MODELLING MORTALITY
BY CONTINUOUS BENEFIT AMOUNT

By S. J. Richards

LONGEVITAS™

Modelling mortality by continuous benefit amount

Richards, Stephen J.*

October 7, 2020

Abstract

Mortality levels vary by benefit amount, and a common simplification is to group by non-overlapping ranges of varying widths. However, this ignores the continuous nature of benefit amounts and leads to discretisation error, i.e. heterogeneity within benefit ranges and step jumps between adjacent ranges. Another drawback of discretisation is that fitted parameters are not easily extrapolated to values outside the range of the experience data. To address these shortcomings it is often better to model mortality continuously by benefit amount. In this paper we present a method of modelling mortality levels with continuous financial covariates, such as pension size. We split the task into (i) a transform to address the presence of extreme benefit amounts in actuarial data sets, and (ii) a response function to model mortality. Using as few as two parameters, the method avoids discretisation error and extrapolates to amounts outside the range covered by the calibrating data set. We illustrate the method by applying it to seven international data sets of pensioners and annuitants.

Keywords: discretisation, binning, concentration risk, excess kurtosis, Hermite splines.

1 Introduction

Actuaries are long used to mortality levels varying by benefit amount. Historical methods of analysis involved weighting deaths and exposure calculations by benefit amount to account for this. Modern actuarial work uses statistical models, and a common approach to benefit amount is to create an ordinal factor based on non-overlapping ranges. However, such factors create artificial discontinuities at range boundaries, while simplistically assuming homogeneity within a benefit range. These issues can be collectively labelled discretisation error, although there will still be a role for discretisation in some portfolios (such as where benefit amounts might be clustered and do not vary continuously).

More problematically for actuaries, the mortality of the holders of the largest benefits is typically of paramount financial significance — a small sub-group of lives typically accounts for a disproportionate share of liabilities, i.e. concentration risk (see Table 2). However, discretisation often means that this critical sub-group is merged with others who might have different mortality characteristics. The alternative is to create more ranges, but this leads to smaller sub-groups with greater relative estimation error. Another drawback of discretised benefit amounts lies in extrapolating to values outside the range in the experience data, such as might be desired when deriving a pricing basis. To avoid both general discretisation error and these specific actuarial concerns, it is useful to model mortality continuously by benefit amount.

In this paper we present a method of modelling mortality levels with continuous financial covariates, such as pension size or sum assured. Using just two parameters estimated from the data, the method addresses the problem of excess kurtosis of benefit amounts in actuarial data sets, avoids discretisation error and extrapolates mortality effects for amounts outside the range covered by the

*stephen@longevitas.co.uk, Longevitas Ltd, 24a Ainslie Place, Edinburgh, EH3 6AJ. www.longevitas.co.uk

calibrating data set. We illustrate the method by applying it to seven international data sets of pensioners and annuitants.

The plan of the rest of this paper is as follows: Section 2 describes the data sets used, while Section 3 provides a quick overview of how to apply Hermite splines to mortality modelling. Section 4 looks at the variation in mortality by pension amount; Section 5 considers aspects and limitations of discretising a continuous covariate like pension size; Section 6 considers transform functions to standardise benefit amounts and deal with excess kurtosis; Section 7 outlines the application of Hermite splines to modelling mortality continuously by transformed benefit amount and gives the results for seven portfolios; Section 8 compares the continuous and discretised approaches, while Section 9 looks at the impact of pension on modelled mortality levels. Section 10 concludes.

2 Data

The data sets used in this paper consist of pensioner records from occupational pension schemes and insurer annuity portfolios. Due to the financial interest in not paying pensions longer than necessary, such portfolios usually maintain accurate records of when pensions commence and cease. Pension schemes CAN BE specific to a particular employer, and so pensioners often share an occupational background and in some cases are geographically concentrated. Due to regulations and tax-reporting requirements, pension schemes often have detailed additional information besides date of birth, gender and (especially) pension amount. Such portfolios are like longitudinal studies with continuous recruitment: as people retire, new benefit records are set up. Upon the death of an annuitant or former employee, a surviving spouse's pension might also be set up. Table 1 gives an overview of the portfolios (the labelling convention is from Richards et al. [2020], which used five of the same portfolios).

Table 1: Overview of portfolio data. The exposure periods vary within 2000–2019, i.e. excluding the period of the COVID-19 pandemic [The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, 2020]. To put all lives within a portfolio on a common value footing, pensions paid to deaths and early exits are revalued to the end of the exposure period.

Country	Label	Age range	Deaths	Reval'n rate	Description
Canada	CAN2	55–100	2,614	1.5%	Single-employer defined-benefit occupational pension plan.
Germany	DE	60–105	30,480	1%	Occupational pension scheme for a mixture of public- and private-sector employers.
England	ENG	60–105	19,435	2.5%	Defined-benefit occupational pension scheme for a single English local authority.
France	FRA	55–99	28,391	0%	Insurer portfolio of voluntary top-up pensions for employees of higher-education institutions around France.
Netherlands	NLD	50–105	4,894	1%	Single-employer occupational pension scheme in the private sector.
USA	USA2	55–100	17,194	0%	Occupational pension plan in the US.
South Africa	ZA	50–100	60,881	5.5%	Multi-employer database of occupational pension funds.

An important aspect of individual records is the potential for duplicates, i.e. two or more pensions or annuities paid to the same person. For the ENG portfolio in Table 1 the records have been deduplicated using the techniques described in Macdonald et al. [2018, Section 2.5]. Individual records also enable detailed data-quality checks to be carried out — see Macdonald et al. [2018, Sections 2.3, 2.4, 2.7 and 2.8]. To avoid the risk of age mis-statements distorting model fits — see Newman [2018a] and Newman [2018b] — any pensioners appearing to exceed age 105 were excluded. The use of individual data makes data-quality checking far easier than with grouped counts. For example, the FRA data set had nearly 700 annuitants appearing to reach age 110; however, since each shared the same date of birth, it was clear that these records were erroneous.

Pension schemes and annuity portfolios are unequal institutions, where pension amounts can vary more than incomes do in the general population. Table 2 shows the Gini coefficient for each portfolio [Gini, 1921], where a coefficient of 0% would arise from everyone having the same pension and a coefficient of 100% would arise from one person having all the income. The ENG, NLD and ZA data sets are particularly unequal with respect to pension amount. One consequence of this inequality is concentration risk, as shown in the second column of Table 2. We define the top decile of pensioners by sorting all lives in order of pension size and examining the proportion of all pensions paid to the 10% of lives with the largest pensions. This represents a small sub-group of lives in a given portfolio, but Table 2 shows that it is a sub-group of outsized financial importance. A related feature of actuarial data sets is ocasional very large and distorting benefit amounts, the presence of which can be measured by the excess kurtosis [Wetherill, 1982, equation (3.22)]. The final column of Table 2 demonstrates extreme excess kurtosis with respect to pension size for five of the seven portfolios; for comparison, the excess kurtosis of the normal distribution is zero.

Table 2: Inequality of pension size and concentration of risk for the portfolios in Table 1.

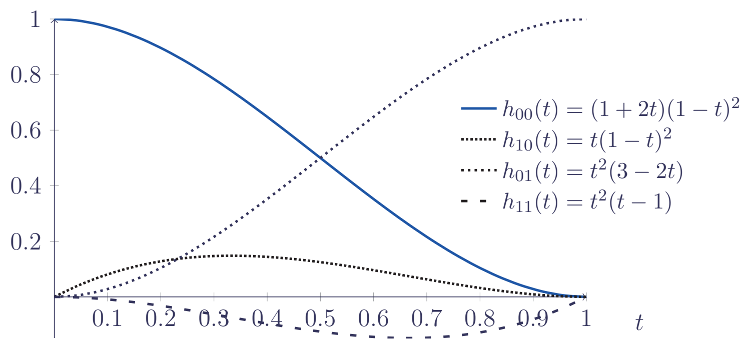
Portfolio	Gini coefficient	Top decile as % of total	Excess kurtosis
CAN2	33.2%	23.4%	46.07
DE	43.4%	26.7%	7.14
ENG	53.4%	37.4%	19.1
FRA	36.0%	25.5%	46.26
NLD	52.9%	37.0%	96.41
USA2	43.8%	25.5%	-0.39
ZA	67.3%	51.9%	76.12

3 A primer on Hermite splines for mortality models

The application of Hermite splines to modelling mortality by age was introduced by Richards [2020], which we briefly recap here. Hermite splines [Kreyszig, 1999] are a basic of four cubic polynomial functions of $t \in [0, 1]$, as depicted in Figure 1. Ignoring the spline function h_{11}, we can use Hermite splines for a model of the mortality hazard, μ_x, as follows:

$$\log \mu_x = h_{00}(t)\alpha + h_{10}(t)m_0 + h_{01}(t)\omega \quad (1)$$

Figure 1: Hermite basis splines for $t \in [0, 1]$.



