A HERMITE-SPLINE MODEL OF POST-RETIREMENT MORTALITY

By S. J. Richards

LONGEVITAS™

A Hermite-spline model of post-retirement mortality

Richards, Stephen J.*†

July 7, 2019

Abstract

We present a model for post-retirement mortality where differentials automatically reduce with increasing age, but without the fitted mortality rates for subgroups crossing over. Selection effects are catered for, as are age-modulated time trends and seasonal variation in mortality. Central to the model are Hermite splines, which permit parsimonious modelling of complex risk factors in even modest-sized portfolios. The model is therefore suitable for the stand-alone analysis of experience data for reinsurance, bulk annuities and longevity swaps. We also illustrate the contrast between the statistical significance of a risk factor and its financial significance, and discuss reasons why one might include risk factors like season that are not directly financially significant.

Keywords: mortality convergence, mortality crossover, selection effects, age-modulated time trend, seasonal variation.

1 Introduction and motivation

In a macro-economic environment of low discount rates, the current and future mortality rates of a portfolio of insured lives assume greater actuarial importance. This is particularly the case where an insurer takes on a block of liabilities *en masse*; examples include Part VII transfers of insured liabilities from one UK life insurer to another, insurers writing large bulk annuities or reinsurers transacting longevity swaps. In each case the (re)insurer has one chance at outset to correctly underwrite the longevity risk of the lives covered; if this assessment fails, the (re)insurer will lose money unless future mortality improvements are lower than priced for. The profitability of such transactions therefore depends on the effective analysis of the portfolio's mortality experience, which is the subject of this paper. Projections of future mortality rates are dealt with elsewhere, see for example Cairns et al. [2009].

We will take the stages of data preparation as given; interested readers can consult Richards [2008] or Macdonald et al. [2018, pp20–44] for a detailed exposition of the steps necessary to validate, deduplicate, profile and check the individual records of a portfolio. The first step thereafter is to identify attributes that explain mortality differentials; see Madrigal et al. [2011] and Richards et al. [2013] for examples of the risk factors available in different portfolios and territories. However, all of these various risk factors share an important feature: their impact varies with age. Specifically, the impact of most post-retirement risk factors decreases with age to the point where the differential essentially vanishes. Indeed, the stronger the initial differential, the faster the rate of convergence. This phenomenon is known to demographers as the compensation law of mortality; see Gavrilov and Gavrilova [1991] for discussion and Gavrilov and Gavrilova [2001] for a theoretical justification. One exception is gender, where the mortality differential between males and females still reduces with age, but never entirely disappears, even at the most advanced ages [Tickle, 1997].

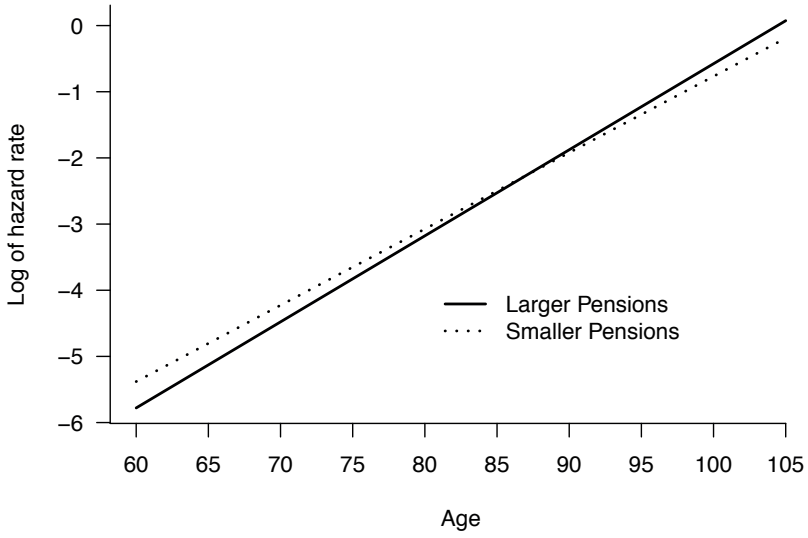
*stephen@longevitas.co.uk

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

Classical modelling of a reducing differential is achieved by fitting an age-interaction term with the main effect. However, this presents a problem, as at some age the fitted mortality rates will cross over. This is illustrated in Figure 1 for the Gompertz model [Gompertz, 1825] — by age 90, the wealthier pensioners who started out with the lightest mortality are modelled as having heavier mortality. This crossover arises purely because of the model structure, not because the data demand it. While some might argue that crossover is theoretically justified [Gavrilov and Gavrilova, 2015], in actuarial work it would not be prudent. Besides, if there is no real difference in mortality beyond age 95, as we might suspect, the structure of the model may be distorting the fit at younger ages which are of greater financial significance. The crossover phenomenon is of course not limited to the Gompertz model, nor to models for μ_x — it arises with any model of a differential that is composed of two separate parameters for the level and slope.

A second issue is that the estimated parameters for mortality level and age slope are usually highly correlated; see Richards et al. [2013]. In particular, the two parameters tend to be correlated such that the mortality differential between two sub-groups (apart from gender) is zero at some point around age 95 or so. This suggests a high degree of redundancy in the age-slope parameter. For example, in the case of the Gompertz model the usual approach is to have baseline mortality represented by parameters α_0 and β_0 thus:

Figure 1: Mortality crossover in $\log \mu_x$ for simple Gompertz model fitted to lives in a UK pension scheme. Source: Macdonald et al. [2018, p120].



$$\log \mu_x = \alpha_0 + \beta_0 x. \tag{1}$$

If we have a set of m risk factors, $R = \{r_1, \ldots, r_m\}$, say, then let α_{r_j} represent the difference in mortality level for risk factor r_j and let β_{r_j} represent the corresponding difference in the age slope. The Gompertz mortality hazard for life i is therefore:

$$\log \mu_{x_i} = \left(\alpha_0 + \sum_{j=1}^{m} \alpha_{r_j} z_{i,j}\right) + \left(\beta_0 + \sum_{j=1}^{m} \beta_{r_j} z_{i,j}\right) x_i \tag{2}$$

where $z_{i,j}$ is an indicator variable taking the value 1 if life i has risk factor j, and zero otherwise. In practice the convergence of mortality rates with age means that each β_{r_j} is always of the opposite sign to α_{r_j}, and usually of a magnitude such that $|\alpha_{r_j}/\beta_{r_j}| \approx 95$. Not only do the values of β_{r_j} cause mortality crossover, but they are also redundant parameters because each β_{r_j} could be replaced with $\text{sgn}(-\alpha_{r_j})/95$; the constant 95 is the age at which mortality differentials are effectively zero.

A third practical problem is signal strength: it is one thing to have enough data to detect the existence of a mortality differential, α_{r_j}, but more data are typically needed to estimate the

corresponding change with age, β_{r_j}. Thus, for smaller portfolios it is sometimes difficult to reliably estimate β_{r_j}, despite strong expectations of what its value is likely to be.

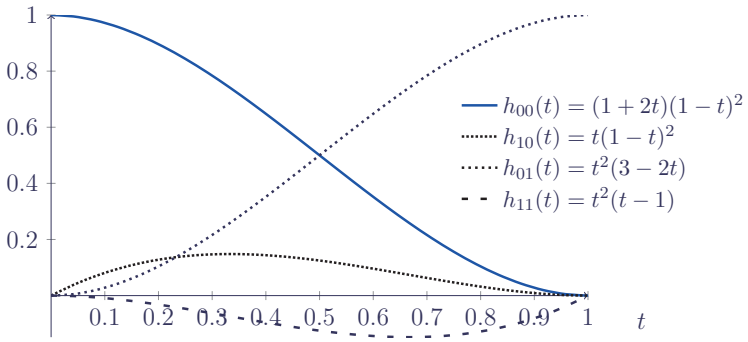
We therefore have three related issues with the classical modelling of age-varying mortality differentials: (i) crossover, (ii) parameter redundancy, and (iii) signal detection. This paper presents an alternative way to model mortality differentials that solves all three problems. In particular, the presented approach describes age-varying mortality differentials with a single parameter, not two. This eliminates parameter redundancy and, since there is no second parameter, it also avoids the problem of weak signal detection. The nature of the model is such that it eliminates crossover by default (although a determined modeller could still re-introduce it).

