British Actuarial Journal (2022), Vol. **27**, e1, pp. 1–22
doi:10.1017/S1357321721000180

SESSIONAL PAPER

Allowing for shocks in portfolio mortality models

Stephen J. Richards

Longevitas Ltd, Edinburgh, EH3 6AJ, UK.
E-mail: stephen@longevitas.co.uk, www.longevitas.co.uk.

Abstract

The COVID-19 pandemic creates a challenge for actuaries analysing experience data that include mortality shocks. Without sufficient local flexibility in the time dimension, any analysis based on the most recent data will be biased by the temporarily higher mortality. Also, depending on where the shocks sit in the exposure period, any attempt to identify mortality trends will be distorted. We present a methodology for analysing portfolio mortality data that offer local flexibility in the time dimension. The approach permits the identification of seasonal variation, mortality shocks and occurred-but-not reported deaths (OBNR). The methodology also allows actuaries to measure portfolio-specific mortality improvements. Finally, the method assists actuaries in determining a representative mortality level for long-term applications like reserving and pricing, even in the presence of mortality shocks. Results are given for a mature annuity portfolio in the UK, which suggest that the Bayesian information criterion is better for actuarial model selection in this application than Akaike's information criterion.

Keywords: COVID-19; seasonal variation; mortality trends; *B*-splines; OBNR

1. Introductionc

The COVID-19 pandemic (The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, 2020) creates the need to allow for mortality shocks in experience analysis performed by actuaries. The intense nature of the repeated COVID-19 mortality shocks in many countries means that traditional methods based around annual q_x-style mortality rates are inadequate: mixing periods of shock and non-shock mortality understates the true intensity of mortality spikes. Furthermore, mortality shocks in the experience data may lead to bias in bases for reserving and pricing. A continuous-time methodology that copes with sharp fluctuations in time is required.

This paper covers the modelling and analysis of portfolio experience data only. The subject of future mortality trends is out of scope, although section 9 briefly considers the topic of year-on-year improvements. In our models, we make extensive use of splines, which are flexible mathematical functions. There are numerous different kinds of spline, each with different properties and thus suitable for different purposes. In this paper, we use two kinds of spline: Hermite splines, which span the interval [0,1], and Schoenberg (1964) splines, which are piecewise local polynomials on the real line. In modern literature, such as de Boor (2001) and Eilers & Marx (2021), references to *B*-splines are synonymous with Schoenberg splines. Since this paper uses two different kinds of spline, we use the term Schoenberg spline, rather than *B*-spline, to distinguish from Hermite splines.

Past mortality modelling using splines in calendar time, such as Currie *et al.* (2006) and Eilers *et al.* (2008), uses stratified grouped counts. In contrast, we use individual records in a survival model so that we can include covariates like gender, pension size and other factors typically

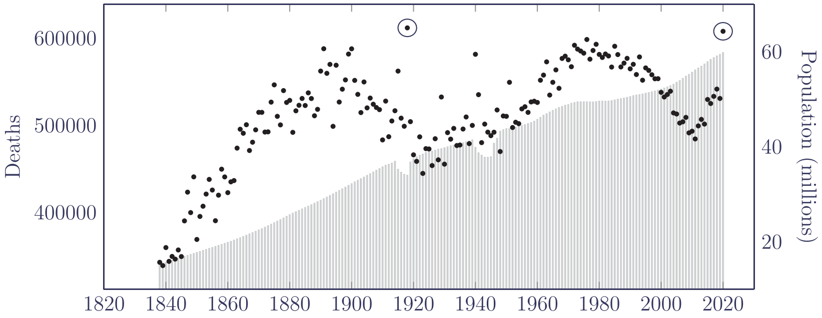

Figure 1. Numbers of deaths in England & Wales (●, left scale) with 1918 and 2020 counts circled. The population size is also shown in grey (right scale). Source: ONS.

available to actuaries. We use Hermite splines to model age- and income-related mortality and we use Schoenberg (1964) splines to model mortality in time. Here, we compress the knot spacing for splines in time to half a year or less to model sharp swings in mortality.

The plan of the rest of this paper is as follows: section 2 presents important features of COVID-19 mortality in the UK and the resulting need for continuous time methods in place of annual q_x rates; section 3 describes the data set used in the paper; section 4 describes the use of Hermite splines for modelling mortality by age, while section 5 describes how to use Hermite splines for modelling mortality by annuity amount. Section 6 recaps the application of Schoenberg splines for modelling mortality levels over time. Section 7 considers the ability of the method to identify seasonal variation, while section 8 looks at the additional requirements for modelling the COVID-19 mortality shocks in the UK. Section 9 looks at further insights that can be derived, such as portfolio-specific improvement rates. Section 10 considers the use of the methodology to allow for the impact of occurred-but-not-reported (OBNR) deaths. Section 11 considers the conditions under which Schoenberg (1964) splines may – and may not – be used for mortality patterns by age. Section 12 discusses the use and limitations of information criteria, and the role of actuarial judgement in selecting models. Section 13 concludes.

2. COVID-19 and Other Mortality Shocks

COVID-19 is a new viral disease whose arrival in the UK in 2020 caused deaths to surge to levels not seen since the global influenza pandemic of 1918–1920 (Spreeuwenberg *et al.*, 2018). Figure 1 shows that England & Wales had sharply higher numbers of deaths in 1918 and 2020 compared to preceding years.

In many places, the 1918–1920 influenza pandemic appeared as a series of three spikes in mortality, as shown in Figure 2 for Scotland, where the second and third spikes were the most severe. As with influenza in 1918–1920, a key feature of COVID-19 mortality in many countries is that it too takes the form of relatively sharp peaks. These peaks are a result of accelerating infection, followed by changes in health policy and behaviour causing infections to fall. For example, without the lockdown in England on 26 March 2020 (Public Health England, 2020), the first COVID-19 peak would have been higher (and if the lockdown had been declared earlier, the peak would likely have been lower). At the time of writing, there have been two such peaks in the UK, as shown in Figure 3. These peaks form and subside over a period of weeks or months, and an annualised approach to mortality will understate such mortality surges. For this reason, it is important to use continuous time methods like the mortality hazard, μ_x, rather than annual q_x-style rates.

COVID-19 made its presence felt in the mortality of annuity portfolios and pension schemes in many countries; Richards (2021b) demonstrated this for annuity portfolios in France, the UK and

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

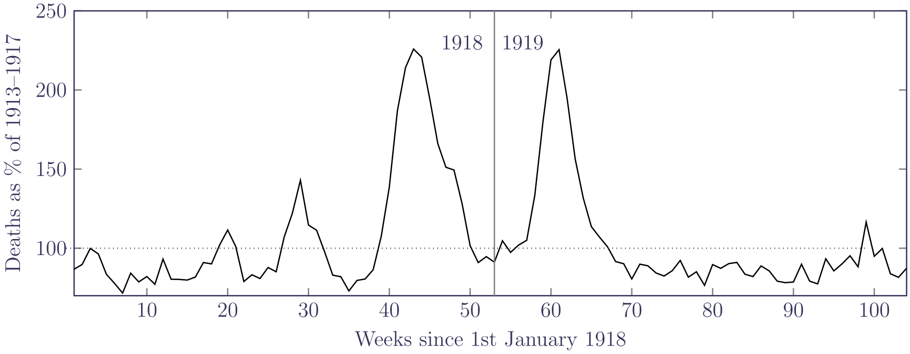

Figure 2. Weekly deaths in Scotland in 1918–1919 as percentage of 1913–1917 average. Source: Craufurd Dunlop and Watt (1915, 1916a, 1916b, 1918, 1919, 1920a, 1920b).