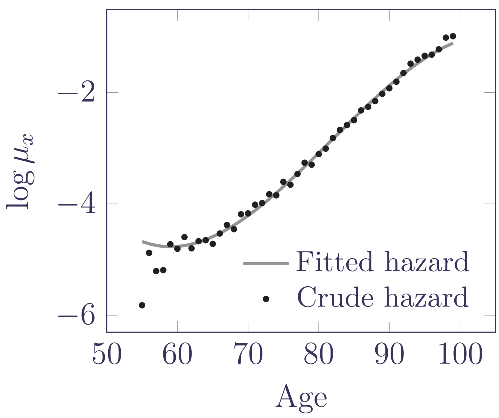
where $t = (x - x_0)/(x_1 - x_0)$, the h functions are shown in Figure 1 and α, m_0 and ω are parameters to be estimated. In this paper we set $x_0 = 50$ and $x_1 = 105$ to cover all the age ranges in Table 1; this means that in equation (1) α and ω are $\log \mu_{50}$ and $\log \mu_{105}$, respectively.

To fit a survival model to individual data we maximise the log-likelihood function in equation (2) [Macdonald et al., 2018, Section 5.3] for the mortality hazard, μ_x, at age x. Each life i enters observation at age x_i and is observed for t_i years. d_i is an indicator variable taking the value 1 if life i is dies at age $x_i + t_i$, or 0 otherwise. $H_x(t)$ is the integrated hazard function in equation (3). For example, for the USA2 data set we have $\hat{\alpha} = -4.19378$, $\hat{\omega} = -0.957487$ and $\hat{m_0} = -7.03612$. The usefulness of Hermite splines in modelling mortality by age is shown in Figure 2. Details of how to perform the numerical integration necessary for $H_x(t)$ are given in Richards [2020, Appendix B], which also gives formulae for (i) the first partial derivatives needed for optimizing the log-likelihood and (ii) the second partial derivatives for estimating the parameter covariances.

$$\ell = -\sum_{i=1}^{n} H_{x_i}(t_i) + \sum_{i=1}^{n} d_i \log \mu_{x_i+t_i} \qquad (2)$$

$$H_x(t) = \int_0^t \mu_{x+r} dr \qquad (3)$$

Figure 2: Crude mortality hazard and fitted hazard using equation (1) for USA2.



Equation (1) is a three-parameter model for mortality by age, which is extended to a five-parameter model allowing for age and gender in equation (4):

$$\log \mu_{x_i} = h_{00}(t)(\alpha + \alpha^{\text{female}} z_i^{\text{female}}) + h_{10}(t) m_0 + h_{01}(t)(\omega + \omega^{\text{female}} z_i^{\text{female}}) \qquad (4)$$

where α^{female} and ω^{female} are the additions to $\log \mu_{50}$ and $\log \mu_{105}$ for females, respectively; z_i^{female} is an indicator variable taking the value 1 if life i is female and zero otherwise. In equation (4) α and ω represent $\log \mu_{50}$ and $\log \mu_{105}$ for males, respectively.

An important point to note about equation (2) is that it is the log-likelihood for a survival model with left-truncated data, which is standard for data encountered in actuarial work; see Macdonald et al. [2018, Section 4.3]. This contrasts with survival models used in medical research, where left-truncation is relatively uncommon and where likelihoods are usually for non-left-truncated data [Collett, 2003, Chapter 6].

4 Mortality by pension amount

We illustrate the variation in mortality by pension amount for the portfolios in Table 1. We do this in two ways: (i) the traditional comparison against a reference table, and (ii) using a statistical model. We start by adapting Macdonald et al. [2018, equations (8.1) and (8.2)] to define a measure of the impact of pension size on mortality, Y, as in equation (5); summation is over single years of age x, d_x is the number of deaths aged x last birthday, E_x^c is the central rate of exposure time between age x and $x+1$ and w_x is total pension payable to lives aged x. μ_x is the mortality hazard according to some reference basis, e.g. a published reference table (the choice of which is largely unimportant). In equation (5) Y represents the standardised reduction of mortality due to pension-size effects, where a positive value means that higher-income pensioners have lower mortality.

$$Y = \frac{\sum_x d_x}{\sum_x E_x^c \mu_{x+1/2}} - \frac{\sum_x d_x w_x}{\sum_x E_x^c \mu_{x+1/2} w_x} \qquad (5)$$

Table 3 shows the impact of pension size using the unadjusted S2PA table from CMI Ltd [2014] for all the portfolios in Table 1. It is clear that mortality weighted by pension size is consistently lighter for every portfolio, and that some dramatic differences exist. However, comparisons against a reference table provide no indication of statistical significance, for which we need some kind of statistical framework.

Table 3: Reduction in mortality, Y, due to impact of pension size.

Portfolio	Males	Females
CAN2	4.5%	4.4%
DE	8.5%	1.7%
ENG	18.1%	7.7%
FRA	3.7%	2.7%
NLD	15.2%	10.0%
USA2	3.1%	3.2%
ZA	32.4%	22.7%

One statistical approach is to fit models for mortality by age and gender, as per equation (4), then examine the deviance residuals by pension band [Macdonald et al., 2018, Section 6.5]. For each portfolio we define twenty size-bands (vigintiles) such that the number of lives in each is as similar as possible. Within each size-band we calculate the expected number of deaths [Macdonald et al., 2018, equation (6.13)] and then use the actual number of deaths to calculate a Poisson deviance residual [Macdonald et al., 2018, equation (6.14)]. A systematic pattern in the deviance residuals is evidence of mortality variation by pension size; specifically, a pattern of positive residuals for lower size-bands and negative residuals for higher size-bands indicates reducing mortality with increasing pension income.

Figure 3: Deviance residuals [McCullagh and Nelder, 1989, page 39] by pension vigintile, with 1 representing the smallest pensions and 20 the largest. Each vigintile within a given portfolio contains a similar number of lives, and the panels show the deviance residuals from a simple model allowing for variation by age and gender[Macdonald et al., 2018, Section 6.5.1]. The residuals for each portfolio suggest a link between increasing pension size and decreasing mortality.