Section 2 of the paper gives a brief primer on Hermite splines, which are central to the paper. Section 3 shows how Hermite splines are applied to the modelling of mortality by age, including how a single parameter specifies a mortality differential with automatic convergence by age without crossover.. Section 4 considers the special case of gender, while Section 5 considers other risk factors. Section 6 shows how Hermite splines can further be used to model selection effects, while Section 7 shows how they can also be used to model age-modulated, time-based changes (improvements) in mortality. Section 8 considers how and when one might allow for seasonal variation in mortality, while Section 9 considers the contrast between the statistical and financial significance of a risk factor. Section 10 concludes the paper. A summary overview of all the model parameters is given in Appendix A, and there are some notes on implementation in Appendix B.

2 Hermite splines

A basis of Hermite splines in one dimension [Kreyszig, 1999, p868] is a collection of four cubic polynomial functions, as shown in Figure 2. Hermite splines have wide application in computer graphics for drawing a smooth path between two points. For given start and end points, $(0, p_0)$ and $(1, p_1)$, say, the set of smooth intermediate points $\{(t, p(t)), t \in [0, 1]\}$ is given by a linear combination of the Hermite basis functions shown in Figure 2:

Figure 2: Hermite basis splines for $t \in [0, 1]$.



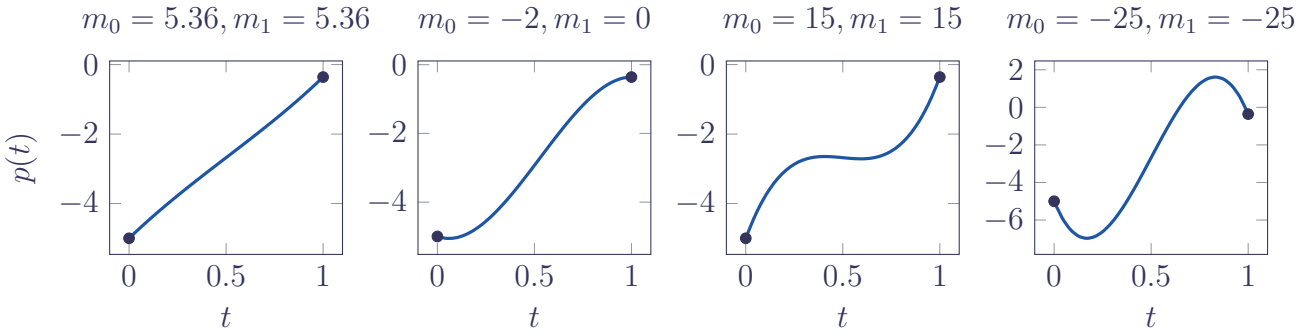
$$p(t) = p_0 h_{00}(t) + m_0 h_{10}(t) + p_1 h_{01}(t) + m_1 h_{11}(t) \tag{3}$$

where m_0 is the initial gradient of the path leaving $(0, p_0)$ and m_1 is the final gradient of the path approaching $(1, p_1)$. Changing the values of m_0 and m_1 allows a wide variety of smooth paths to be taken between the start and end points, as illustrated in Figure 3. The cubic Hermite polynomials in Figure 2 can be alternatively expressed as Bernstein polynomials of order 3; see de Boor [2001, p89], while equation (3) can also be expressed as a Bézier curve [Kreyszig, 1999, p868] with four control points expressed in terms of p_0, p_1, m_0 and m_1.

3 Applying Hermite splines to mortality modelling by age

We now turn to the application of Hermite splines to mortality modelling. The notion of drawing a smooth path between two points in Section 2 is analogous to the actuarial task of graduating a life table. If we scale the age range of interest, $[x_0, x_1]$, to the interval $[0, 1]$, then the "smooth path" is

Figure 3: Influence of m_0 and m_1 on the Hermite-interpolated path in equation (3) with $p_0 = -5$ and $p_1 = -0.36$ in each of the four panels.



the smoothed mortality hazard. We therefore define our model for the logarithm of the mortality hazard as follows:

$$\log \mu_x = \alpha h_{00}(t) + m_0 h_{10}(t) + \omega h_{01}(t) + m_1 h_{11}(t), \quad x_0 < x < x_1 \tag{4}$$

where $t = (x - x_0)/(x_1 - x_0)$ and α, ω, m_0 and m_1 are parameters to be estimated. x_0 and x_1 are limits set in advance by the modeller, e.g. the lower and upper limits of the age range for Hermite-modelled rates. The h functions are the Hermite functions defined in the legend of Figure 2. In practice most models for retirement mortality look like the second panel in Figure 3, and we use a simple extrapolation of $\log \mu_x = \alpha$ for $x < x_0$ and $\log \mu_x = \omega$ for $x > x_1$; other approaches are possible, but this assumption is simple and does not affect the crucial "no-crossover" property of the Hermite family.

Figure 3 shows that equation (4) offers plenty of flexibility to describe $\log \mu_x$ over the post-retirement age range: the first panel is a straight line, i.e a Gompertz model for $\log \mu_x$, while the second panel shows the sort of logistic curve typically seen for $\log \mu_x$ above age 50; see Perks [1932] and Beard [1959]. The third and fourth panels of Figure 3 are not realistic patterns for $\log \mu_x$, but they do demonstrate that the shape at one end of the curve is only weakly affected by flexing at the other end. Indeed, Figure 3 shows that the full, four-parameter Hermite model offers more flexibility than we strictly need for modelling late-life mortality. Accordingly, we can eliminate unnecessary flexibility by implicitly setting some parameters to zero, as shown in Table 1.

Table 1: The Hermite model family.

Model	Parameters
Hermite I	$\alpha, \omega.\ m_0 = m_1 = 0$
Hermite II	$\alpha, \omega, m_0.\ m_1 = 0$
Hermite III	$\alpha, \omega, m_1.\ m_0 = 0$
Hermite IV	α, ω, m_0, m_1

The Hermite-spline approach means that some of the parameters have a natural interpretation: if we use a logarithmic scale in Figure 3, then $p_0 = \log \mu_{x_0}$ and $p_1 = \log \mu_{x_1}$. As with many other regression-type methodologies, the Hermite-spline approach allows extrapolation outside the range of data, in particular to advanced ages like 120 to "close" a fitted life table. Indeed, bearing in mind the paucity of data above age 105, analysts can often just set a value for p_1 to reflect their own beliefs with regard to what $\log \mu_{x_1}$ is at the end of the life table; Gampe [2010] proposed a limiting value of $\log 0.7 \approx -0.36$, but see Gavrilov and Gavrilova [2015], Barbi et al. [2018], Newman [2018a] and Newman [2018b] for the vigorous ongoing debate as to whether a limiting value for $\log \mu_x$ even exists.

To compare the Hermite model fits in this paper we also need a comparison benchmark. In a survey of seventeen alternative models for the mortality experience of a large UK annuity portfolio, Richards [2012] identified the Makeham-Beard model as the best-fitting; this is a reparameterised version of the model proposed by Perks [1932]:

$$\log \mu_x = \frac{e^{\epsilon} + e^{\alpha+\beta x}}{1 + e^{\alpha+\rho+\beta x}} \tag{5}$$

where α, β, ϵ and ρ are parameters to be estimated.

The role of α in the models in equations (4) and (5) is very similar. In each case it is the value of $\log \mu_0$, and so is given the name "Intercept" (see Appendix A for further parameter-naming conventions). If we have a set of m risk factors, $R = \{r_1, \ldots, r_m\}$, say, then the operation of each main effect is achieved by building an individual value of α_i for each life thus:

$$\alpha_i = \alpha_0 + \sum_{j=1}^{m} r_j z_{i,j} \tag{6}$$

which is the same as the first term in equation (2) — α_0 is the baseline, r_j is the differential for risk factor j, and $z_{i,j}$ is a zero-one indicator variable for whether life i has risk factor j. The combination of risk factors in α_i therefore shifts $\log \mu_x$ up or down by the same amount at each age for the Gompertz model; in order to have an age-varying differential the Gompertz model needs an equivalent definition for β_i, as in equation (2). However, with the Hermite model the same definition of α_i leads to a reducing differential due to the shape of $h_{00}(t)$ in Figure 2. It is this crucial distinction that gives the Hermite model automatic convergence with a single parameter, where traditional models like Gompertz require two. To illustrate this, consider two lives who follow a Hermite mortality law, but differ only in their main effects, i.e. life 1 has $\alpha_1 = \alpha_0 + \sum_{j=1}^{m} r_j z_{1,j}$ and life 2 has $\alpha_2 = \alpha_0 + \sum_{j=1}^{m} r_j z_{2,j}$ as per equation (6). Crossover can only occur if the difference between $\log \mu_x^{(1)}$ and $\log \mu_x^{(2)}$ changes sign at some age. However, under equation (4) we have:

$$\log \mu_x^{(2)} - \log \mu_x^{(1)} = (\alpha_2 - \alpha_1) h_{00}(t), \quad x_0 < x < x_1 \tag{7}$$

where $t = (x - x_0)/(x_1 - x_0)$. As shown in Figure 2, $h_{00}(t)$ is a strictly positive decreasing function over (x_0, x_1), and so the differential narrows but its sign never switches. Thus, mortality crossover will not occur in the Hermite family of models when only the main effects differ. Mortality crossover can only take place in a Hermite model when a main effect is interacted with a parameter other than α in Table 8 in Appendix A.