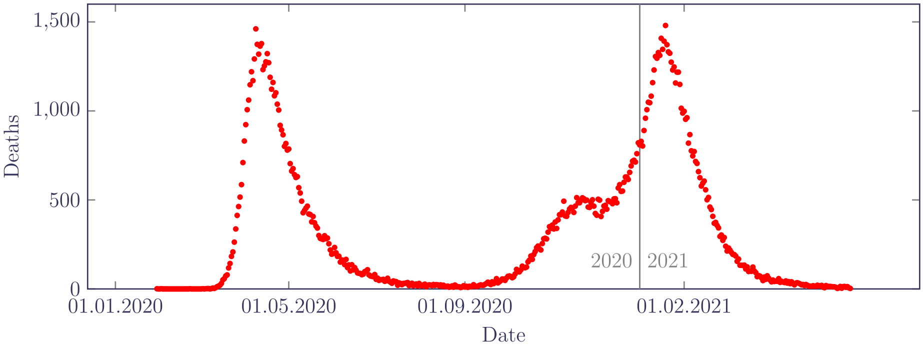

Figure 3. Daily deaths in UK where the death certificate mentions COVID-19 as one of the causes. Note that the first lockdown legally commenced in England on 26 March 2020. Source: ONS (2021).

the USA. COVID-19 therefore creates a problem for actuaries analysing experience data to set bases: the extra mortality of 2020–2021 risks either under-reserving (in the case of liabilities already on the balance sheet) or under-pricing (in the case of insurers writing bulk annuities and longevity swaps). Unfortunately, portfolio administrators seldom record the cause of death, so data like Figure 3 are typically unavailable. Excluding periods of COVID-affected data is also an unsatisfactory approach – pension schemes looking to transact bulk annuities or longevity swaps may only have experience data for the most recent 3–5 years, and discarding one or more of those years' data is an unaffordable luxury for a pricing actuary. Actuaries therefore require a methodology that works with all-cause mortality data, but which flexibly tracks mortality levels over time to allow for COVID-19 spikes. This way all available experience data can be used, and the actuary can then exercise judgement as to what point in time is most representative for calibrating a mortality basis.

3. Data Description

The data used in this paper comprise individual records from a mature annuity portfolio of the UK insurer. At the end of June 2021, a direct extract was made from the administration systems, as recommended by Macdonald *et al.* (2018, section 2.2) to get the most up-to-date data. Policy records were validated using the checks described in Macdonald *et al.* (2018, sections 2.3 and 2.4).

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

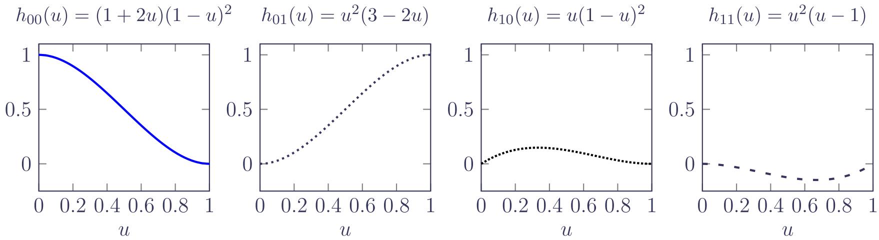

Figure 4. Four Hermite basis splines for $u \in [0,1]$. Note that various potentially useful equalities exist, such as $h_{00}(u) + h_{01}(u) = 1$ and $h_{10}(u) + h_{11}(1-u) = 0$.

Insurer annuity portfolios often contain two or more policies per person, and so deduplication is required for the independence assumption that underpins statistical modelling; see Macdonald *et al.* (2018, section 2.5) for discussion of various approaches to deduplication. As is common for UK annuity portfolios, there are many such duplicates – of 351,947 records passing validation, 124,420 records were found to be someone's second or third annuity, and there were 3 people with 29 annuities each. The tendency to have multiple annuities is correlated with wealth and socio-economic status, so deduplication is an essential step in building a statistical model for actuarial purposes. There were 729 annuities where the annuitant was marked alive on one annuity and dead on another; these annuities were excluded from the data for modelling.

The data set used in this paper is an updated extract of the UK3 data set in Richards (2021b). In this paper, we will continue to refer to it as UK3 for continuity.

4. Hermite Splines for Mortality by Age

A basis of Hermite splines in one dimension (Kreyszig, 1999, p. 868) is a collection of four cubic polynomial functions, as shown in Figure 4. A basis of cubic Hermite splines will produce the same fitted curve as a basis of cubic Bézier curves or Bernstein polynomials of degree 3, albeit with different coefficients.

Richards (2020) proposed using Hermite splines for modelling mortality by mapping an age range $[x_0, x_1]$ onto $[0, 1]$ and forming the log hazard as a linear combination of the spline functions in Figure 4. The model for the mortality hazard at age x, μ_x, is given in equation (1):

$$\begin{aligned} \log \mu_x &= \alpha h_{00}(u) + \omega h_{01}(u) + m_0 h_{10}(u) + m_1 h_{11}(u) \\ u &= \frac{x - x_0}{x_1 - x_0} \end{aligned} \tag{1}$$

where the intermediate variable u maps age $x \in [x_0, x_1]$ onto $[0, 1]$ and α, ω, m_0 and m_1 are parameters to be estimated. In practice, the h_{11} spline is seldom needed for mortality work and so we often set $m_1 = 0$. Doing so forces the mortality hazard in equation (1) to be monotonic at advanced ages, a topic we will return to in section 11. We can further set $m_0 = 0$ to obtain a simple two-parameter alternative to the Gompertz (1825) model that is strictly monotonic at all ages, as per equation (2):

$$\log \mu_x = \alpha h_{00}(u) + \omega h_{01}(u) \tag{2}$$

Actuarial survival data are typically left-truncated; see Macdonald *et al.* (2018, section 1.9) for a discussion of the differences between actuarial survival models and those in medical research, such as described by Collett (2003). To fit a survival model for n individual lifetimes, we use the log-likelihood, ℓ, in equation (3):

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

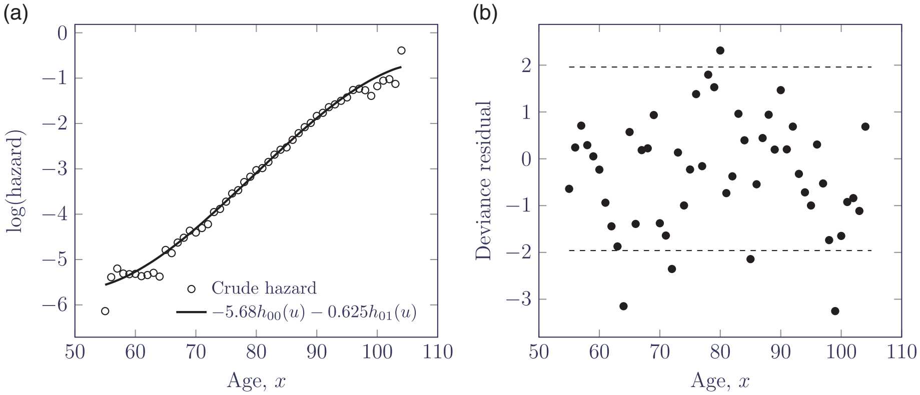

Figure 5. (a) Crude mortality hazard (○) for UK3 data set with fitted curve from equation (2). (b) Deviance residuals (●) from the model fit with dashed lines showing the 95% confidence limits for N(0,1) variates. Data cover ages 55–105 years across the period 2015–2020.

$$\begin{aligned} \ell &= -\sum_{i=1}^{n} H_{x_i}(t_i) + \sum_{i=1}^{n} d_i \log \mu_{x_i+t_i} \\ H_{x_i}(t_i) &= \int_0^{t_i} \mu_{x_i+s} ds \end{aligned} \tag{3}$$

where x_i is the exact age when life $i = 1, \ldots, n$ enters observation, t_i is the time in years that life i is observed and d_i is an indicator variable taking the value 1 if life i is dead at age $x_i + t_i$ and 0 otherwise; $H_{x_i}(t_i)$ is the integrated hazard function. The log-likelihood in equation (3) is maximised to find the maximum-likelihood estimates of the parameters underlying μ_x; see Appendix 14 for the technical details of implementation and model-fitting.

Figure 5(a) shows the simplified Hermite spline model of equation (2) fitted to the UK3 data set. Figure 5(b) shows the deviance residuals (McCullagh & Nelder, 1989, p. 39), indicating that Hermite splines are an effective approach to modelling post-retirement mortality, especially where the linear assumption of Gompertz (1825) does not hold.