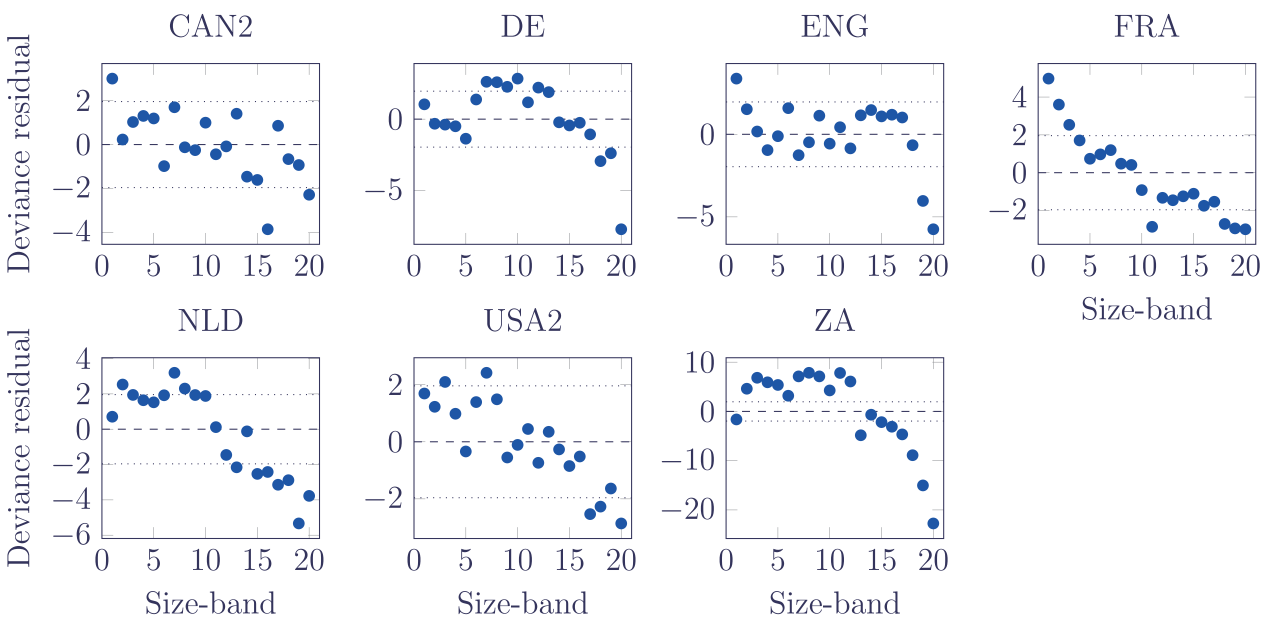
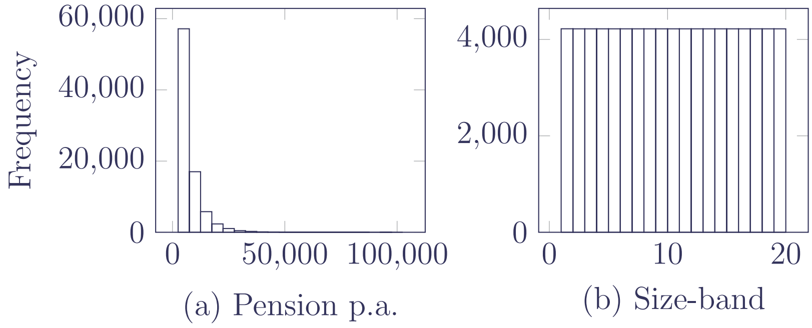


Figure 3 shows the deviance residuals from the various model fits after allowing for age and gender. All the residual tests of normality of Macdonald et al. [2018, Section 6.6] fail at the 5% level, although the portfolios vary considerably in the strength of association between pension size and mortality. One reason for this lies in the different taxation systems and benefit practices among the various countries [Richards et al., 2020, Section 7]. However, the implication of Figure 3 is clear — pensioner mortality usually varies by pension size. Table 2 shows the financial importance of allowing for this, and so such variation must be handled in any model for actuarial purposes.

5 Discretisation: features and limitations

Table 2 shows extreme excess kurtosis for five of the seven portfolios. This is also illustrated for the ENG portfolio in Figure 4(a), where the presence of pensions over 23 times the average distorts the scale of the plot. These large pensions are relatively rare, but Table 2 shows that they have an outsized financial importance.

Figure 4: Histograms of ENG portfolio by (a) pension and (b) pension size-band.



One approach to handling excess kurtotis when analysing mortality by pension size is to assign each pension to a size-band. The continuous nature of pension size is discarded as we sort the lives in ascending order of benefit size and select boundaries such that there are equal numbers of lives in each range (or as close to equal as is possible). This is shown for the ENG portfolio in Figure 4(b), where the binning process has curtailed the excess kurtosis; the excess kurtosis of the uniform distribution is -1.2, compared to 19.1 for the raw pension amounts.

Figure 5: ENG portfolio by pension vigintile, with 1 representing the smallest pensions and 20 the largest. Each vigintile contains a similar number of lives. Panel (a) shows the time lived in years and panel (b) shows the number of deaths. Panel (c) shows the deviance residuals [Macdonald et al., 2018, Section 6.5.1] from a simple model allowing for age and gender, demonstrating the link between increasing pension size and decreasing mortality.

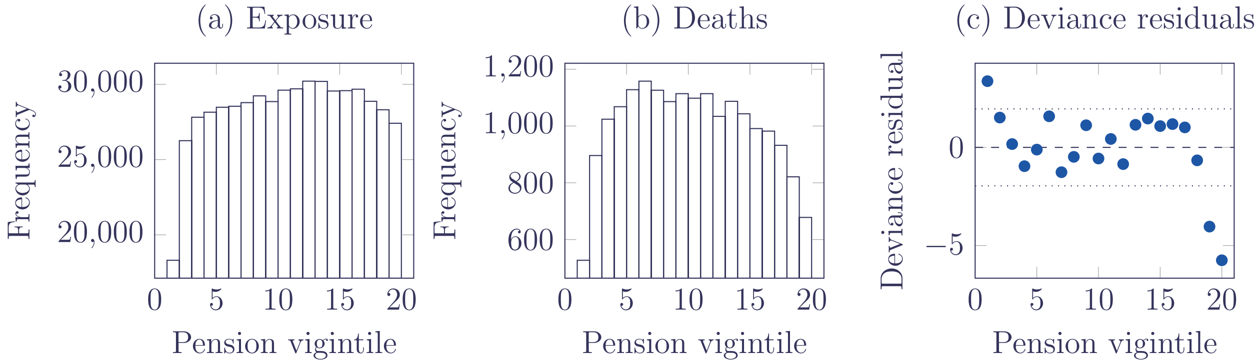


Figure 5 shows some further aspects of this equal-counts binning. Although the numbers of lives in each vigintile are similar, panel (a) shows that the first vigintile has much less exposure time than the others; this is due to the commutation of very small pensions to save administrative expense. Panel (b) shows that the first vigintile also has far fewer deaths than the other vigintiles for the same reason. Panel (c) plots the deviance residuals [Macdonald et al., 2018, Section 6.5.1] by vigintile, showing a general tendency for mortality to fall with increasing pension size, with particularly stark drops in mortality level for the largest pensions.

One feature of such discretisation (or binning) is that it leads to widely different benefit ranges. For example, dividing the ENG dataset into twenty near-equal bins produces 4,221 individuals with annual pensions in £0–£289.09 p.a. at the smallest end (vigintile 1) but 4,208 individuals in £15,284.91–£103,110.53 p.a at the highest end (vigintile 20). The latter vigintile is a problem from an actuarial perspective: the assumption is that everyone in the vigintile has the same mortality, and yet we could reasonably suspect that someone with a pension of £100,000 p.a. might be of different socio-economic status and mortality than someone with a pension of £16,000 p.a. A further actuarial issue would be in using this approach to calibrate a pricing model for wealthier individuals still; even assuming the estimated effect for vigintile 20 were correct for lives up to £100,000 p.a., would it necessarily apply to an individual with a pension of £1 million p.a.? And, if not, how would one extrapolate a suitable effect for a pricing basis?

Vigintiles 18 and 19 in Figure 5(c) pose a problem for actuaries — while their mortality effects appear to be part of a smooth progression, there is clearly a large difference between the two vigintiles on average. The boundary between vigintiles 18 and 19 is £11,115.13, so a hypothetical Pensioner A with £11,100 p.a. would fall into the higher-mortality vigintile 18, while Pensioner B with £11,200 would fall into the lower-mortality vigintile 19. With just £100 p.a. difference we would reasonably expect the mortality of these two pensioners to be near-identical, all other things being equal, and yet the vagaries of discretisation leads to them being treated differently. This is discretisation error.

The obvious solution to the boundary problem — increasing the number of size-bands — is not in fact a viable one. Figure 6 shows the deviance residuals for an age-gender model when the ENG portfolio is subdivided into a hundred size-bands. The underlying continuous nature of the link between pension size and mortality is clear, but the signal is noisier compared with Figure 5(c). If used in a model, there would be higher standard errors for the estimates (to say nothing of the gross over-parmeterization of a model with 99 discrete pension-size effects). Some kind of smoothing would be required to counteract the variance of the resulting estimates. One option might be to use a suitable basis of splines with knot points at the size-band edges. However, it seems perverse to discretise a continuous variable only to have to add a smoothing process to combat the noise thus introduced. Another option would be to optimise the grouping of centiles into a smaller number of ranges, but this would be computationally expensive due to the number of combinations, nor would it solve the problem of step jumps in mortality at range boundaries, and nor would it extrapolate to amounts outside the data set.