Returning to our benchmarking exercise, in Table 2 we see that the Hermite family performs well against the Makeham-Beard model for the experience data of a medium-sized pension scheme. In contrast, the Gompertz model is a significantly poorer fit, as it under-states mortality at younger ages and over-states it at the oldest ages. The models are all in age only, i.e. without any risk factors, so the next task is to compare performance when allowing for gender. Before we do, we note in passing that the Hermite models could be fitted to population mortality data as a sequence of period models, and the parameters projected as a time series in a similar manner to the Gompertz model in Cairns et al. [2006]. This is a topic for further research.

4 Modelling the gender mortality differential

In modern developed countries females have lower mortality rates throughout life, from birth to the very oldest ages [Tickle, 1997]. However, at retirement ages the differential narrows with increasing age [Richards, 2012, Figure 5]. Gender is therefore a tricky risk factor to model, as there are three requirements: we want a clear differential, but one that reduces with age and yet never converges entirely (or at least doesn't converge until some very advanced age). In the case of the simple

Table 2: Model comparisons with age as the sole risk factor ($x_0 = 50$ and $x_1 = 110$). The AICs [Akaike, 1987] are for models fitted to the experience data of a medium-sized, public-sector pension scheme in Scotland, ages 50-105 observed over the period 2000-2008 (3,488 deaths out of 16,780 lives with 100,092 life-years of exposure time). The small-sample correction for the AIC is unnecessary for models with such a large number of observations relative to the number of parameters [Macdonald et al., 2018, Table 6.1].

Mortality law	Function form for μ_x	Parameter count	AIC
Gompertz	$e^{\alpha+\beta x}$	2	27,318.5
Makeham-Beard	$\dfrac{e^{\epsilon}+e^{\alpha+\beta x}}{1+e^{\alpha+\rho\beta x}}$	4	27,276.3
Hermite I	Equation (4) with $m_0 = m_1 = 0$	2	27,279.4
Hermite II	Equation (4) with $m_1 = 0$	3	27,280.9
Hermite III	Equation (4) with $m_0 = 0$	3	27,280.2
Hermite IV	Equation (4)	4	27,275.4

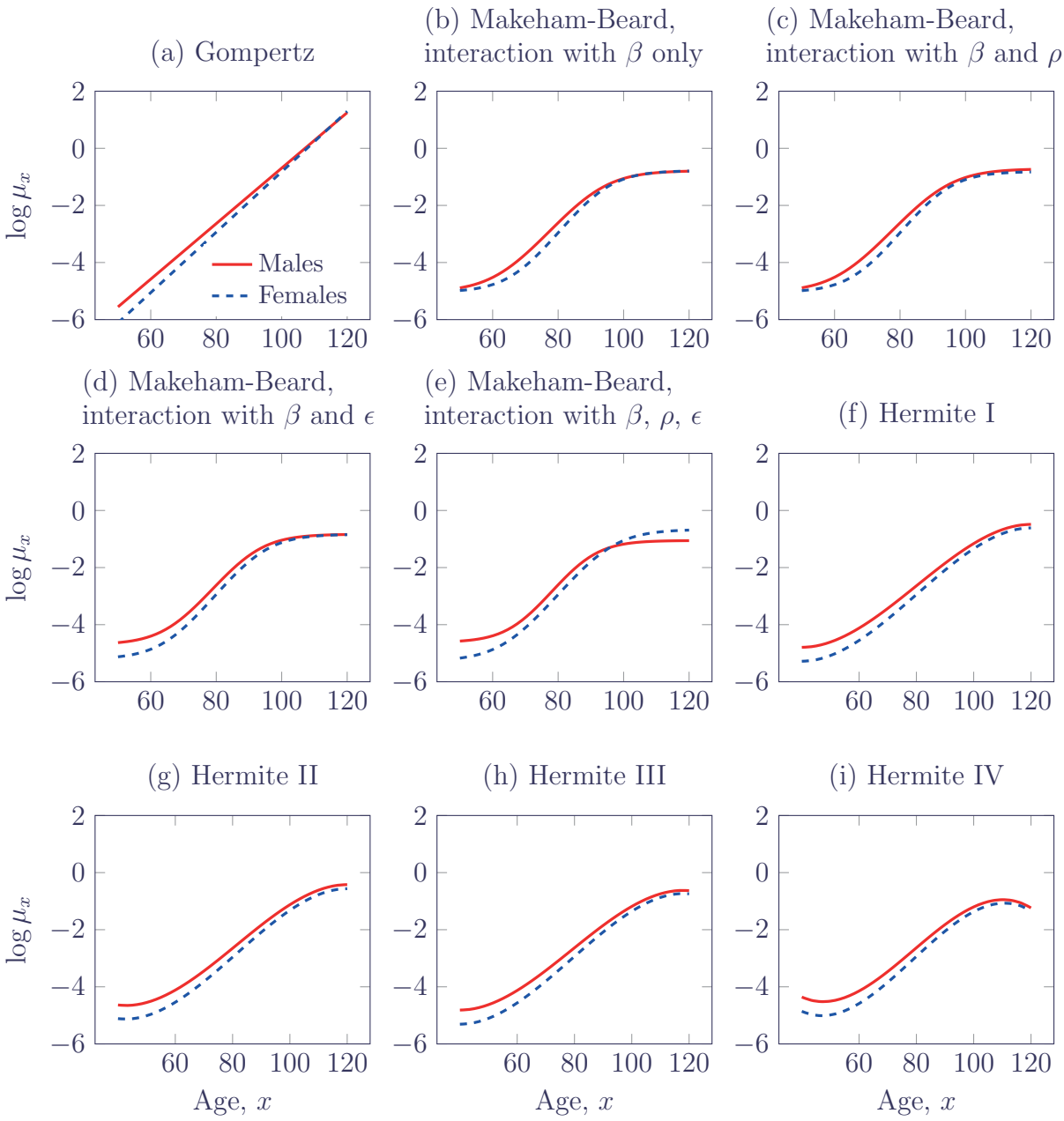
Gompertz model this is impossible — if the model has a non-zero age interaction in equation (2), then mortality rates will cross over. In the case of the Makeham-Beard model it is possible to achieve the three desired aspects of gender mortality, but only by using a model with more interactions: instead of just interacting the gender effect with the age parameter (β), we could also interact it with the Beard parameter (ρ) and the Makeham parameter (ϵ). Table 3 shows a selection of models for the gender differentials of the pensioners of a medium-sized Scottish pension scheme, while Figure 4 shows their behaviour graphically.

Table 3: Model comparisons with age and gender as risk factors ($x_0 = 50$ and $x_1 = 110$). AICs are for models fitted to the same data as in Table 2. Figure 4 shows the corresponding fitted and extrapolated mortality rates. In each case we have $\alpha_i = \alpha_0 + \alpha_{Female} z_{i,Female}$, plus the additional stated interactions.