Figure 5 shows that equation (2) does an acceptable job of explaining mortality variation by age. We therefore extend equation (2) to include gender as a risk factor as shown in equation (4), where $z_{i,\text{female}}$ is an indicator variable taking the value 1 if the annuitant is female and 0 if male. α_0, α_{female}, ω_0 and ω_{female} are parameters estimated by maximum likelihood using equation (3).

$$\begin{aligned} \log \mu_{x_i} &= \alpha_i h_{00}(u) + \omega_i h_{01}(u) \\ \alpha_i &= \alpha_0 + \alpha_{\text{female}} z_{i,\text{female}} \\ \omega_i &= \omega_0 + \omega_{\text{female}} z_{i,\text{female}} \end{aligned} \tag{4}$$

5. Hermite Splines for Mortality by Annuity Amount

Figure 5 shows that even just using two of the Hermite basis splines works well when fitting a mortality curve by age. This is because mortality rates tend to gradually and monotonically increase with age after retirement. Figure 6 shows the deviance residuals by size band, which suggest a more complicated relationship between annuity level and mortality. However, the lower-than-expected mortality for the smallest size bands is of low materiality for actuarial applications: the five size bands covering the decile of lives with the smallest annuities account for less than 0.5% of total annuity payments. We are therefore only interested in the general downward trend in mortality with increasing annuity amount, and especially the sharply lower mortality of those receiving the very largest annuities. As with age, we can again consider mortality by annuity amount to be monotonic without material loss of actuarial relevance.

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

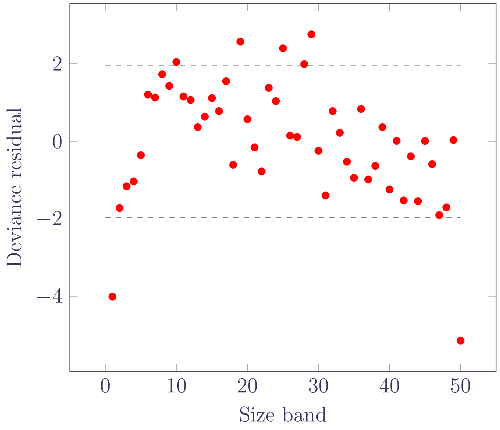

Figure 6. Deviance residuals by size band for the model specified by equation (4), that is, after accounting for variation by age and gender. Each of the 50 size bands contains around 2% of the lives in the UK3 portfolio. The dashed lines show the 95% confidence limits for N(0,1) variates.

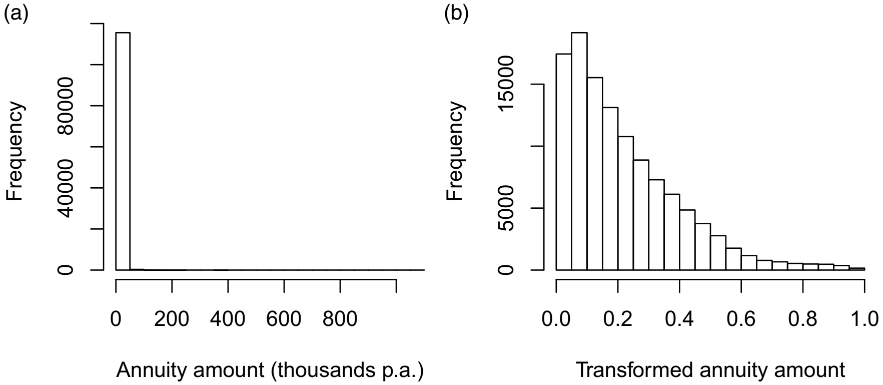

Figure 7. Histograms of annuity amounts. (a) The untransformed amounts, a_i, displaying extreme kurtosis (there are four individuals with pensions in excess of £ 800,000 p.a.). (b) The amounts transformed by $a_i e^{-8.58082}/(1 + a_i e^{-8.58082})$, showing a marked reduction in kurtosis. The value of $\hat{\lambda}_0 = -8.58082$ comes from the model in Table 1, where it was estimated with reference to the mortality characteristics of the UK3 portfolio.

We adopt a two-part approach: (i) a monotonic transform of the annuity amount onto the interval $[0, 1]$ and (ii) modelling mortality by the transformed amount. For the first step, we transform the annuity amount $a \geq 0$ onto the interval $[0, 1]$ so that we can use Hermite splines in the second step. We use a parameterised transform function, $\tau(a, \lambda_0)$, defined in equation (5):

$$\tau(a, \lambda_0) = \frac{ae^{\lambda_0}}{1 + ae^{\lambda_0}} \tag{5}$$

The operation of the transform function in equation (5) is demonstrated in Figure 7, where the parameter λ_0 allows the transform function to adapt simultaneously to the kurtosis of the portfolio and the mortality effect of annuity amount.

The transformation in the right panel of Figure 7 leads to a deliberately unequal distribution of lives and deaths, as shown in Figure 8(a) and (b). However, a useful actuarial consequence is that the unequal distribution focuses on the financially signficant annuities: the interval $[0.95, 1]$ contains the 155 largest annuities with 9.57% of total annuity amounts, whereas the similar sized interval $[0, 0.05]$ has the 18,437 smallest annuities with just 0.96% of the total annuity amounts.

We treat a zero annuity amount as the baseline and use the single h_{01} Hermite spline to model the change (reduction) of mortality with increasing annuity amount on the transformed scale. We

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

Table 1. Parameter Estimates for UK3 Portfolio, Together with Numbers of Contributing Lives and Deaths. The Model is Specified in equation (6). Ages Covered are 55–105 years, using Data from 1 January 2015 to 31 December 2020. AIC = 187,693 and BIC = 187,751

Parameter	Name	Estimate	Std. Err.	*Z*-value	Lives	Deaths
α_0	Intercept	−5.45242	0.03244	−168.08	116,056	25,081
α_{female}	Gender. F	−0.60228	0.05370	−11.22	34,875	6,429
ω_{female}	Gender. F: Oldest	−0.20897	0.03787	−5.52	34,875	6,429
ω_0	Oldest	−0.53007	0.02021	−26.23	116,056	25,081
ω_{amount}	Amount Ultimate	−0.66319	0.11206	−5.92	116,056	25,081
λ_0	Amount Transform Parameter	−8.58082	0.27120	−31.64	116,056	25,081

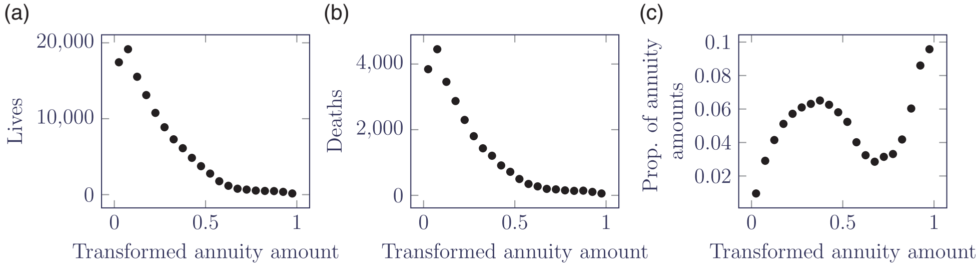

Figure 8. (a) Lives, (b) deaths and (c) proportion of annuity amounts by transformed annuity amount.

thus further extend the model in equation (4) to include a continuous allowance for annuity amount, as in equation (6):

$$\begin{aligned}\log \mu_{x_i} &= \alpha_i h_{00}(u) + \omega_i h_{01}(u)\\ \alpha_i &= \alpha_0 + \alpha_{\text{female}} z_{i,\text{female}} + \omega_{\text{amount}} h_{01}(\tau(a_i, \lambda_0))\\ \omega_i &= \omega_0 + \omega_{\text{female}} z_{i,\text{female}}\end{aligned} \tag{6}$$

The parameter ω_{amount} represents the maximum mortality reduction from a near-infinite annuity income relative to zero income. In equation (6) the function $\tau()$ is the logistic transform in equation (5), although Richards (2021a) considers some alternative transform functions.