Figure 6: Deviance residuals of ENG portfolio by pension centile (1 represents the smallest pensions and 100 the largest).

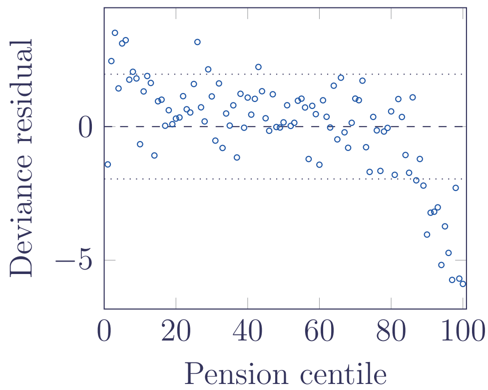


Figure 7: Deviance residuals of ENG portfolio for pension centiles 86–100.

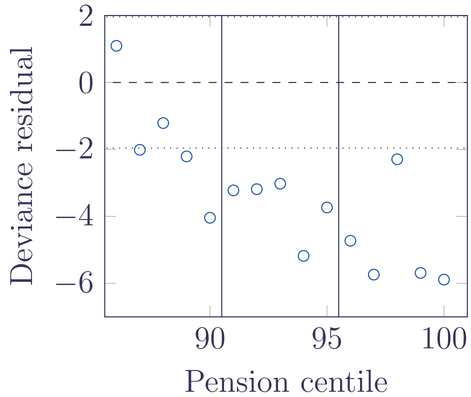


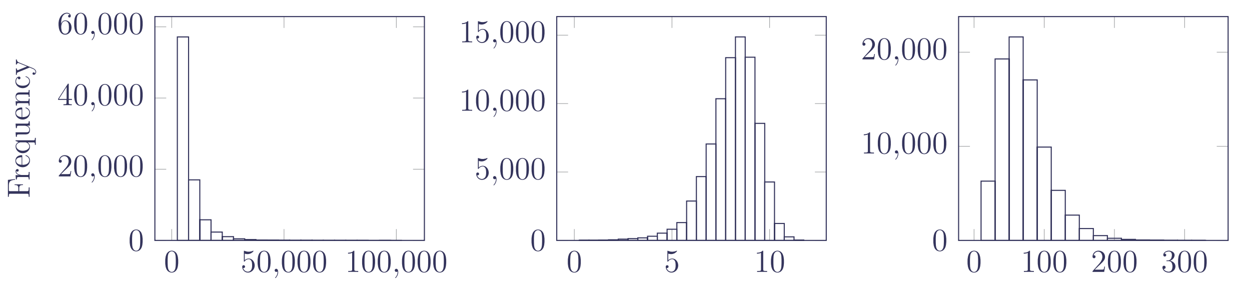
Figure 7 focuses on the deviance residuals for centiles 86–100 with vertical lines corresponding to the boundaries of vigintiles 18–20 in Figure 5. Although there is random variation, it is clear that the vigintiles are not homogeneous with respect to mortality. The centile residuals in Figure 7 suggests that mortality rates are falling continuously within each of vigintiles 18, 19 and 20. Under such circumstances, a statistical model fitting an effect for each vigintile will get the mortality level correct by lives, but it will naturally tend to under-state the mortality measured by amounts.

While discretisation of benefit amounts into non-overlapping ranges is useful for modelling, it has several issues: (i) discontinuities at range boundaries, (ii) the false assumption of homogeneity within wide bands, especially for the largest pensions, and (iii) the challenge of extrapolation. These problems could all be solved by modelling mortality continuously by benefit size. We therefore seek a continuous approach that reduces excess kurtotis in a similar manner to the binning in Figure 5, but without the discretisation error that accompanies it. Some candidate transform functions are discussed in the next section.

6 Transforms for benefit amounts

We have two requirements of a continuous transform of benefit amount: (i) that it addresses the excess kurtosis in the distribution of benefit amounts, and (ii) that it acts somewhat like the binning in Figure 5 in evenly distributing lives. Figure 8 shows that variance-stabilising transforms [Bartlett, 1947] like the logarithm and square root work fairly well in reducing the skew shown in panel (a), but the resulting distributions still have more excess kurtosis than we might like.

Figure 8: Histograms of ENG portfolio by pension size and various transforms thereof.



(a) Pension p.a. (b) $\log(1+\text{Pension})$ (c) $\sqrt{\text{Pension}}$

Instead, we consider a transform function, $\tau(s)$, that maps the benefit amount, $s \geq 0$, onto the interval $[0, 1)$. The transform function must preserve the relationship between benefit amounts, i.e. if $s_0 < s_1$ then $\tau(s_0) < \tau(s_1)$. A practical point for pensions and annuities is that $\tau(0)$ should exist, as there are often some records in experience data with zero benefit amounts. The cumulative distribution function of any non-negative random variable would satisfy these requirements. Equations (6)–(10) list the transform functions used in this paper, where $\lambda_0 \in \mathbb{R}$ is a parameter that controls the distribution of the transformed amounts and $\Phi(x)$ is the N(0,1) cumulative distribution function. Equations (6)–(10) are illustrated in Figure 9 for $\lambda_0 = -10$. The transforms reduce the excess kurtosis of the pension amounts: $s \in [0, 25,000)$ maps relatively evenly onto the lower half of $[0, 1)$,

$$\text{Logistic} : se^{\lambda_0}/(1 + se^{\lambda_0}) \tag{6}$$

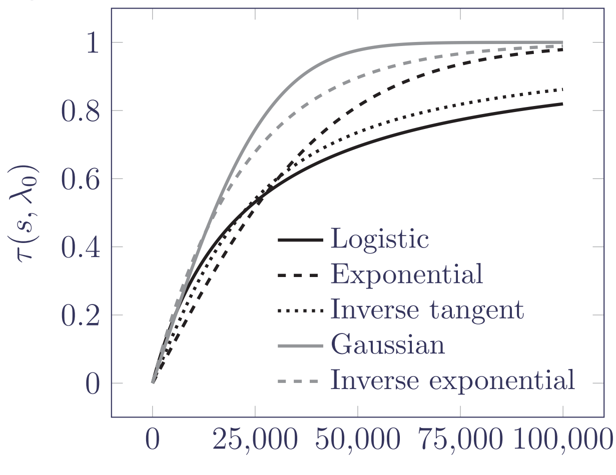
$$\text{Exponential} : 1 - \exp(-se^{\lambda_0}) \tag{7}$$

$$\text{Inverse tangent} : \frac{2}{\pi}\tan^{-1}(se^{\lambda_0}) \tag{8}$$

$$\text{Gaussian} : 2\Phi\left(se^{\lambda_0}\right) - 1 \tag{9}$$

$$\text{Inverse exponential} : 2\left(1 + \exp\left(-se^{\lambda_0}\right)\right)^{-1} - 1 \tag{10}$$

Figure 9: Transform functions with $\lambda_0 = -10$.



while $s > 25,000$ maps onto the upper half of $[0, 1)$. The value of 25,000 was chosen by eye from inspection of Figure 9, and is dependent on the value of λ_0. Useful values of λ_0 will be determined by the monetary scale of the pensions, i.e. the pension levels and the currency unit.