	Mortality law	Gender interactions	Parameter count	AIC
(a)	Gompertz	$\beta_i = \beta_0 + \beta_{Female} z_{i,Female}$	4	27,233.7
(b)	Makeham-Beard	$\beta_i = \beta_0 + \beta_{Female} z_{i,Female}$	6	27,193.7
(c)	Makeham-Beard	As (b), but also $\rho_i = \rho_0 + \rho_{Female} z_{i,Female}$	7	27,195.6
(d)	Makeham-Beard	As (b), but also $\epsilon_i = \epsilon_0 + \epsilon_{Female} z_{i,Female}$	7	27,190.4
(e)	Makeham-Beard	As (c), but also $\epsilon_i = \epsilon_0 + \epsilon_{Female} z_{i,Female}$	8	27,191.1
(f)	Hermite I	$\omega_i = \omega_0 + \omega_{Female} z_{i,Female}$	4	27,192.5
(g)	Hermite II	$\omega_i = \omega_0 + \omega_{Female} z_{i,Female}$	5	27,193.2
(h)	Hermite III	$\omega_i = \omega_0 + \omega_{Female} z_{i,Female}$	5	27,194.2
(i)	Hermite IV	$\omega_i = \omega_0 + \omega_{Female} z_{i,Female}$	6	27,189.0

Figure 4(a) shows that the Gompertz model is incapable of allowing for age interactions without crossover. Figure 4(b) shows that a simple age interaction with the Makeham-Beard model has less pronounced crossover, but cannot properly express differentials at the youngest ages; interacting the gender effect with both β and ρ in Figure 4(c) is only a modest improvement. The most satisfactory Makeham-Beard model comes from an interaction of gender with both β and ϵ in Figure 4(d), while the least-satisfactory Makeham-Beard model comes from an interaction of gender with β, ρ and ϵ in Figure 4(e). This latter point is interesting because the least-satisfactory Makeham-Beard model in visual terms is one of the best-fitting of the Makeham-Beard models in Table 3. This is because goodness-of-fit statistics only measure fit where there are data, and by definition they cannot detect

Figure 4: Selected models allowing for male-female differentials only (Table 3 compares the model fits).



poor-quality extrapolations beyond the data region. An analogous point applies to projection models, as goodness of fit to the data is no guarantee of a sensible projection; see Cairns et al. [2009].

Figures 4(f)-(h) show that the simpler Hermite models produce good extrapolations, but with fewer parameters. When one considers that a difference in AIC of 4 or less is not statistically significant, Table 3 shows that the Hermite and Makeham-Beard models all produce similar fits. As expected, however, the Hermite IV model in Figure 4(i) is a little too flexible at the oldest ages, producing an undesirable reduction in mortality above age 110 or so. This is a data artefact from the small number of deaths above age 100, which could be dealt with either by restricting the age range or, as in Figures 4(f) and (g), by dropping the m_1 parameter where flexibility is unnecessary. The simpler Hermite models I and II provide the best balance of goodness of fit, extrapolation to higher ages and parsimony. As the number of risk factors in the model increases later in the paper, the parsimony of the Hermite models will become ever more apparent.

5 Modelling other mortality differentials

In Section 4 we considered the special case of gender, where differences in mortality are present at even the most advanced ages and therefore often require an interaction with the Hermite ω term to avoid these differentials vanishing. However, few mortality differentials are as persistent at high ages as gender, so for most risk factors a single parameter per main effect is all that is required in the Hermite family. Figure 5 shows a Hermite II model with age, gender and pension size-band as risk factors. As expected, the mortality differentials narrow with age but do not cross over within a given gender, as they do with traditional models. Table 4 shows just how parsimonious this model is: with just six parameters we have a model for age, gender and pension size-band, where differentials narrow with age but do not cross over. Achieving this with traditional models would take more parameters, assuming that they could even be supported by the data. However, Table 4 also shows a particular actuarial benefit of the parsimony of the Hermite model: there are just 81 deaths amongst those with the largest 5% of pensions. The mortality differential of this small-but-financially-key sub-group can only practically support a single parameter, not the two or more required by traditional models. Further risk factors can of course be added, with a single parameter allowing for age-varying differential without crossover as before.

Figure 5: Hermite II model for $\log \mu_x$ by gender and pension size. The parameter estimates underlying the curves are given in Table 4.

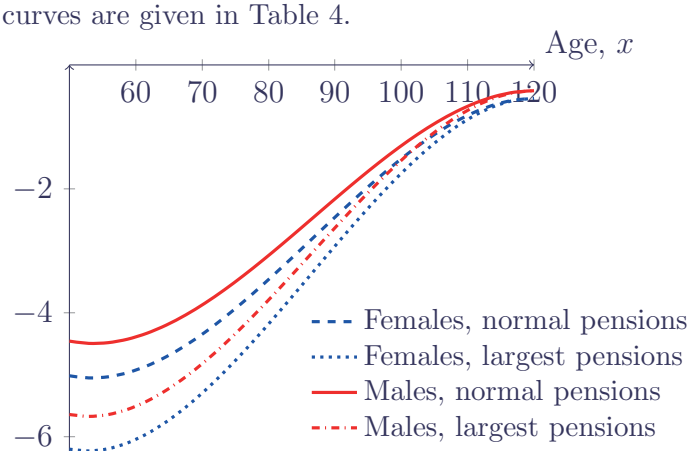
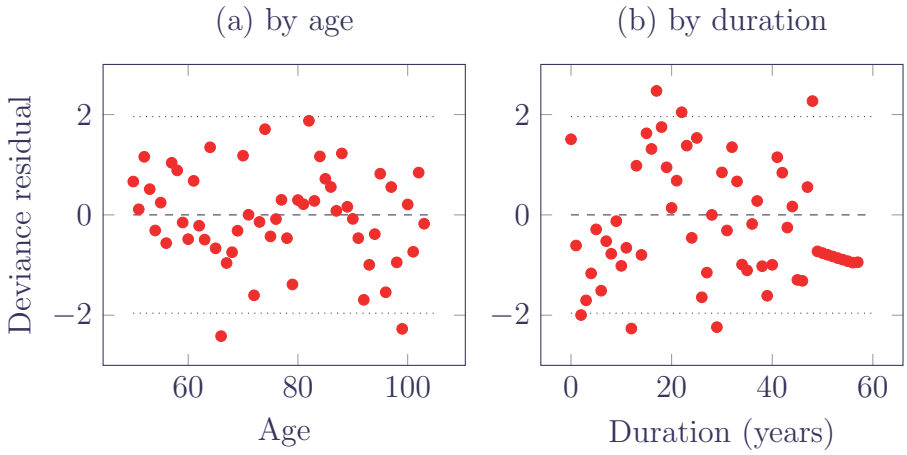


Table 4: Hermite II model for age, gender and pension size for data in Table 2. The fitted and extrapolated mortality rates are shown in Figure 5. Source: own calculations using the methodology outlined in Appendix B.3.

Parameter	Name	Estimate	Standard error	Lives	Deaths
m_0	AgeGradientYoungest	−1.49309	0.8204	16,780	3,488
α_{Female}	Gender.F	−0.559514	0.0821	9,433	1,752
ω_{Female}	Gender.F:Oldest	−0.123822	0.0909	9,433	1,752
α_0	Intercept	−4.45737	0.1445	16,780	3,488
ω_0	Oldest	−0.416038	0.0893	16,780	3,488
$\alpha_{Largest}$	Pension.Largest	−1.18428	0.1925	838	81

We use the deviance residuals [McCullagh and Nelder, 1989] to check the model fit in Table 4 — these are plotted in Figure 6 using the procedure outlined in Macdonald et al. [2018, pp100–101] to assess the fit of a survival model for individual lives. In Figure 6(a) we can see that the residuals by age are a plausible N(0,1) scatter, and so we conclude that the basic model accounts for the age-related patterns in mortality. However, Figure 6(b) shows that we still need to account for patterns in mortality since pension commencement, which we consider in Section 6.

Figure 6: Deviance residuals for the model in Table 4. The residuals are plausibly N(0,1) by age according to (i) a χ^2 test, (ii) a runs test, (iii) a standarised-deviations test test and (iv) a bias test [Macdonald et al., 2018, pages 102–108]. However, the residuals by duration fail all four tests, indicating that there are important unmodelled mortality patterns by the time since pension commencement. The data are left-truncated, and bias may be introduced from a mortality-relevant event itself being the trigger for pension commencement [Macdonald et al., 2018, pp.60–61].