The above approach to mortality varying by annuity amount means that the amounts effect can be handled statistically without discrete size bands. This avoids discretisation error and also means that we can extrapolate mortality effects for annuity amounts above the upper limit in the calibrating data set. This latter aspect is useful when calibrating a pricing basis, since an insurer may encounter pension sizes during quotation that are above the largest pension amount in its own experience data. Table 1 shows the parameter estimates for the UK3 portfolio; it is a parsimonious model accounting for three age-varying risk factors with just $p = 6$ parameters for $n = 116,056$ lives. The information criterion from Akaike (1987) is shown ($AIC = -2\ell + 2p$), along with the Bayesian information criterion (BIC) (Schwarz, 1978) ($BIC = -2\ell + p\log n$). For further details of this continuous approach to the amounts effect, see Richards (2021a). Alternatively, see van Berkum *et al.* (2020) for a different approach using thin-plate splines.

Figure 9 shows the deviance residuals by transformed annuity amount before and after fitting the continuous amounts factor. The overall fit of the model has improved, as measured by the Akaike's information criterion (AIC), BIC and χ^2 test statistic by size band. However, we have made an important trade-off in quality of fit: improved fit for the larger annuity amounts comes at the cost of a worsened fit for the smallest annuity amounts. Actuarially we are content with this

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

Table 2. Information Criteria (ICs) from Stepwise Inclusion of Age, Gender and Annuity Amount as Risk Factors

Risk factors	Equation number	Parameter count	AIC	BIC	$\sum_m r_m^2$
Age only	2	2	188,413	188,433	60.7
Age and gender	4	4	187,774	187,813	67.8
Age, gender and annuity amount	6	6	187,693	187,751	45.0

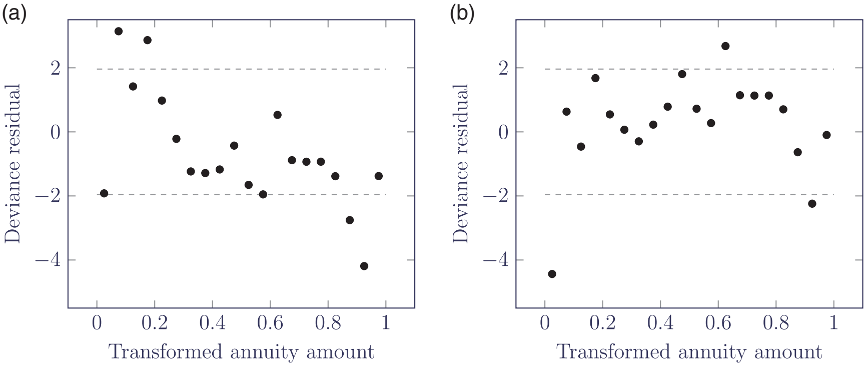

Figure 9. Deviance residuals, $\{r_m : m = 1, \ldots, 20\}$, by transformed annuity amount for two models. (a) For a model including age and gender only with $\sum_m r_m^2 = 67.8$. (b) For a model with age, gender and annuity level with $\sum_m r_m^2 = 45.0$. The deviance residuals are calculated for 20 intervals each of length 0.05, which have very different numbers of lives and deaths, as shown in Figure 8. Note that the grouping here was performed purely for the purpose of residual calculation, and that the underlying mortality model by annuity amount is fully continuous. The dashed lines show the 95% confidence limits for N(0,1) variates.

trade-off due to materiality, as Figure 8(c) shows that the worsened fit accounts for just 0.96% of the annuity amounts and is thus a tiny proportion of overall liabilities.

Table 2 shows the development of model fit from adding age, gender and annuity amount as mortality risk factors. As expected, the information criteria reduce as statistically significant risk factors are added. The sum of squared residuals does not have a similarly monotonic progression: moving from a model with age only to one with age and gender increases $\sum_m r_m^2$ calculated against pension size. This is not an aberration: the same phenomenon is observed when calculating deviance residuals against duration (not shown).

Mortality levels tend to vary monotonically by age. They also can be approximated as such for annuity amount, even if the fit for the very smallest annuities is not ideal. Thus, mortality by age and annuity amount can both be modelled using Hermite splines once each continuous variable is mapped onto the real interval $[0, 1]$. However, modelling mortality variation in calendar time is not monotonic, for which we need another kind of spline.

6. Schoenberg Splines for Mortality Levels over Time

Mortality levels can fluctuate sharply over short periods of time, as shown in Figure 10, where seasonal variation is evident along with the COVID-19 shock in April 2020. Unlike mortality by age and annuity amount, patterns of mortality in time are not monotonic. We therefore require a method with sufficient local flexibility to reflect the rise and fall of mortality levels over time. For this, we use the splines of Schoenberg (1964). Schoenberg's splines are based on polynomials of degree *bdeg* spanning $(bdeg + 2)$ knot points. The knot points are unique values on the real line and in many

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

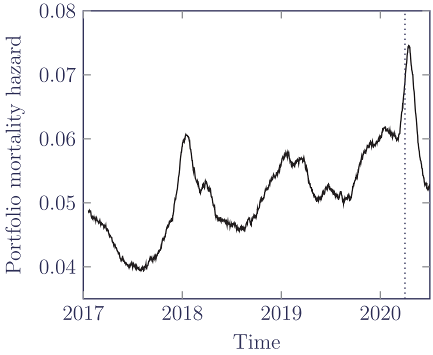

Figure 10. Mortality level in time for a mature US annuity portfolio using a semi-parametric estimator. The vertical dotted line indicates 1 April 2020. Source: Richards (2021b, Figure 3(c)).

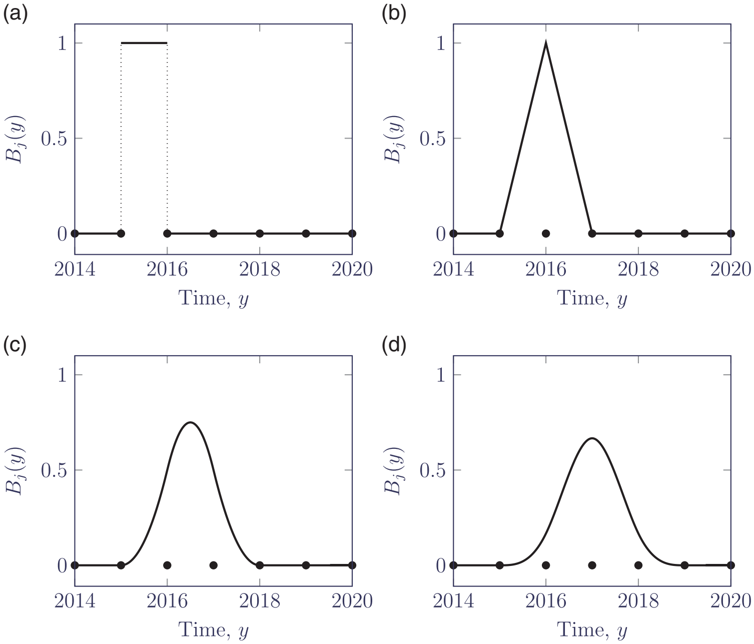

Figure 11. Schoenberg (1964) splines with 1-year spacing between knot points (marked •). (a) $bdeg = 0$, (b) $bdeg = 1$, (c) $bdeg = 2$ and (d) $bdeg = 3$. In this specimen example, each spline is 0 before the knot at 2015 and 0 above the knot $(bdeg + 1)$ knots to the right. The area under each spline is 1.

applications they are uniformly spaced. However, uniform knot spacing is not mandatory (Kaishev *et al.*, 2016) and section 8 will demonstrate an application of unequally spaced knots.

Outside of a spline's starting and end knots the spline takes the value 0, making it a purely local function. Figure 11 shows four Schoenberg splines of varying degrees, and more detail on such splines can be found in de Boor (2001) and Eilers & Marx (2021).

Schoenberg splines are not new in mortality modelling: Eilers *et al.* (2004) applied them to modelling mortality trends in grouped counts. This is done by forming a basis of cubic Schoenberg splines (B-splines) in calendar time, as shown in Figure 12.