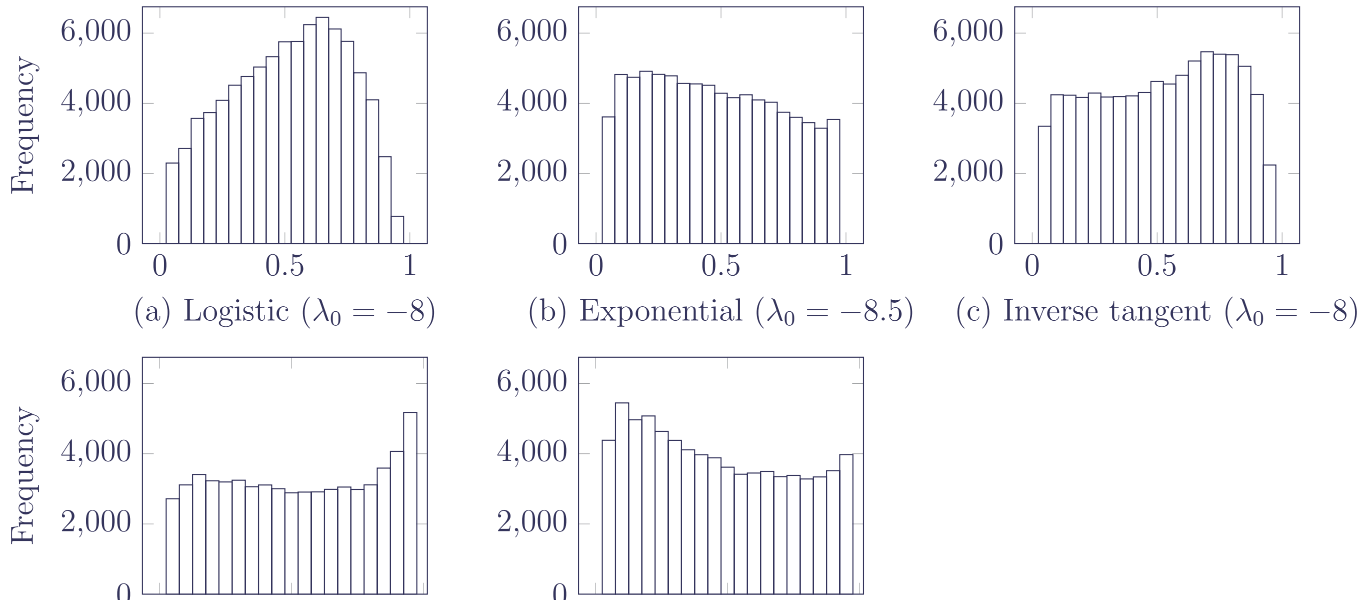
Equations (6)–(9) were adapted from the inverse link functions for a binomial GLM, as shown in Table 4. The two are not identical, however, as the transforms only map half of the real interval: equations (6)–(9) map $s \geq 0$ onto $[0, 1)$, whereas the inverse link function for a GLM maps $\eta \in \mathbb{R}$ onto $[0, 1)$. The transform in equation (10) was additionally chosen because it is based on the the distribution function for the logistic distribution, which has greater kurtosis than the normal distribution.

Table 4: Relationship between equations (6)–(9) and link functions for binomial GLMs [Aitken et al., 1989, p.169].

Transform	Related link function
Logistic	Logit [Berkson, 1944]
Exponential	Complementary log-log [Fisher, 1922]
Inverse tangent	Cauchy [Morgan and Smith, 1992]
Gaussian	Probit [Bliss, 1934]

In practice λ_0 in equations (6)–(10) will be estimated from the data, and so will vary between portfolios depending on the kurtosis of benefit amounts. Figure 10 shows examples applied to the pension amounts in the ENG portfolio.

Figure 10: Histograms of ENG portfolio by transformed pension size, $\tau(s)$, using equations (6)–(10). In contrast to Figure 8, these parameterised transforms spread the benefit amounts relatively evenly over $[0, 1)$.

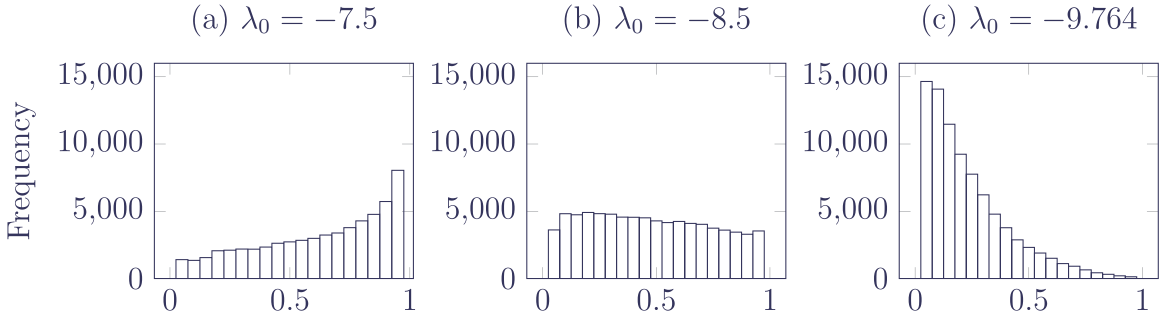


(a) Logistic ($\lambda_0 = -8$) (b) Exponential ($\lambda_0 = -8.5$) (c) Inverse tangent ($\lambda_0 = -8$)

(d) Gaussian ($\lambda_0 = -8$) (e) Inverse exponential ($\lambda_0 = -8$)

The histograms in Figure 10 are not perfectly uniform, and we might need different transforms (possibly with a second transform parameter) to achieve greater uniformity. However, perfect uniformity is not a goal of itself and indeed a deliberate lack of it could be an advantage — in the example of vigintile 20 in Section 5, there was a question mark over the homogeneity of a group with such a wide pension range. There may therefore be improvements in fit from selecting a value of λ_0 that distributes amounts non-uniformly, as in Figure 11(a) and (c). As we will see, it can be advantageous to use an estimation procedure to find the value of $\hat{\lambda}_0$ that optimises the fit with respect to mortality, rather than produce uniformity over $[0,1)$. It can also be useful to permit different values of λ_0 for different categories, such as males v. females or retirees v. surviving spouses. The continuous allowance for mortality by transformed amount is the subject of the next section.

Figure 11: Histograms of ENG portfolio by pension size using the exponential transform in equation (7) with varying values of λ_0. A non-uniform distribution, such as in panel (c), may actually be beneficial in terms of distributing lives according to their mortality characteristics (the value $\lambda_0 = -9.764$ comes from the model fit in Table 6).



7 Continuous modelling of mortality by benefit amount

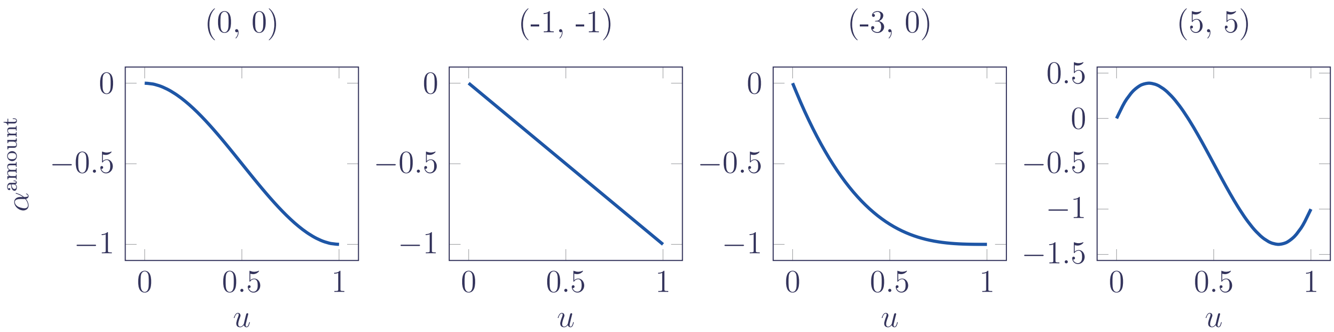
We assume that we have selected a transform function, $\tau(s, \lambda_0)$, from equations (6)–(10). For a given benefit amount, s, we calculate a transformed value $u = \tau(s, \lambda_0)$. We then use this to calculate a mortality effect, $\alpha^{\text{amount}}(u)$, that can be added to $\log \mu_x$ to raise or lower risk depending on benefit size. To do so continuously we need a response function operating on u. Since $u \in [0, 1)$ we can use the same basis of Hermite splines in Figure 1 to flexibly model mortality patterns by transformed benefit size.

Without loss of generality we take a benefit size of zero to be the baseline, which means we do not need the Hermite basis spline h_{00}. Our continuous mortality response, $\alpha^{\text{amount}}(u)$, is therefore:

$$\alpha^{\text{amount}}(u) = \omega^{\text{amount}} h_{01}(u) + m_0^{\text{amount}} h_{10}(u) + m_1^{\text{amount}} h_{11}(u) \tag{11}$$

where ω^{amount} is the ultimate effect of an infinite benefit amount (AmountUltimate), m_1^{amount} is the gradient of the mortality effect approaching ω^{amount} (AmountGradientUltimate) and m_0^{amount} is the initial direction of the mortality effect from a zero benefit amount (AmountGradientInitial). Figure 12 shows some of the possible response shapes using equation (11).