6 Selection

Figure 6(b) shows that we need to extend the model to allow for mortality patterns by time since pension commencement; the negative residuals at shorter durations suggest selection effects of at least ten years. We therefore extend the Hermite model with an allowance for selection as follows:

$$\log \mu_{x,r} = \log \mu_x + \begin{cases} \theta h_{00}(u) + m_2 h_{10}(u), & r < r_1 \\ 0, & r \geq r_1 \end{cases} \tag{8}$$

where $\log \mu_x$ is defined in equation (4), r is the duration since the contract outset, $u = r/r_1$ is the duration in $[0, r_1]$ standardised to the interval $[0, 1]$ and θ, m_2 and r_1 are parameters to be estimated. θ is the initial selection (positive or negative), m_2 is the initial direction of the selection effect and r_1 is the selection term. After r_1 there is assumed to be no selection, and the gradient of the selection effect approaching r_1 is also assumed to be zero. Extending equation (7) shows that the inclusion of selection effects as per equation (8) does not affect the "no crossover" feature of the Hermite model family.

Equation (8) contains three parameters to be estimated from the data: θ, m_2 and r_1. In many circumstances it will be enough to set $m_2 = 0$ and just estimate the other two parameters. We do this for our Scottish pension scheme and find a surprisingly long selection term of $\hat{r}_1 = 21.2892$ years. The resulting modelled selection effect is plotted in Figure 7 with the updated deviance residuals by duration in Figure 8. The allowance for selection effects has improved the model fit and thus the residuals.

However, with a short period of experience data there is always a risk of conflating cohort effects [Willets, 1999] with true duration-based effects. We can allow for cohort effects by fitting the year of birth as a risk factor, but this does not materially change either the selection term or the initial selection effect. We also fitted a similar Makeham-Beard model and got a similar residual result to Figure 6(b). We therefore conclude that this pension data set does indeed seem to have a surprisingly long selection term.

Figure 7: Addition to $\log \mu_x$ in respect of selection effects. For the Scottish pension scheme $\hat{\theta} = -0.32334$ and $\hat{r}_1 = 21.2892$ (m_2 was left implicitly zero for simplicity).

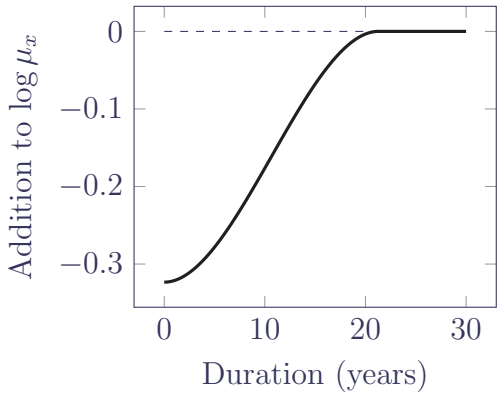
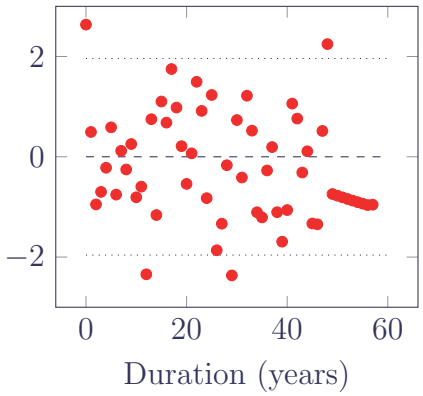


Figure 8: Deviance residuals by duration. In contrast to Figure 6(b), all four residual tests are passed, suggesting that we have allowed for the most important aspects of mortality patterns by time since pension commencement.



7 Age-modulated time trend

In Section 1 we discussed the difficulty in estimating age-varying mortality differentials with modest amounts of data. It is even trickier to estimate any time-based variation in small portfolios — exposure periods are often only a handful of years long, and the improvements we are trying to detect are weak compared with the impact of the risk factors. As with the approach to narrowing differentials by age, we need a solution that uses the minimum of parameters to robustly allow for whatever mortality improvements might be detectable in the portfolio experience data.

Figure 9 shows the annual improvement rates for males and females in the populations of the UK and the Netherlands. The shape of improvements can be broadly described as follows: (i) an initial rate of improvement at the youngest age, (ii) a peak rate of improvement at a central age, and (iii) improvements tailing off to zero at some age just above 100. We can model this by defining x_2 as the age below which mortality improvements are deemed to be constant at rate ϕ, x_3 as the age above which improvements are zero, and κ as the peak rate of change at age s_1 in the range $[x_2, x_3]$. This simple model will not capture all the flexibility exhibited in Figure 9, but it will reproduce the essential features. In any case, the subtler shape variations in Figure 9 will not be detectable with the limited amount of data typically available for most portfolios.

The Hermite model for age-modulated time trend in mortality is as follows:

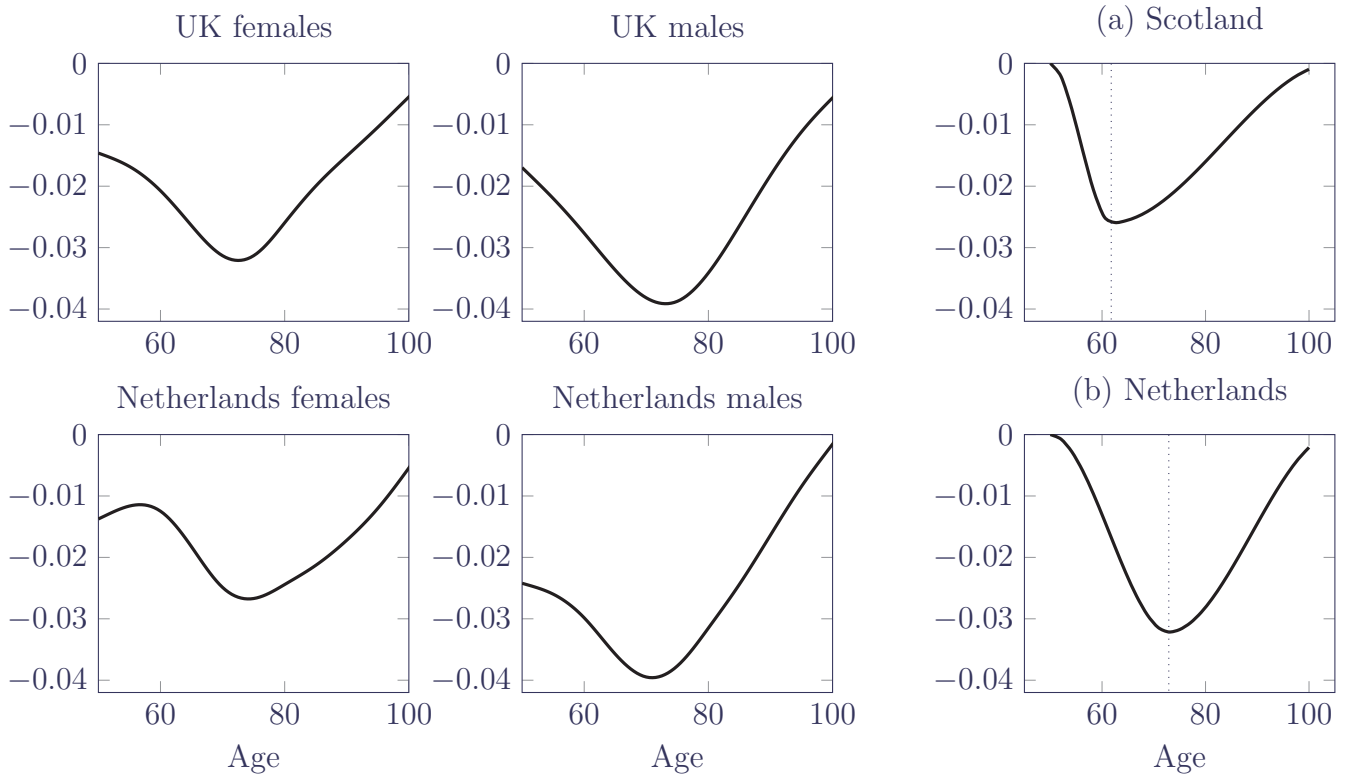
$$\log \mu_{x,r,y} = \log \mu_{x,r} + \begin{cases} \phi y, & x \le x_2 \\ \left[\phi h_{00}(v_1) + \kappa h_{01}(v_1)\right] y, & x_2 < x \le s_1 \\ \kappa h_{00}(v_2) y, & s_1 < x \le x_3 \\ 0, & x > x_3 \end{cases} \tag{9}$$

where $\log \mu_{x,r}$ is defined in equation (8) and y is calendar time in years (usually after deducting an offset like 2000 or 2010 to keep the other parameters well-scaled; see Richards [2008]). $v_1 = (x - x_2)/(s_1 - x_2)$ is the age in the range $[x_2, s_1]$ standardised to the interval $[0, 1]$, while $v_2 = (x - s_1)/(x_3 - s_1)$ is the age in the range $[s_1, x_3]$ standardised similarly. ϕ, κ and s_1 are parameters to be estimated. Extending equation (7) shows that the inclusion of an age-modulated time trend as per equation (9) does not affect the "no crossover" feature of the Hermite model family.

