5: Orthogonality and parameter correlation

**15.1** In ¶6.8 we noted a high correlation between the estimates of the intercept ( $\hat{\alpha}_0$ ) and age parameter ( $\hat{\beta}_0$ ). In the case of the simple two-parameter Gompertz model, this is a direct result of the choice of parameterisation in Equation 10. However, we can attempt to orthogonalise  $\hat{\alpha}_0$  and  $\hat{\beta}_0$  by re-parameterising the model as in Equation 23:

$$\mu_{x_i} = e^{\alpha_0 + \beta_0(x_i - o)} \quad (23)$$

where  $o$  is an offset constant. In a model for grouped mortality counts, Cairns et al. (2006) used  $o = \bar{x}$ , i.e. the mean of the age range. However, in Equation 23 setting  $o$  to the mid-point of the age range is not enough to eliminate the correlation between  $\hat{\alpha}_0$  and  $\hat{\beta}_0$  — the distribution of time lived and deaths is not even, as shown in Figure 9. We can calculate the correlation between  $\hat{\alpha}_0$  and  $\hat{\beta}_0$  in Equation 23 for various values of  $o$ , as shown in Figure 12. We see that  $o = 82$  delivers a correlation of -3%, while a value of  $o$  slightly greater than 82 will give zero correlation.

Figure 12: Correlation between  $\hat{\alpha}_0$  and  $\hat{\beta}_0$  for various values of offset age,  $o$ , in Equation 23. Data as in Table 2.



**15.2** In practice it would be necessary to determine  $o$  afresh for each portfolio (or for each altered age range for the same portfolio) if one wanted to eliminate the correlation between  $\hat{\alpha}_0$  and  $\hat{\beta}_0$  in Equation 23. This alone might argue for the simplicity of Equation 10. However, the offset  $o$  also needs to be changed whenever the model changes. For example, we can introduce gender as a risk factor as in Equation 24:

$$\mu_{x_i} = e^{\alpha_0 + \alpha_{\text{male}} I_{i,\text{male}} + \beta_0 (x_i - o)} \quad (24)$$

where  $I_{i,\text{male}}$  is an indicator variable taking the value 1 for males and 0 for females. Table 16 shows the correlations between the parameters, and the correlation between  $\hat{\alpha}_0$  and  $\hat{\beta}_0$  has strengthened due to the change in model. Furthermore, the correlation between  $\hat{\alpha}_0$  and  $\hat{\alpha}_{\text{male}}$  is large and cannot be easily eliminated. In fact, setting  $o = 83$  reduces the correlation between  $\hat{\alpha}_0$  and  $\hat{\beta}_0$  to -1%, but leaves the correlation between  $\hat{\alpha}_0$  and  $\hat{\alpha}_{\text{male}}$  almost unchanged at -64%. Thus, our mis-estimation methodology must take account of parameter correlations as they cannot be avoided. We therefore prefer to use the simpler model structure of Equation 10 without the orthogonalising constant in Equation 23.

Table 16: Percentage correlations ( $c_{i,j} \times 100\%$ ) between coefficients in Equation 24 with  $o = 82$ . The data are the same as in Table 2.

|                              | $\hat{\beta}_0$ | $\hat{\alpha}_{\text{male}}$ | $\hat{\alpha}_0$ |
|------------------------------|-----------------|------------------------------|------------------|
| $\hat{\beta}_0$              | 100             | 12                           | -10              |
| $\hat{\alpha}_{\text{male}}$ | 12              | 100                          | -65              |
| $\hat{\alpha}_0$             | -10             | -65                          | 100              |

Note: As with the variance-covariance matrix, the correlation matrix is symmetric about the leading diagonal. The leading diagonal is 100% because a parameter value is perfectly correlated with itself.