Figure 12: Influence of $(m_0^{\text{amount}}, m_1^{\text{amount}})$ on the shape of α^{amount} in equation (11) with $\omega^{\text{amount}} = -1$.



The full flexibility shown in Figure 12 is not always required. Indeed, for some portfolios it may suffice to replace equation (11) with a straight line, i.e. $\alpha^{\text{amount}} = \omega^{\text{amount}}u$, and allow the value of λ_0 to adapt the transform so that benefit amounts are distributed such that the linear assumption holds. In many cases m_0^{amount} and m_1^{amount} can be dropped and it will suffice to just estimate λ_0 and ω^{amount}, as in the models in this paper. In such cases λ_0 will be optimised so that the mortality of the transformed pension amounts follows the $h_{01}(u)$ spline, and so the estimates $\hat{\lambda}_0$ and $\hat{\omega}^{\text{amount}}$ will be highly correlated.

Our mortality model for age, gender and pension size is then given in equation (12), with $\log \mu_x$ defined in equation (4), $t = (x - x_0)/(x_1 - x_0)$ and $\alpha^{\text{amount}}(u)$ defined in equation (11). The multiplication of $\alpha^{\text{amount}}(u)$ by $h_{00}(t)$ allows the effect of pension size to automatically reduce with increasing age; see Richards [2020, Section 5]. We can then refit the mortality models using equation (2) to estimate not just the parameters that allow for age and gender, but also estimate the values of λ_0 and ω^{amount} that best reflect mortality by pension size.

$$\log \mu_x^* = \log \mu_x + \alpha^{\text{amount}}(u) h_{00}(t) \quad (12)$$

Table 5 shows the stepwise improvements in fit from adding first gender and then a continuous pension-size effect. In all cases the addition of pension size improves the fit measured by the AIC [Akaike, 1987]. This is the case even for CAN2, which had the weakest association between pension size and mortality in the deviance residuals in Figure 3. Of particular interest is the NLD portfolio, where pension size explains more of the mortality variation than gender does.

Table 5: Stepwise development of AIC from including gender as categorical factor and pension size as continuous variable. For simplicity the pension-size effect only uses λ_0 and ω^{amount}, i.e. $m_0^{\text{amount}} = m_1^{\text{amount}} = 0$. The pension transform function, $\tau(s, \lambda_0)$, is the inverse tangent in equation (8). Final AICs might not match due to rounding.

Model	$\log \mu_x$	CAN2	DE	ENG	FRA	NLD	USA2	ZA
Age only	eq. (1)	20,667	240,900	151,257	264,900	36,701	125,439	448,237
+ gender	eq. (4)	-111	-1,893	-993	-1,340	-123	-267	-2,339
+ pension size	eq. (12)	-32	-115	-302	-124	-137	-38	-1,737
Age+gender+pension size		20,524	238,892	149,962	263,436	36,441	125,134	444,161

One consideration is the impact of alternative transform functions from equations (6)-(10). Table 6 shows that the choice of transform function does not make a large difference in fit for the ENG portfolio. However, from experimentation with other data sets (not shown) the inverse tangent tends to provide a good all-round fit within a reasonable number of iterations.

Table 6: Results of various transforms when modelling pension-varying mortality of ENG data set.

Transform	AIC	$\hat{\lambda}_0$	$\hat{\omega}^{\text{amount}}$
Logistic	149,965	-10.170	-2.840
Exponential	149,963	-10.050	-1.881
Inverse tangent	149,962	-9.764	-2.050
Gaussian	149,962	-9.863	-1.515
Inverse exponential	149,962	-9.411	-1.578

The deviance residuals from the models in Table 5 are shown in Figure 13, which can be directly compared with Figure 3. The patterns by pension size-band are substantially reduced in all portfolios, showing that in most cases the mortality variation by pension amount has been allowed for. The exception is the ZA portfolio, where many of the residuals are still too large to be consistent with random variation.

Figure 13: Deviance residuals by pension vigintile from models in Table 5 allowing for age and gender and continuous pension size in Table 5.

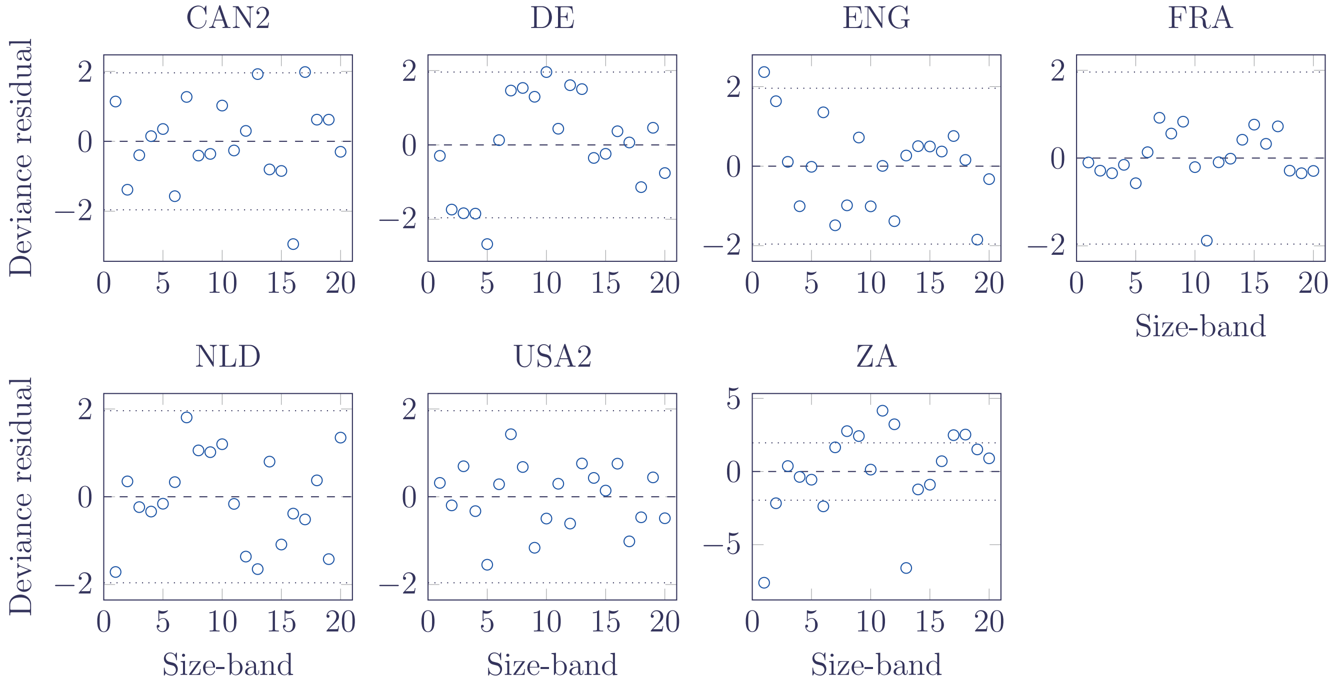


Figure 14 shows the centile residuals 86–100 for the ENG portfolio in Figure 13 in more detail. As can be seen by comparison with Figure 7, the continuous amounts model has made a smooth allowance for these all-important large pensions without step jumps and without material bias at any point (although there appears to be a slight negative bias overall, it is just consistent with random variation). It is reasonable to assume that the model could then sensibly extrapolate mortality effects for larger pensions than the maximum observed, thus making the model a suitable starting point for a pricing basis. Finally, Table 7 shows that the Hermite-spline response function is generally a better fit than a simple linear one.

Figure 14: Deviance residuals of ENG portfolio for pension centiles 86–100 using model for age, gender and pension size.