Equation (9) has three parameters, which may be more than can be supported by portfolios with only a modest amount of experience observed over a few years. Under such circumstances it would make sense to drop ϕ, i.e. implicitly set it to zero and just use the two-parameter age-modulated time trend involving κ and s_1. This is shown for our Scottish pension scheme in Figure 10, where the time-trend effect is weaker than the UK population time trend in Figure 9.

Figure 9: Average annual change in $\log \mu_x$ over the period 2000–2010. Difference in $\log \mu_{x,y}$ smoothed using P-splines for population data. Source: own calculations using 2D P-spline model applied to national data from Human Mortality Database.

Figure 10: Age-modulated time-trend in equation (9) for pension schemes in (a) Scotland and (b) the Netherlands. Males and females combined.



8 Seasonal variation

Seasonal variation in mortality is a well-established phenomenon in the UK and elsewhere. It is not just a feature of mortality in the colder climes of Northern Europe either — Figure 11 shows persistent seasonal variation in mortality in Australia over a twenty-year period. Seasonal mortality is highest in the winter, so the peaks in the Southern Hemisphere are shifted approximately six months relative to the peaks observed in the Northern Hermisphere. Figure 12 further shows that certain causes of death exhibit particularly strong seasonal patterns, such as circulatory and respiratory causes, while others have no meaningful seasonal variation, such as cancer. Seasonal patterns in mortality exist at all ages, but Figure 13 shows that the fluctuations are greater as age increases.

Seasonal patterns in mortality are strong, but they obviously balance over longer periods of time. Allowing for seasonal variation will therefore seldom directly impact long-term liabilities, so it might not seem worthwhile including them in a model for actuarial calculations. However, it can be useful to allow for them to permit seasonally unbalanced exposure periods when examining time trends. For

Figure 11: Percentage of average daily number of deaths in Australia, all causes, 1979–1999. All-cause winter mortality peaks at around 110-120% of the average level. Source: de Looper [2002].

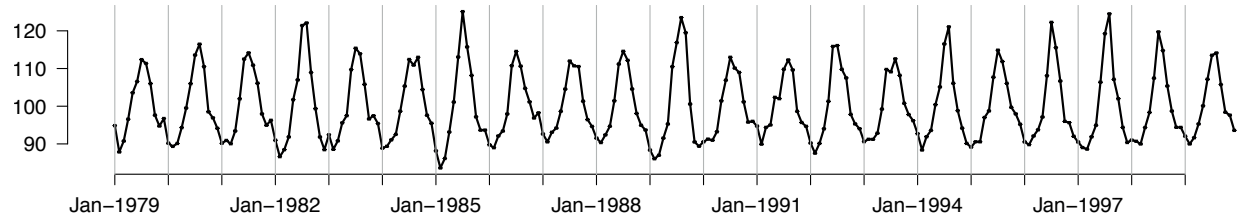


Figure 12: Percentage of average daily number of deaths for selected causes in Australia, 1979–1999. Source: de Looper [2002].

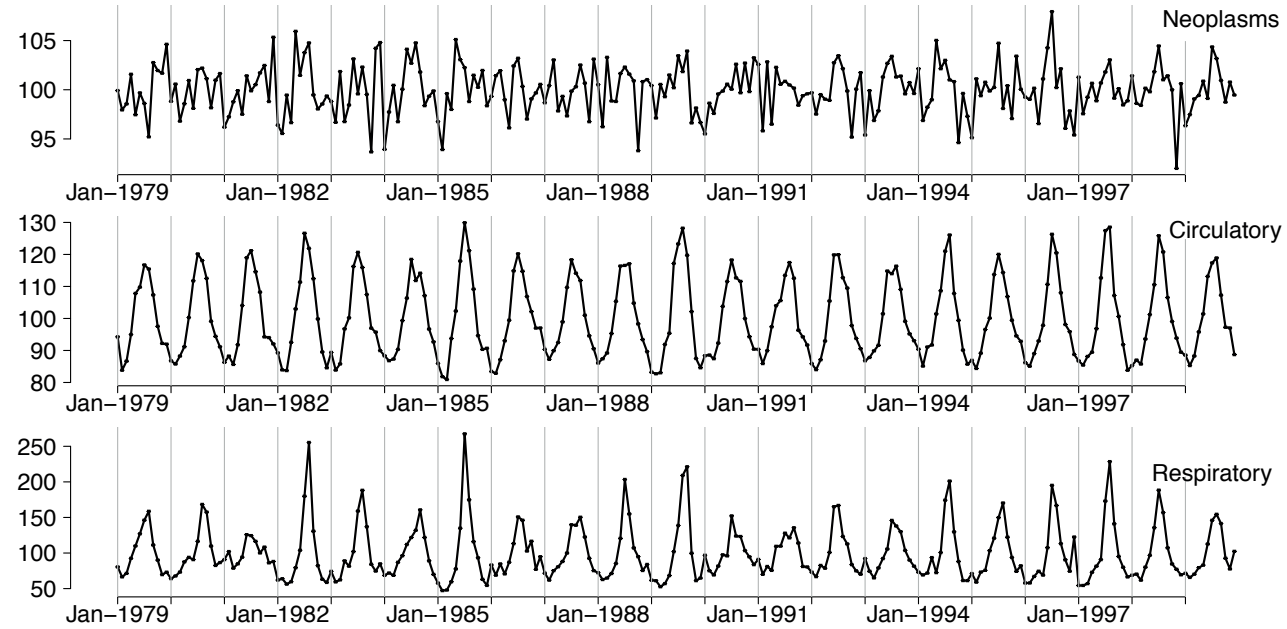
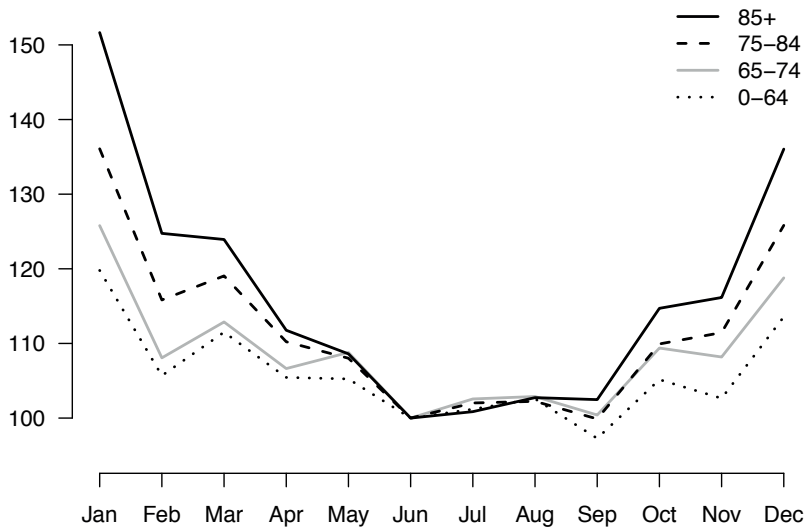


Figure 13: Deaths in England and Wales by month of occurrence and age group as percentage of June count, 2015–2017, showing that seasonal variation increases with age. Note that, as a territory in the Northern Hemisphere, the seasonal peak in mortality is offset by around half a year compared to the Southern Hemisphere in Figures 11 and 12. Source: Own calculations using data from ONS.



example, consider the example where a cedant makes experience data available from 1st January 2014 to September 2019 for a transaction planned to conclude on 31st December 2019. The reinsurer wishes to analyse the improvements in the portfolio to see if they are consistent with the improvements in the pricing basis. Figure 11 shows that an analysis without seasonal effects will need to include equal numbers of each season to avoid biasing the estimated trend. An analysis without seasonal effects would therefore compel the reinsurer to discard part of the experience data to balance the seasons, thus losing potentially useful information. However, if the analysis includes a seasonal effect, then more of the experience data can be used without biasing the estimated trend.