We define $B_j(y)$ as the j^{th} basis spline evaluated at time y. Whereas formulae for Hermite h-splines are explicit, as in Figure 4, Schoenberg B-splines are typically defined by means of a recurrence relation; see de Boor (2001, p. 90). We further define $\mu_{x,y}$ as the mortality hazard at exact age x and calendar time y. We can then use a B-spline basis as in Figure 12 to form an age-period model for $\mu_{x,y}$ as per equation (7):

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

Table 3. Estimates of $\kappa_{0,j}$ for $j=1,2,\ldots,8$ for UK3 Portfolio, Using Data for the Age Range 55–105 years from 1 January 2015 to 31 December 2020. $\kappa_{0,0}=0$ by Construction because it is Absorbed into the Baseline Hazard

j	1	2	3	4	5	6	7	8
$\hat{\kappa}_{0,j}$	−4.30805	−3.73912	−3.91987	−3.69781	−4.0887	−3.76673	−3.98538	−4.23435

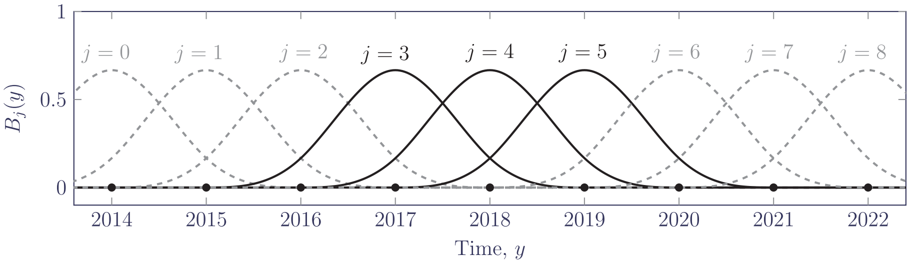

Figure 12. A basis of nine equally-spaced cubic B-splines spanning 1 January 2015 to 31 December 2020, indexed $j=0,1,\ldots,8$. Splines in solid black lie completely within the period, while splines in dashed grey are edge splines that only partly lie in the period being spanned. Knot points are marked •. Note that at any time-point $y\in[2015,2021]$, there are always four non-zero splines that sum to 1. This is not true outside $[2015,2021]$, meaning the method cannot extrapolate outside the calibration interval.

$$\log\mu_{x,y}=\log\mu_x+\sum_{j\geq 1}\kappa_{0,j}B_j(y) \tag{7}$$

The parameter $\kappa_{0,j}$ corresponds to the j^{th} B-spline, and summation is from $j=1$ as the spline $j=0$ is absorbed into the baseline hazard. We use the zero subscript for the series $\{\kappa_{0,j}\}$ to distinguish it from the κ term used in a simpler approach to time-varying mortality in Richards (2020, section 7). To estimate the $\kappa_{0,j}$ along with any other parameters, we maximise the log-likelihood in equation (8):

$$\begin{aligned}\ell&=-\sum_{i=1}^{n}H_{x_i,y_i}(t_i)+\sum_{i=1}^{n}d_i\log\mu_{x_i+t_i,y_i+t_i}\\ H_{x_i,y_i}(t_i)&=\int_0^{t_i}\mu_{x_i+s,y_i+s}ds\end{aligned} \tag{8}$$

where life i enters observation at exact age x_i at exact year y_i, and is observed for exactly t_i years; $H_{x_i,y_i}(t_i)$ is the integrated hazard that now depends on both age at entry and year of entry. Some sample estimates of $\kappa_{0,j}$ are given in Table 3, which are then applied to the B_j in Figure 13 to show the local flexibility. Finally, the various $\hat{\kappa}_{0,j}B_j$ products are summed in Figure 14 to show how the estimated mortality level varies over time.

Note that absorbing $\kappa_{0,0}$ into the baseline as in equation (7) is just one way of specifying the necessary identifiability constraint, in this case $\kappa_{0,0}=0$. That a constraint is required comes from the identity in equation (9):

$$\begin{aligned}\log\mu_{x,y}&=\alpha_0h_{00}(u)+\omega_0h_{01}(u)+\sum_{j\geq 0}(\kappa_{0,j}-c)B_j(y)\\ &=\alpha_0h_{00}(u)+\omega_0h_{01}(u)-c+\sum_{j\geq 0}\kappa_{0,j}B_j(y),\quad\text{as }\sum_{j\geq 0}cB_j(y)=c,\forall c\in R\\ &=(\alpha_0-c)h_{00}(u)+(\omega_0-c)h_{01}(u)+\sum_{j\geq 0}\kappa_{0,j}B_j(y),\quad\text{as }h_{00}(u)+h_{01}(u)=1\end{aligned} \tag{9}$$

Without an identifiability constraint, there is an infinite choice of real-valued c that can be deducted from each $\kappa_{0,j}$ and added to α_0 and ω_0 in equations (1) or (2) and still yield the same

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

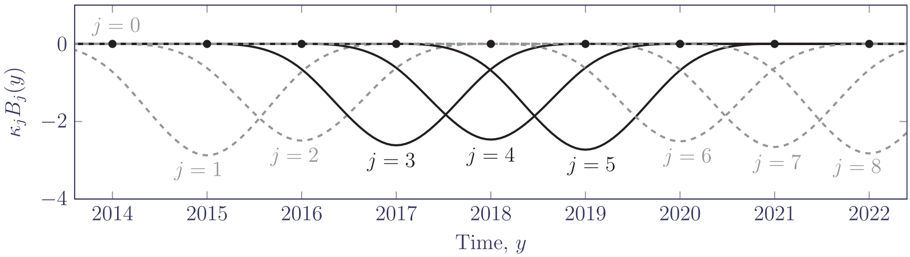

Figure 13. $\hat{\kappa}_{0,j}B_j(y)$ using the nine basis splines in Figure 12 and the estimates in Table 3. Knot points are marked •.

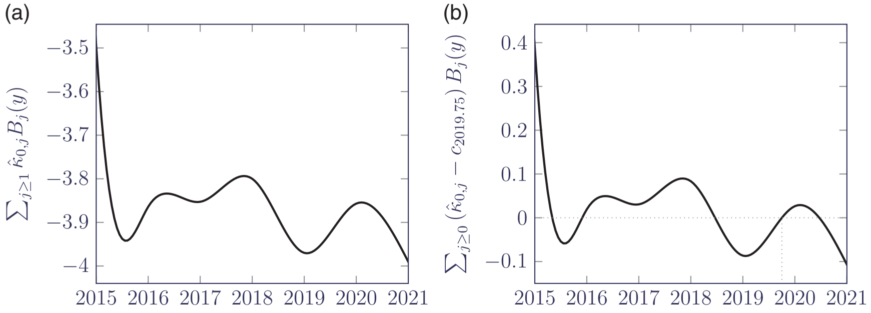

Figure 14. Schoenberg time spline function spanning 1 January 2015 to the end of 2020 using the basis splines in Figure 12 and the $\hat{\kappa}_{0,j}$ in Table 3. Panel (a) shows the unadjusted spline function, while panel (b) shows the function normalised at 0 at 2019.75, the last mid-point between a summer trough and winter peak before the COVID-19 pandemic. Note that summation in panel (b) is from $j=0$ because the coefficient of B_0 is no longer 0.

fit. As a consequence, the vertical scale in Figure 14(a) is somewhat arbitrary and depends in large part on the mortality experience covered by the right-hand part of spline B_0 (see e.g. the spline $j=0$ in Figure 12 relative to the period from 1 January 2015).

We can use equation (9) to normalise the Schoenberg spline function to take the value 0 at a particular point in calendar time. This is useful if we regard that point in time as having a meaning and we want to compare subsequent mortality levels (as is done in Figure 17, for example). For example, 1 October 2019 (2019.75 decimalised) represents the last mid-point between a summer trough and a winter peak before the COVID-19 pandemic, and so might be regarded as the most recent suitable time point unaffected by seasonal swings, pandemic shocks and unreported deaths. If we define $c_{2019.75}=\sum_{j\geq 1}\hat{\kappa}_{0,j}B_j(2019.75)$, we can then deduct this value from each $\hat{\kappa}_{0,j}$ and normalise the Schoenberg spline function without distorting the model fit as long as we also add $c_{2019.75}$ to the estimates $\hat{\alpha}_0$ and $\hat{\omega}_0$ ($c_{2019.75}$ of course needs to be recalculated whenever the knot points in the spline basis change). Figure 14(b) shows the resulting normalised Schoenberg spline.