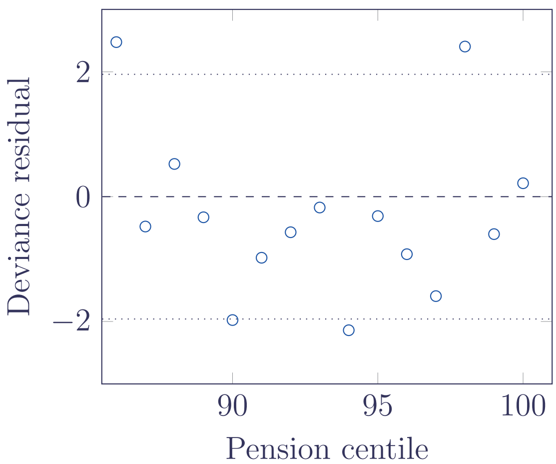


Table 7: Comparison of AICs using linear and Hermite-spline response functions with inverse tangent transform.

Response function	Form of $\alpha^{\text{amount}}(u)$	CAN2	DE	ENG	FRA	NLD	USA2	ZA
Linear	$\omega^{\text{amount}}u$	20,532.8	238,923	149,944	263,568	36,464.7	125,129	444,261
Hermite spline	$\omega^{\text{amount}}h_{01}(u)$	20,524.5	238,892	149,938	263,567	36,453.8	125,129	444,224
Advantage for Hermite spline		8.3	31	6	1	10.9	0	37

8 Comparison of discretised and continuous approaches

Table 8 shows the model fits using two different approaches to modelling mortality by pension size. The first six rows of Table 8 show the AIC and BIC [Schwarz, 1978] from splitting the population into equal-sized groups by pension size, while the last row shows the information criteria for the simplest model for the continuous approach. The continuous model is the best-fitting according to both the AIC and BIC.

Table 8: Results of different approaches to pension-varying mortality of ENG data set (age and gender effects are also allowed for). Factor levels contain equal numbers of lives. The continuous model uses an inverse tangent transform and the Hermite-spline response function.

Approach to pension	Factor levels	Lives in each level	Information criterion: (a) AIC	(b) BIC
Factor	2	42,193	150,179	150,235
	3	28,128	150,129	150,195
	4	21,094	150,085	150,160
	5	16,876	150,042	150,126
	10	8,437	149,973	150,104
	20	4,208	149,972	150,197
Continuous	n/a	n/a	149,962	150,028

An alternative approach is to optimise the range breakpoints by starting with 20 bands of equal numbers of lives and merging adjacent bands with similar mortality. We do this by searching for merges that produce the lowest AIC for a given target number of levels (targeting the BIC would produce the same result, as the number of parameters is held constant). The process is iterative and the time taken to consider all possible merges increases with both the target number of levels and the initial number of size-bands. For this reason, we adopt a more limited searching algorithm for targeting four or more factor levels. In our implementation we have further reduced run-times by using parallel processing to spread calculations over 63 threads [Butenhof, 1997]. The results of this optimisation process are shown in Table 9. We can see that, for a given number of factor levels, the model fit is materially improved compared with the same number of levels in the equally-sized bands of Table 8. The best-fitting optimised factor in Table 9 has six levels, and it is the equal of the continuous model in Table 8 in terms of the AIC (but not the BIC). However, we still have discretisation error, and the granularity is limited by the number of initial bands. A further refinement for reducing discretisation error might be to optimise from (say) 100 initial discretised size-bands instead of 20, although this would substantially increase the run-time in searching for the optimal breakpoints.

Table 9: Results of optimising breakpoints so that factor levels contain unequal numbers of lives. Applied to ENG data set with age and gender also allowed for in the model.

Factor levels	Lives in top level	Information criterion: (a) AIC	(b) BIC
2	12,648	150,019	150,075
3	12,648	149,994	150,059
4	12,648	149,997	150,072
5	12,648	149,986	150,071
6	4,208	149,962	150,056
7	4,208	149,962	150,065
8	4,208	149,962	150,074
9	4,208	149,963	150,085
10	8,440	149,975	150,105

9 Impact of pension size on modelled mortality

Table 10 shows the parameter estimates for the model fitted to the ENG data set with allowance for age effects, gender differentials and a continuous pension-size effect. The model is very parsimonious, containing just seven parameters.

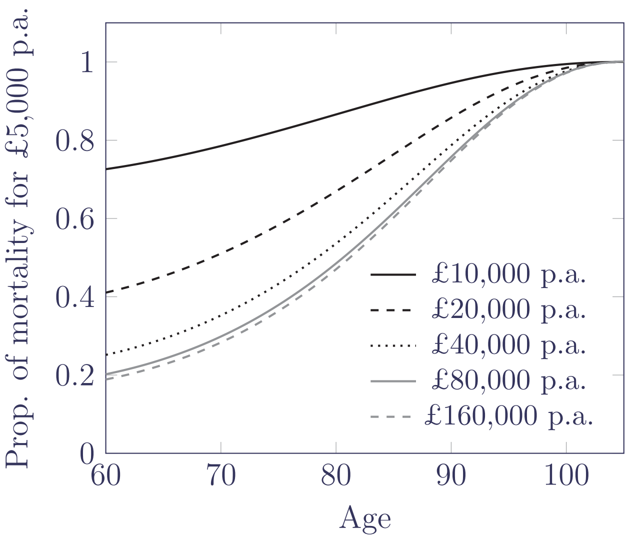
Table 10: Mortality model fitted to ENG portfolio allowing for variation by age, gender and continuous pension-size (inverse tangent transform and Hermite-spline response function). *** denotes a p-value of 0.1% or lower.

Parameter name	Type	Estimate	Standard error	Z-value	Sig.	Lives	Deaths
AgeGradientYoungest	m_0	-4.306	0.5497	-7.83	***	84,389	19,435
AmountTransformParameter	λ_0	-9.764	0.1279	-76.37	***	84,389	19,435
AmountUltimate	ω^{amount}	-2.050	0.2831	-7.24	***	84,389	19,435
Gender.F	α^{female}	-0.977	0.0415	-23.53	***	51,064	10,690
Gender.F:Oldest	ω	-0.156	0.0327	-4.76	***	51,064	10,690
Intercept	α	-3.904	0.1088	-35.90	***	84,389	19,435
Oldest	ω	-0.772	0.0335	-23.09	***	84,389	19,435

Figure 15 shows the modelled mortality rates by pension size for the ENG portfolio, expressed as a proportion of the mortality rates for lives receiving £5,000 p.a. The convergence of mortality rates with increasing age is a deliberate design feature of the Hermite-spline approach to mortality modelling [Richards, 2020, Section 5].

Figure 15 shows that large reductions in mortality happen with modest absolute increases in annual pension: doubling a £5,000 p.a. pension reduces mortality at age 60 by a quarter, while doubling again to £20,000 p.a. reduces mortality rates to 40% of the original level. However, there is a diminishing return: doubling again to £40,000 p.a. and £80,000 p.a. produces ever-smaller reductions. Doubling from £80,000 p.a. to £160,000 p.a. results in only a tiny reduction. A cash increase of £X at £5,000 p.a. produces a much larger proportional change in mortality-relevant life circumstances than it can at much larger incomes; this is probably because, at small income levels, the extra is spent on basic needs.

Figure 15: Ratio of mortality rates by pension size relative to those with £5,000 p.a. according to model in Table 10.



One confounding element in Figure 15 is the role of time: the model used does not allow for time trends, although these can be added [Richards, 2020, Section 7]. Nor does the model allow for year-of-birth (cohort) effects [Willets, 1999], which are important because smaller pensions will tend to be received by those in earlier cohorts who retired long ago. This may lead to exaggeration of the impact of pension size on mortality in Figure 15.

Similarly, there is greater diversity amongst small and medium-sized pensions than there is for larger pensions: while a very large pension is an unambiguous signal of an individual with high income, small and mid-sized pensions are often only weak indicators of overall retirement income.

To see why, consider two pensioners of the same age and gender with a pension of £10,000 p.a. — for the first pensioner this may be their whole retirement income after an entire working life on a modest salary, whereas for the second the pension may be the result of short period of service with a high salary. In such circumstances, the effect of pension size on mortality in Figure 15 might be under-stated. To improve a model for small and mid-sized pensions, one might add a geodemographic factor based on postcode (UK, Richards [2008]), postal code (Canada and the Netherlands) or nine-digit ZIP code (USA). Other territories can achieve similar geodemographic profiling using the whole address. Adding such factors to the model would change the impact of pension size on mortality in Figure 15.