To allow for seasonal variation in mortality, we can define a cyclic factor for the mortality hazard. Figure 13 shows that we should perhaps consider an age-related amplitude, given the propensity for seasonal fluctuations to increase with age. However, for simplicity we start with an age-independent approach. The Hermite mortality hazard with selection, time trend and seasonal variation is therefore defined as follows:

$$\log \mu^*_{x,r,y} = \log \mu_{x,r,y} + e^{\zeta} \cos\left(2\pi(y - \tau)\right) \tag{10}$$

where τ represents the proportion of the year after January 1st when mortality peaks and where e^{ζ} is the peak additional mortality at that time (on a logarithmic scale). Extending equation (7) shows that the inclusion of an age-modulated time trend as per equation (10) does not affect the "no crossover" feature of the Hermite model family.

The definition of equation (10) is such that there will be recurrent mortality peaks at $\tau + k, k \in \mathbb{Z}$ and equivalently-sized mortality troughs at $\tau + \frac{k}{2}$. Since equation (10) is for an addition to $\log \mu_x$, this means that the mortality hazard will be multiplied by a seasonal factor that fluctuates smoothly around unity during the year. In the Northern Hemisphere $\hat{\tau}$ will typically take values in $(-0.1, 0.2)$, while in the Southern Hemisphere $\hat{\tau}$ will typically take values in $(0.4, 0.7)$. Table 5 shows the seasonal mortality for six different portfolios in Northern Europe, where the peak winter mortality of 111–122% of the baseline level corresponds closely to the amplitudes of the peaks in Figure 11.

Table 5: Seasonal peak mortality for six portfolios in Northern Europe. Source: own calculations using Hermite models for age, gender and season.

Country	Portfolio nature	Period covered	$\hat{\zeta}$	$\hat{\tau}$	Peak mortality: (a) as % of average	Peak mortality: (b) time of year
Scotland	Pension scheme	1998–2007	−1.62	0.092	122%	Feb 1st
UK	Insurer annuities	2009–2012	−2.00	0.001	114%	Dec 30th
England	Pension scheme	2002–2012	−2.02	0.071	114%	Jan 25th
Netherlands	Pension scheme	2006–2012	−2.25	0.055	111%	Jan 19th
England	Pension scheme	2000–2013	−2.29	0.048	111%	Jan 16th
UK	Insurer annuities	1998–2007	−2.27	0.086	111%	Jan 30th

The cosine approach in equation (10) is by no means the last word on modelling seasonal variation. Figure 13 shows that we could usefully extend equation (10) to include variation by age, while Figure 11 suggests that the mortality peaks in winter are spikier than the shallower summer troughs. These are both topics for future research.

9 Statistical v. financial significance

The models fitted in this paper are survival models for individual lifetimes. Although these are statistical models, there are important features of actuarial data that statisticians do not often

encounter; see Macdonald et al. [2018, pages 16–17] for an overview of the salient differences in data for actuarial applications, and how these drive specific modelling requirements.

A further important difference between actuaries and statisticians lies in what counts as a significant risk factor. Statistical significance can be defined as the change in an information criterion exceeding a threshold value, as in the third column of Table 6. However, a risk factor for mortality can be statistically significant without being financially significant. An example is the seasonal effect in Table 6, which has a large statistical impact (the AIC improves by 36.3) and yet makes the smallest change to the discounted cashflow valuation. One reason to nevertheless include all statistically significant risk factors is enhancement [Currie and Korabinksi, 1984], i.e. the phenomenon whereby the inclusion of a risk factor in a model increases the explanatory power of risk factors already present. A risk factor that is not directly financially significant on its own may nevertheless be indirectly financially significant through modulating the action of other risk factors. Enhancement also makes the changes in Table 6 dependent on the order of inclusion.

Table 6: Analysis of change in AIC and discounted cashflow valuation from adding risk factors. Cashflow valuation is of annual pensions paid to survivors, discounted using a net 0% discount rate and with no allowance for future mortality improvements, i.e. period rates until the model in the last line of the table. Source: own calculations using the same pension scheme as Tables 2-4.

		Change	PV of cashflows (£m):			Change in
Model	AIC	in AIC	Males	Females	Total	total PV
Hermite I, age only	27279.4	n/a	532.7	302.6	835.3	n/a
+gender	27192.5	−86.9	495.0	321.4	816.4	−18.9
+widow(er) status	27175.7	−16.8	495.8	328.4	824.2	7.8
+early-retirement status	27161.7	−15.9	485.6	324.6	810.2	−14.1
+pension size	27099.5	−62.2	538.5	333.6	872.1	61.9
+selection	27082.1	−17.4	534.7	328.3	863.0	−9.1
+season	27045.8	−36.3	532.3	326.8	859.1	−3.9
change from period mortality to forecast mortality:						
+age-related time trend	27044.0	−1.8	579.5	352.9	932.4	73.3

A risk factor can also be financially significant without being statistically significant: an example is the age-related time trend in Table 6, which is the least statistically significant of the risk factors, yet it has the largest impact on the valuation of all factors bar age.

10 Conclusions

The Hermite family of mortality models provides a good fit to pensioner mortality data with automatic convergence of differentials by age. The Hermite models are more parsimonious than traditional models, as they do not need a second parameter for each differential to allow for convergence by age. The simpler members of the Hermite family of models produce sensible extrapolations to ages where data are sparse or non-existent. Hermite splines additionally allow parsimonious modelling of both selection effects and age-modulated time trends in mortality rates. The modelling approach is efficient, allowing even modest-sized portfolios to have effective and robust bespoke mortality tables. The criteria used by actuaries to decide on the risk factors in a model are different to those that might be used by statisticians, however, as there is an important difference between statistical significance and financial significance.

Acknowledgments

The author thanks Wendy Ho of the AIHW for the data on seasonal mortality in Australia, together with Iain Currie, Torsten Kleinow, Angus Macdonald and Gavin Ritchie for helpful comments. Any errors or omissions remain the responsibility of the author. Models were fitted using Longevitas. Graphs were done in R, tikz and pgfplots, while typesetting was done in LaTeX.

References

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

E. Barbi, F. Lagona, M. Marsili, J. W. Vaupel, and K.W. Wachter. The plateau of human mortality: Demography of longevity pioneers. *Science*, 2018;360:1459–1461, 2018. ISSN 0036-8075. doi: 10.1126/science.aat3119.

R. E. Beard. Note on some mathematical mortality models. In G. E. W. Wolstenholme and M. O'Connor, editors, *The Lifespan of Animals*, pages 302–311. Little, Brown, Boston, 1959.

A. J. G. Cairns, D. Blake, and K. Dowd. A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, 73:687–718, 2006. doi: https://doi.org/10.1111/j.1539-6975.2006.00195.x.

A. J. G. Cairns, D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, A. Ong, and I. Balevich. A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, 13(1):1–35, 2009. doi: 10.1080/10920277.2009.10597538.

I. D. Currie and A. Korabinksi. Some comments on bivariate regression. *The Statistician*, 33:283–293, 1984.

C. de Boor. *A practical guide to splines*, volume 27 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2001. ISBN 978-0-387-95366-3.

M. de Looper. *Seasonality of death*, volume Bulletin No. 3. Australian Institute of Health and Welfare, 2002. ISBN 978–1–74024–209–7.

J. Gampe. Human mortality beyond age 110. *Supercentenarians. Demographic Research Monographs*, 2010. doi: https://doi.org/10.1007/978-3-642-11520-2_13.

L. A. Gavrilov and N. S. Gavrilova. *The Biology of Life Span: A Quantitative Approach*. Harwood Academic Publisher, 1991. ISBN 3718649837.

L. A. Gavrilov and N. S. Gavrilova. Biodemography of old-age mortality in humans and rodents. *Journal of Gerontology: Biological Sciences*, 70(1):1–9, 2015.

L. A. Gavrilov and N. S. Gavrilova. The reliability theory of aging and longevity. *Journal of Theoretical Biology*, 213:527–545, 2001. doi: https://doi.org/10.1006/jtbi.2001.2430.

B. Gompertz. The nature of the function expressive of the law of human mortality. *Philosophical Transactions of the Royal Society*, 115:513–585, 1825.

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, 2018. ISBN 978-1-107-04541-5.