7. Modelling Seasonal Variation and COVID-19 Shocks

Mortality has long been known to have a major seasonal component; Rau (2007, Chapter 2) gives a comprehensive introduction, covering both historical and modern aspects. Seasonal variation is also pronounced in pensioner mortality, as demonstrated by Richards *et al.* (2020) with a recurring annual cosine term. Such an approach estimates an average seasonal effect within each year, with peak mortality in winter and low mortality in summer. However, this approach cannot account for years with heavier-than-average winter mortality, nor will it account for slight shifts in the timing of the winter peak.

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

Table 4. Information Criteria (ICs) for Various Knot Densities with Equal Spacing. UK3 Portfolio for 1 January 2015 to the End of 2020. ICs and Parameter Counts can be Compared with Table 2 as the Underlying Data are Identical. The Optimal AIC and BIC Values are Marked in Bold

Knots per year	Parameter count	AIC	BIC
1	14	187,594	187,729
2	20	187,412	**187,605**
4	32	187,324	187,634
10	68	**187,244**	187,901

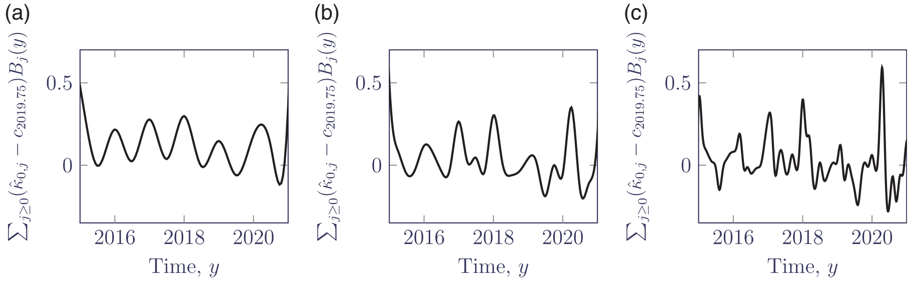

Figure 15. Addition to log(hazard) using equally spaced knots, normalised at 0 for 1 October 2019. (a) Two2 knots per year, (b) four knots per year and (c) ten knots per year.

We can address these issues by increasing the number of knots per year. This increases the number of splines needed to span 2015–2020, with the resulting mortality level shown in Figure 15 for two, four and ten knots per year.

Figure 15(a) shows that using two knots per year picks up the seasonal variation with the winter peaks around January of each year 2015–2019. The exception is 2020, where the expected winter peak in January has merged with the COVID-19 mortality spike of April and May 2020, resulting in an aggregate peak shifted more towards the time of the first COVID-19 shock. Another feature of Figure 15(a) is the deep trough leading up to the end of 2020; this could be either a real feature of the data, say due to the COVID-19 mortality being partly due to brought-forward deaths, or else an artefact from the *B*-spline basis having insufficient flexibility to cope with the short-term intensity of the two COVID-19 shocks in Figure 3. To resolve this, we can increase the knot density further, as in Figure 15(b) and (c), but we face a difficult balancing act – increasing the knot density allows the nature of the COVID-19 spike to come through, but at the expense of weakening the signal at other times.

Table 4 summarises the model fits using various equally spaced knot densities. According to the AIC, the model with ten knot points per year is the best fit, which is rather contradicted by the fluctuations in Figure 15(c). In contrast, the BIC indicates that two knot points per year is the best fit. The conflicting messages are likely due to the large number of parameters not being penalised as heavily in the AIC. In mainstream statistical work, a small-sample correction to the AIC is often used because of this (Macdonald *et al.*, 2018, p. 98). However, in actuarial work, the sample size is typically not a problem due to there being tens of thousands of data points ($n = 116,056$ for the UK3 portfolio here). Rather, the issue is a large number of parameters for a given risk factor – with four or more knot points per year, many of the $\kappa_{0,j}$ parameters prior to the mortality shock are not explaining enough variation to justify their inclusion. Ye (1998, p. 120) notes that "flexibility often leads to substantial overfitting." A possible explanation for the failure of the AIC is given by Owen (1991, pp. 102–103), albeit

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

in the context of a different model — where a statistically significant parameter is being estimated, the estimation costs one degree of freedom; however, if the parameter is statistically insignificant, the estimation costs more than one degree of freedom. Thus, where unnecessary spline parameters are introduced, the standard definition of the AIC may fail to properly penalise this. We will consider this drawback of the AIC again in sections 11 and 12.

8. Increasing the Knot Density Around Shock Times

Figure 15(c) shows that increasing the density of the knot points allows the COVID-19 mortality shock to be clearly identified in terms of height and timing. However, the development from Figure 15(a) to (c) also suggests the introduction of random variation for the non-COVID period. It is undesirable to add knot points where flexibility is not required, so we address this by providing extra knots only where they are needed. One approach is to use half-yearly spaced knots for all years, but to add extra knots around the time of the first COVID-19 spike in April and May 2020, as shown in Figure 16. We justify this from our a priori knowledge of the population COVID-19 mortality in Figure 3.

Figure 17 shows the resulting mortality levels in time, with the expected seasonal fluctuations and the first COVID-19 spike around April 2020. The mixed approach with two knots per year for pre-COVID years and additional knots for the pandemic shock works well in capturing the salient features in time. Figure 17(b) further shows the usefulness of normalising the Schoenberg splines at 0 at a point in time – when converted to the hazard scale the multiplier is 1 at the reference point (2019.75 in this case) and we can see that the peak of the COVID-19 spike is nearly double the reference level of mortality. Following the shock in April 2020, there is also an unusually deep summer trough in 2020, which could be a result of “harvesting” due to brought-forward deaths of the frail. Figure 17(b) also shows that the pre-pandemic seasonal trough-to-peak variation varies between around 15–30%, which is consistent with the results in Richards *et al.* (2020, Table 2) for a variety of international annuity and pension portfolios.

The BIC for the model behind Figure 17 is 187,489, which is considerably lower than any of the BICs in Table 4. This suggests that the basis of variably spaced knots in Figure 16 has provided flexibility only where it is needed. The time signal is relatively strong: dropping the factors gender and annuity amount from the model leaves Figure 17 largely unchanged, which justifies the age plus period nature of equation (7). However, the BIC for the model behind Figure 17 does not take into account the fact that the knots were selected with reference to data and judgement (specifically, a comparison of the plots in Figure 15 and the distribution of COVID-19 deaths in Figure 3). The question of defining and using information criteria is discussed further in section 12.

There are at least two other alternative approaches that would be possible. The first is the optimisation of the number and position of knots by an algorithm targeting a measure of fit; Kaishev *et al.* (2016) propose a knot-addition algorithm and also give a comprehensive overview of related knot optimisation research. The second approach is to deliberately use too many knots and use a tuning parameter to minimise the variability exhibited in Figure 15(c); Eilers & Marx (1996) give details. However, both of these methods have huge computational cost when applied to actuarial data sets, where the observations often number hundreds of thousands. For example, the UK3 data set in this paper has $n = 116,056$ records and the model behind Figure 17 has 19 Schoenberg splines, giving a knot-to-data ratio of around 1:6,000. In contrast, the data set from Kimber *et al.* (2009) reworked by Kaishev *et al.* (2016) involves just 1,151 observations and has 227 knots, giving a knot-to-data ratio of 1:5. The model behind Figure 17 took 1.5 hours to fit using parallel processing over 63 threads (Butenhof, 1997), so adding lots more splines – as per Eilers & Marx (1996) – would result in much longer run times. Similarly, the repeated refitting with new knots – as per Kaishev *et al.* (2016) – would also be time-consuming when applied to actuarial data sets.