Bearing in mind the limitations of a mortality model with just three risk factors, Table 11 shows the impact of pension size on period life expectancy, showing a wide spread from low to high incomes and a reducing male-female differential as income increases.

Table 11: Period life expectancies implied by the model in Table 10.

Pension (£ p.a.)	Life expectancy at age 60: Males	Females
5,000	20.50	25.01
10,000	22.04	26.18
20,000	24.42	27.96
40,000	26.13	29.25
80,000	26.81	29.75
160,000	27.01	29.91

Another benefit of the continuous approach lies in the ability to fit interactions. For example, we could vary the effect of pension size by gender (not shown). With the most basic two-parameter form of the continuous model we could also test the addition of separate parameters for λ_0 and ω^{amounts} for males and females. To do this with a six-level ordinal factor would require adding five parameters, instead of just two for the continous approach. The parsimony of the continuous approach allows far simpler models compared to discretised factors, especially if interactions with other variables are required.

10 Conclusions

We can model mortality continuously by benefit amount using a combination of (i) a simple transform and (ii) a Hermite-spline response function. The transform addresses the high excess kurtosis of benefit amounts in actuarial data, while the response function models the mortality level continuously across the transformed amounts. At its simplest, there are just two parameters that need to be estimated from the data, although a further two optional parameters exist for more complex patterns of mortality. In addition to fitting the existing experience data, the combination of transform and Hermite-spline response function permits extrapolation of results to benefit amounts outside the range covered by the calibrating data set. This makes the method useful for creating pricing models for very large benefit amounts.

Acknowledgments

The author thanks the following for their support in providing data: Denis Dupont; Hélène Queau and Clément Frappier; Jay Wang; Susanne Rosenbusch. The author also thanks Gavin Ritchie, Torsten Kleinow, Stefan Ramonat and Gary Velcich for helpful comments on earlier drafts. Any errors or omissions remain the responsibility of the author. All models were fitted with Longevitas, graphs were done in tikz and pgfplots, while typesetting was done in LaTeX. Data preparation was carried out with R [R Core Team, 2017].

References

M. Aitken, D. Anderson, B. Francis, and J. Hinde. *Statistical modelling in GLIM*. Oxford University Press, 1989. ISBN 0-19-852204-5.

H. Akaike. Factor analysis and AIC. *Psychometrica*, 52:317–333, 1987. ISSN 0033–3123. doi: https://doi.org/10.1007/BF02294359.

M. S. Bartlett. The use of transformations. *Biometrics*, 3(1):39–52, 1947. ISSN 0006341X, 15410420. URL `http://www.jstor.org/stable/3001536`.

J. Berkson. Application of the logistic function to bio-assay. *Journal of the American Statistical Association*, 39(227):357–365, 1944. doi: 10.2307/2280041.

C. I. Bliss. The method of probits. *Science*, 79(2037):38–39, 1934. doi: 10.1126/science.79.2037.38.

D. R. Butenhof. *Programming with POSIX Threads*. Addison-Wesley, Boston, 1997. ISBN 978-0-201-63392-4.

CMI Ltd. *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, 2014.

D. Collett. *Modelling Survival Data in Medical Research*. Chapman and Hall, Boca Raton, second edition, 2003. ISBN 1-58488-325-1.

R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London*, 222:309–368, 1922. doi: 10.1098/rsta.1922.0009.

C. Gini. Measurement of inequality of incomes. *The Economic Journal*, 31(121):124–126, 1921.

E. Kreyszig. *Advanced Engineering Mathematics*. John Wiley and Sons, eighth edition, 1999. ISBN 0-471-33328-X.

A. S. Macdonald, S. J. Richards, and I. D. Currie. *Modelling Mortality with Actuarial Applications*. Cambridge University Press, Cambridge, 2018. ISBN 978-1-107-04541-5.

P. McCullagh and J. A. Nelder. *Generalized Linear Models*, volume 37 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London, second edition edition, 1989. ISBN 0-412-31760-5.

B. J. T. Morgan and D. M. Smith. A note on Wadley's problem with overdispersion. *Journal of the Royal Statistical Socieity, Series C*, 41(2):349–354, 1992.

S. J. Newman. Errors as a primary cause of late-life mortality deceleration and plateaus. *PLoS Biology*, 16(12):e2006776, 2018a. doi: https://doi.org/10.1371/journal.pbio.2006776.

S. J. Newman. Plane inclinations: A critique of hypothesis and model choice in Barbi et al. *PLoS Biology*, 16(12):e3000048, 2018b. doi: https://doi.org/10.1371/journal.pbio.3000048.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL `https://www.R-project.org/`.

S. J. Richards. Applying survival models to pensioner mortality data. *British Actuarial Journal*, 14(II):257–326 (with discussion), 2008. doi: https://doi.org/10.1017/S1357321700001720.

S. J. Richards. A Hermite-spline model of post-retirement mortality. *Scandinavian Actuarial Journal*, 2020:2:110–127, 2020. doi: 10.1080/03461238.2019.1642239.

S. J. Richards, S. J. Ramonat, G. Vesper, and T. Kleinow. Modelling seasonal mortality with individual data. *Scandinavian Actuarial Journal*, pages 1–15, 2020. doi: 10.1080/03461238.2020.1777194.

G. E. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6 (2):461–464, 1978.

The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) — China, 2020. *China CDC Weekly*, 2:113, 2020. ISSN 2096-7071. URL `http://weekly.chinacdc.cn//article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51`.

G. B. Wetherill. *Elementary Statistical Methods*. Chapman and Hall, third edition, 1982. ISBN 0-412-24000-9.

R. C. Willets. Mortality in the next Millennium. *Staple Inn Actuarial Society, London*, 1999.

An interactive online tool to explore the functional forms presented in this paper can be found at https://www.longevitas.co.uk/site/Hermite/HermiteAmount.html

Appendices

A Parameters

There are two kinds of parameters to be set for a Hermite model of mortality by age: (i) configuration parameters, whose values are decided in advance by the analyst, and (ii) parameters whose values are estimated from the data.

A.1 Parameters set by the analyst

Table 12 sets out the configuration parameters that are set in advance by the analyst, i.e. they are not estimated from the data. The values used in the main body of the paper are given.

Table 12: Configuration parameters for the Hermite model family.

Parameter	Value	Description and role
x_0	50	Age below which $\log \mu_x$ is deemed constant in age; see equation (1).
x_1	105	Age above which $\log \mu_x$ is deemed constant in age.

In addition to these values, the analyst must also decide which transform function to use; see Figure 9.

A.2 Parameters estimated from the data

Table 13 sets out the parameters whose values are estimated from the data, i.e. by maximising the log-likelihood function in equation (2).

Table 13: Overview of parameters.

Parameter	Name	Description and role of parameter
α	Intercept	log(mortality) for lives aged $x \leq x_0$; see equation (4).
ω	Oldest	log(mortality) for lives aged $x \geq x_1$.
α^{female}	Gender.F	Addition to log(mortality) for females aged $x \leq x_0$; see equation (4).
ω^{female}	Oldest:Gender.F	Addition to log(mortality) for females aged $x \geq x_1$.
m_0	AgeGradientYoungest	Gradient of log(mortality) leaving age x_0; see equation (1).
m_0^{amount}	AmountGradientInitial	Initial direction of log(mortality) for lives with zero benefit amounts; see equation (12).
m_1^{amount}	AmountGradientUltimate	Gradient of log(mortality) approaching maximum benefit effect, ω^{amount}; see equation (12).
λ_0	AmountTransformParameter	Parameter used in mapping benefit amount onto $[0, 1)$; see Figure 9.
ω^{amount}	AmountUltimate	Ultimate effect of infinite benefit amount on log(mortality) relative to those with zero amounts; see equation (12).

Contact

More information including case studies, latest features, technical documentation and demonstration videos can be found on our website at www.longevitas.co.uk

24a Ainslie Place, Edinburgh, EH3 6AJ
Telephone 0131 315 4470
Email info@longevitas.co.uk

LONGEVITAS™