A. Madrigal, F. Matthews, D. Patel, A. Gaches, and S. Baxter. What longevity predictors should be allowed for when valuing pension scheme liabilities? *British Actuarial Journal*, 16(I):1–62 (with discussion), 2011. doi: 10.1017/S1357321711000018.

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.

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.

W. Perks. On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries*, 63:12–40, 1932.

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C++: The Art of Scientific Computing*. Cambridge University Press, 2005. ISBN 0-521-75033-4.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

S. J. Richards. A handbook of parametric survival models for actuarial use. *Scandinavian Actuarial Journal*, 2012 (4):233–257, 2012. doi: https://doi.org/10.1080/03461238.2010.506688.

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. Mis-estimation risk: measurement and impact. *British Actuarial Journal*, 21 (3):429–457, 2016. doi: 10.1017/S1357321716000040. URL `http://dx.doi.org/10.1017/S1357321716000040`.

S. J. Richards, K. Kaufhold, and S. Rosenbusch. Creating portfolio-specific mortality tables: a case study. *European Actuarial Journal*, 3 (2):295–319, 2013. doi: https://doi.org/10.1007/s13385-013-0076-6.

L. G. Tickle. The sex differential in mortality rates and underlying factors. *Actuarial Studies and Demography Department Research Paper Series*, 006/97, 1997.

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

Data were downloaded on 21 May 2019 from the Human Mortality Database (HMD). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de.

An interactive online tool to explore the functional forms presented in this paper can be found at http://www.richardsconsulting.co.uk/Hermite/

Appendices

A Overview of model parameters

There are two kinds of parameters to be set for a Hermite model of mortality: (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 7 sets out the configuration parameters that need to be 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 7: 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 Section 3
x_1	110	Age above which $\log \mu_x$ is deemed constant in age.
x_2	50	Age below which $\log \mu_x$ changes in time at constant value ϕ.
x_3	105	Age above which $\log \mu_x$ does not change in time; see Section 7.

A.2 Parameters estimated from data

Table 8 sets out the parameters whose values are estimated from the data, e.g. by maximising a likelihood function such as equation (13) in Appendix B.

Table 8: Estimated parameters in the Hermite model family.

Parameter	Name	Description and role
α	Intercept	log(mortality) at and below age x_0; see Section 3.
ω	Oldest	log(mortality) at and above age x_1.
m_0	AgeGradientYoungest	Gradient of log(mortality) leaving age x_0.
m_1	AgeGradientOldest	Gradient of log(mortality) approaching age x_1.
r_1	SelectionTerm	Duration in years after which the selection effect is zero.
θ	SelectionInitial	Initial selection effect at duration zero; see Section 6.
m_2	SelectionGradient	Gradient (direction) of initial selection effect from duration zero.
ϕ	TrendYoungest	Rate of change in time for mortality at age x_2 and below; see Section 7.
κ	TrendPeak	Peak rate of change in time for mortality at age s_1.
s_1	TrendPeakAge	Age at which rate of change of mortality peaks with value κ.
ζ	SeasonalExcess	Amplitude of seasonal peak mortality from baseline (log scale); see Section 8.
τ	SeasonalPeak	Time of year of peak seasonal (winter) mortality, expressed as fraction of year from 1$^{\text{st}}$ January.

B A general approach to fitting survival models

The models fitted in this paper are for the lifetimes of individuals, rather than grouped counts; the benefits of doing this are discussed in Macdonald et al. [2018, pp 11–15]. One aspect of fitting models to individual lifetimes is the requirement to integrate the hazard function. For many mortality models this can be done explicitly; see Richards [2012, Tables 1 and 2] for the integrated hazard functions for seventeen different models. However, not every integrated hazard function can be expressed in a closed form, so numerical integration is sometimes required. This appendix describes how to fit any survival model to any required level of accuracy, not just the Hermite models in this paper. All that is required is a small number of partial derivatives of a single function, $\log \mu_x$, and a suitable algorithm for numerical integration.

B.1 Introduction

Macdonald et al. [2018, p78] give the likelihood for a survival model based on n individual lifetimes as follows:

$$L \propto \prod_{i=1}^{n} \exp\left(-\int_0^{t_i} \mu_{x_i+s} ds\right) \mu_{x_i+t_i}^{d_i} \tag{11}$$

where μ_{x_i} is the instantaneous mortality hazard for life i aged exactly x_i at the start of observation. d_i is an indicator variable taking the value 1 if life i is dead at age $x_i + t_i$, and zero otherwise. The data are therefore left-truncated and right-censored.

It is usually more convenient to work with the log-likelihood function [Macdonald et al., 2018, p79]:

$$\ell = -\sum_{i=1}^{n} \int_0^{t_i} \mu_{x_i+s} ds + \sum_{i=1}^{n} d_i \log \mu_{x_i+t_i} \tag{12}$$

Note that equations (11) and (12) are entirely specified by μ_x. Also, no specific form for μ_x is required; the equations hold irrespective of the mortality law that applies.

B.2 Maximising the (log-)likelihood

To maximise the log-likelihood we want the gradients, i.e. the first partial derivatives with respect to each parameter. We start by noting that equation (12) can be rewritten as follows:

$$\ell = -\sum_{i=1}^{n} \left(\int_0^{t_i} e^{\log \mu_{x_i+s}} ds\right) + \sum_{i=1}^{n} d_i \log \mu_{x_i+t_i} \tag{13}$$

Equation (13) emphasises that the entire log-likelihood depends only on $\log \mu_x$. In particular, even if there is not a closed-form expression for the integral in equation (13), the log-likelihood can be evaluated to any required level of accuracy using quadrature; see Press et al. [2005, pages 133–166], Kreyszig [1999, pages 876-878], R Core Team [2013, `integrate()` function] or Macdonald et al. [2018, pages 344–345].

The first partial derivative of equation (13) with respect to any parameter, θ, is therefore:

$$\frac{\partial}{\partial \theta}\ell = -\sum_{i=1}^{n} \left[\int_0^{t_i} e^{\log \mu_{x_i+s}} \left(\frac{\partial}{\partial \theta} \log \mu_{x_i+s}\right) ds\right] + \sum_{i=1}^{n} d_i \frac{\partial}{\partial \theta} \log \mu_{x_i+t_i} \tag{14}$$

Equation (14) holds because we can differentiate through the integral in equation (13), since the limits of integration do not depend on the parameter being differentiated. As before, knowledge of $\log \mu_x$ and $\frac{\partial}{\partial \theta} \log \mu_x$ completely specifies the first partial derivative, and the integral in equation (14) can be evaluated to any required level of accuracy using quadrature. All that is required is to work out the first partial derivative of $\log \mu_x$ with respect to each parameter. In the case of the Hermite model in equation (4), $\log \mu_x$ is linear in all the parameters and so these first partial derivatives are trivial.

B.3 Information matrix (and thus covariance matrix)

We approximate the covariance matrix by inverting the negative information matrix [Richards, 2016, Section 5]. We evaluate the information matrix as the Fisher information, i.e. the matrix of second partial derivatives of the log-likelihood function [Macdonald et al., 2018, pp 83, 86]. We can repeat the trick of differentiating equation (14) through the integral to get any second partial derivative with respect to parameters θ_1 and θ_2 as follows:

$$\frac{\partial^2}{\partial \theta_1 \partial \theta_2} \ell = -\sum_{i=1}^{n} \left[\int_0^{t_i} e^{\log \mu_{x_i+s}} \left[\frac{\partial^2}{\partial \theta_1 \partial \theta_2} \log \mu_{x_i+t_i} + \left(\frac{\partial}{\partial \theta_1} \log \mu_{x_i+s} \right) \left(\frac{\partial}{\partial \theta_2} \log \mu_{x_i+s} \right) \right] ds \right] + \sum_{i=1}^{n} d_i \frac{\partial^2}{\partial \theta_1 \partial \theta_2} \log \mu_{x_i+t_i} \quad (15)$$

We already have $\log \mu_x$ and the first partial derivatives of $\log \mu_x$ from the gradient calculation in equation (14). All that are additionally required are the second partial derivatives of $\log \mu_x$. As before, we can use quadrature to evaluate the information matrix to any required level of accuracy, even if we cannot derive closed-form expressions for the integral in equation (15). In the case of the Hermite model in equation (4), $\log \mu_x$ is linear in all the parameters, so the mixed second partial derivatives are simply zero.

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™