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

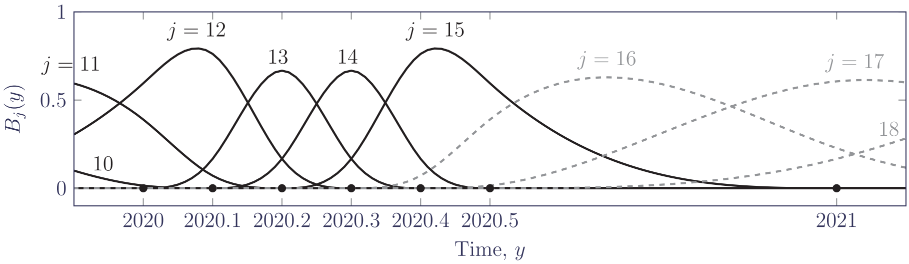

Figure 16. Part of a basis of nineteen variably spaced cubic B-splines spanning 1 January 2015 to the end of 2020, indexed $j = 0, 1, \ldots, 18$ (only splines $j = 10, \ldots, 18$ are shown). Splines in solid black lie completely within the period, while splines in dashed grey are edge splines that only partly lie in the period being spanned. Knot points are marked ●.

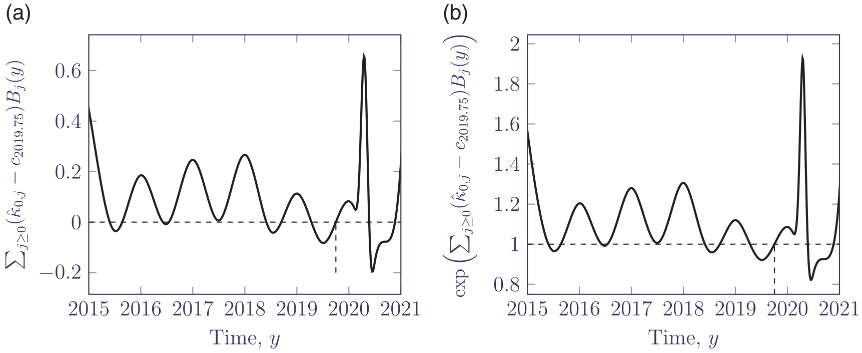

Figure 17. Mortality level modelled using the spline basis depicted in Figure 16. Panel (a) shows the addition to log(mortality), while panel (b) shows the multiplier of the mortality hazard.

9. Estimating Portfolio-Specific Improvements

An advantage of equation (7) is that it can be used to estimate the portfolio-specific mortality improvement (*PSMI*). This can be estimated by selecting either two winter peaks or two summer troughs at time points $y_1 < y_2$ and using equation (10):

$$PSMI = \left[1 - \exp\left(\frac{\sum_{j\geq 1} \kappa_{0,j}(B_j(y_2) - B_j(y_1))}{y_2 - y_1}\right)\right] \times 100\% \tag{10}$$

Our preference is to use summer troughs due to the tendency for sharp peaks in winter and relatively flat troughs in summer (Marx *et al.*, 2010), meaning that winter peaks can be more extreme and variable. From Figure 17(a), the period 2015.5–2019.5 seems to be the most suitable period for estimating the portfolio-specific improvement rate, as the trough in the summer of 2020 may be unusually deep due to brought-forward deaths. With $y_1 = 2015.5$ and $y_2 = 2019.5$, equation (10) gives $PSMI = \left[1 - \exp\left(\frac{-1.00971-(-0.96219)}{2019.5-2015.5}\right)\right] \times 100\% = 1.18\%$ p.a. This is an age-independent mortality improvement rate, as equation (7) is an "age plus period effect" model.

Table 5 shows the improvement rates for various combinations of summer troughs in Figure 17. Of interest are the strong improvement rates ending in summer 2020 (2020.5), indicating the depth of the mortality dip following the first COVID-19 shock in April 2020. This

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

Table 5. Annualised Improvement Rates (*PSMI*) between Summer Troughs in Figure 17

Period	Period end				
Start	2016.5	2017.5	2018.5	2019.5	2020.5
2015.5	−2.8%	−2.1%	0.1%	1.2%	2.4%
2016.5		−1.4%	1.5%	2.5%	3.6%
2017.5			4.3%	4.3%	5.2%
2018.5				4.4%	5.7%
2019.5					7.0%

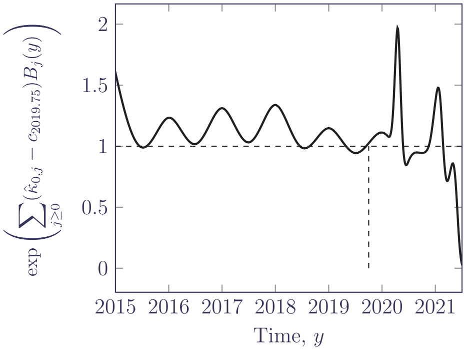

Figure 18. Hazard multiplier using the spline basis depicted in Figure 16 and extended with more knots to cover Q1 2021. The identifiability constraint is that at 1 October 2019 (2019.75 decimalised) the multiplier is normalised at 1.

phenomenon is unrelated to late-reported deaths, which is the subject of section 10. However, it is an open question as to what value such volatile improvement rates have when summer troughs can be almost as variable as winter peaks.

10. OBNR Deaths

Although the UK3 portfolio data include experience to the end of June 2021, we have so far only used the experience to the end of 2020 to minimise the impact of delays in death reporting. Richards (2021b, Figure 7) examined the reporting delays for the same portfolio and found minimal impact on mortality experience a quarter of a year or more before the extract date, so discarding the most recent half-year of experience should be more than enough to eliminate OBNR effects. However, discarding experience data is undesirable if it can be avoided, and Richards (2021b) proposed a parametric model for late-reported deaths as a means of using all available data without distorting the final results.

However, the reduction in reported mortality leading up to the extract date is just another pattern in time, which means that the Schoenberg spline function in equation (7) is also an alternative means of allowing for late-reported deaths. We can therefore extend the spline basis and include all available experience data, including the period most affected by late-reported deaths. The result is presented in Figure 18, which shows not only a multiplicative factor falling to 0 at the extract date at the end of June 2021, but also the second COVID-19 spike of January 2021. Figure 3 suggests that the second COVID-19 spike should be around the same height as the first, and so it is possible that the lower spike in January 2021 in Figure 18 is due to some OBNR deaths. We could in theory use the forecasting method of Richards (2021b, section 7) to adjust for this, but a parametric OBNR model will not easily fit due to the high correlation with the last spline parameter. However, it is more likely that the first peak in Figure 3 is due to COVID-19 deaths not being recorded as such.

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

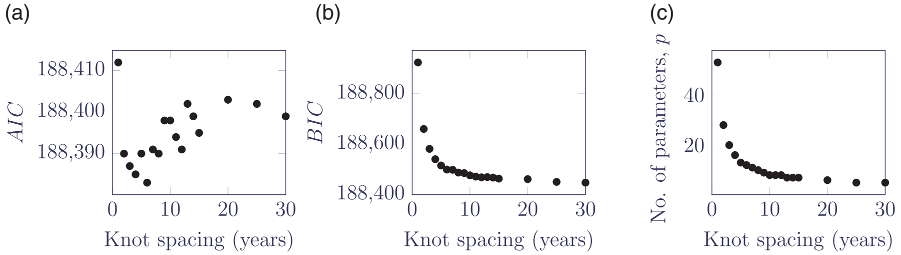

Figure 19. (a) AIC, (b) BIC and (c) number of spline parameters for UK3 mortality rates using model in equation (11) and various equidistant knot spacings. Ages covered are 55–105 years from 1 January 2015 to 31 December 2020.

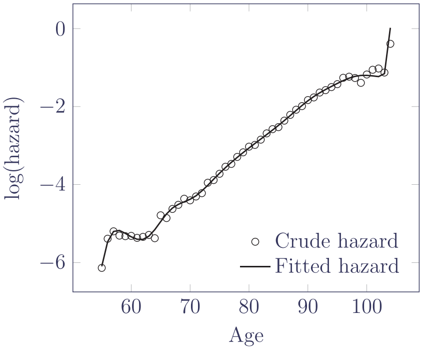

Figure 20. Crude mortality hazard at each age, x, and fitted curve using equation (11) and a 4-year knot spacing, showing over-fitting. Source: own calculations using UK3 data.

11. Schoenberg Splines for Mortality by Age

The demonstrable effectiveness of Schoenberg splines when applied to time-varying mortality raises the question whether we should not also use them for mortality by age. This is not a new idea – McCutcheon (1979) used cubic splines by age when graduating mortality tables, for example. However, the approach of McCutcheon (1979) was designed for stratified grouped data, and not the inclusion of additional covariates per life.

For application to a survival model for individual lives, we therefore use equation (11):

$$\log \mu_x = \alpha_0 + \sum_{k\geq 1} \alpha_{0,k} B_k(x) \tag{11}$$

where α_0 is the baseline, $B_k(x)$ is the k^{th} B-spline evaluated at exact age x and $\alpha_{0,k}$ is the corresponding spline parameter. As before, summation is from $k = 1$ because the spline $k = 0$ is absorbed into the baseline.

Figure 19 shows the AIC and BIC for various knot spacings for $B_k(x)$. As with Table 4, the two information criteria give conflicting guidance on which model to pick: the AIC indicates that the best knot spacing in age is 4 years, while the BIC suggests that the optimum knot spacing is 25 years or more.

One reason to reject the AIC-recommended 4-year knot spacing is shown in Figure 20. The model is over-fitted at both the youngest and oldest ages: we have no reason to believe that mortality rates decline at the oldest ages, and the single death at the age of 55 years is not grounds enough for the fitted shape. This is then an advantage of the Hermite spline approach of equation (1) — it has fewer parameters and less flexibility, thus only permitting slow and stable changes with age. Alternatively, a basis of widely spaced Schoenberg splines will behave similarly. However, we note that the Hermite spline approach of equation (6) is better suited to the modelling of post-retirement mortality due to automatic convergence of differentials with increasing age (Richards, 2020, section 3).

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

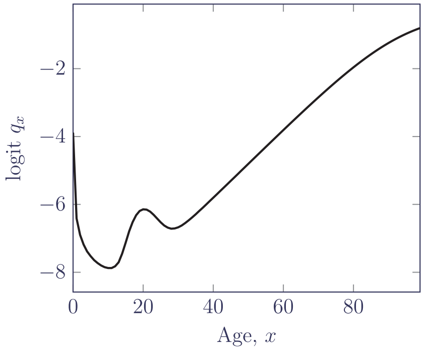

Figure 21. Graduated mortality rates for Australian males, showing accident hump around the age of 20 years. Source: Heligman & Pollard (1980, Figure 12).

In Figure 20, we see that excessive flexibility from closely spaced Schoenberg splines in age is undesirable – the fit might be superficially better quantitatively in terms of the AIC, but is poorer qualitatively when one takes into account data quality and reasonable prior expectations. However, there will be occasions when the greater flexibility of closely spaced Schoenberg splines is useful, such as where younger ages are included. Figure 21 shows an example of a non-monotonic pattern with age, for which Hermite splines will be unsuitable and for which greater flexibility is required.

12. Information Criteria and Actuarial Judgement

In this paper, we have seen two instances where the AIC led to a qualitatively poorer choice of knot spacing for Schoenberg splines for mortality: first for a time-varying function in Figure 15(a) and again for an age-varying function in Figure 20. In both cases, the AIC led to an excessive number of parameters and over-fitting. Using the small-sample correction for the AIC (Hurvich & Tsai, 1989) would not have led to a different outcome due to the large number of observations (Macdonald *et al.*, 2018, Table 6.1).

The definitions of the AIC and BIC used in this paper are the usual ones given in section 4 and used in Figures 19(a) and (b). However, it is worth noting that more rigorous definitions are given in equation (12), where ℓ is the log-likelihood from equation (3), n is the number of lives and df is the number of degrees of freedom used. In equation (12), it is common to use $df = p$, where p is the number of parameters; for a simple linear regression model, the two are synonymous. However, Owen (1991, p. 103) noted that it is possible for a non-linear model to have $df > p$, while Ye (1998, p. 122) notes that even a linear smoother can have fewer degrees of freedom than there are parameters ($df < p$). In the field of mortality modelling, Macdonald *et al.* (2018, section 11.6) give an example of a penalised Generalized Linear Model where the effective degrees of freedom are $df = 9.0$ with $p = 20$ regression coefficients. Assuming that $df = p$ is a simplifying assumption only, and actuaries should not automatically pick a model solely because of its low information criterion.

$$\begin{aligned} AIC &= -2\ell + 2df \\ BIC &= -2\ell + df \log n \end{aligned} \tag{12}$$

This problem of the AIC leading to over-fitting may be restricted to relatively large numbers of parameters for a given risk factor. However, relying only on the BIC as an alternative is not a complete solution either for actuarial work. For example, there is a distinction between risk factors that are statistically significant and those that are financially significant; Richards (2020, section 9) gives several contrasting examples. In Figure 6, the large negative residuals for the lowest and highest size bands are of equal significance statistically, yet actuaries will tolerate a poorer fit for smaller amounts if it means better explaining the mortality of those with the largest annuity

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

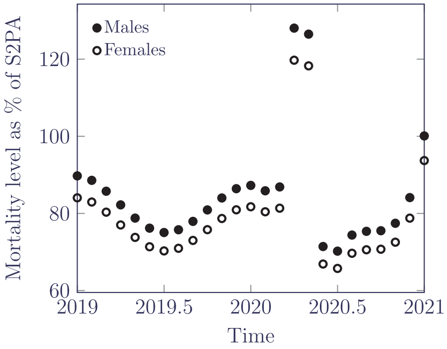

Figure 22. Percentages of S2PA implied by mortality levels over 2019–2020. Liabilities for UK3 at 1 January 2021 equated as per equation (13) using a net discount rate of 0% p.a. ($v^t = 1, \forall t$).

amounts. While the BIC might be the better information criterion, it cannot be the sole arbiter of model selection for actuarial purposes.

Leaving aside the complexities of model selection, there is also the practical question of how to turn a complex, multi-factor model into a mortality basis. In the bulk annuity and longevity swap markets, specialist valuation software is required to handle the complexities of UK pensioner benefits. However, such valuation systems are seldom capable of handling multi-factor models, and usually a table of mortality rates by age and gender is all that can be accommodated. A useful conversion tool is the equivalent reserve method of Willets (1999), which expresses one mortality basis in terms of another via the medium of the liabilities in question. For the sake of example, assume that we want to express current mortality levels in terms of the equivalent percentages of some third-party table. We are interested in the mortality rates appplying at a point in time, so we ignore future mortality improvements, as these are usually a separate basis item. The approach of Willets (1999) is to solve equation (13) for males and females separately:

$$\sum_i a_i \int_0^\infty v^t\,{}_tp_x^{\text{model}}dt = \sum_i a_i \int_0^\infty v^t\,{}_tp_x^{\text{table}}dt \tag{13}$$

where a_i is the annuity amount for life i, v^t is the net discount function to apply to a payment in t years, ${}_tp_x^{\text{model}}$ is the survival probability according to the model and ${}_tp_x^{\text{table}}$ is the survival probability according to the published table. We do not need to worry about the distorting effect of ignoring mortality improvements, as long as they are ignored on both sides of equation (13). In general, ${}_tp_x = \exp\left(-\int_0^t \mu_{x+s}ds\right)$, and we use the definition of the mortality hazard in equation (14):

$$\mu_x^{\text{table}} = f\mu_x^{\text{S2PA}} \tag{14}$$

where μ_x^{S2PA} is the mortality hazard according to the table S2PA (CMI Ltd, 2014), and f is the percentage of that table to be solved for in equation (13). Figure 22 shows the results using the modelled mortality levels applying at various points in time over 2019–2020, assuming – somewhat unrealistically – that those mortality levels continue indefinitely, that is, for the calculation using mortality levels at outset time y the integrated hazard is $H_x(t) = \displaystyle\int_0^\infty \mu_{x+s,y}$.

Figure 22 shows the extent of the challenge for actuaries in setting the best-estimate mortality basis for pricing bulk annuities or longevity swaps. Under normal circumstances, one might pick the most recent mid-point between a summer trough and a winter peak; 75–76% for males and 70–71% for females in October 2020, say. As it happens, mortality at this point in time was largely unaffected by COVID-19 (see Figure 3), but the equivalent percentages for April and May 2020 show the discontinuities possible due to pandemic mortality. The equivalent percentages for January 2020 show an emerging second discontinuity due to the second COVID-19 shock as shown in Figures 3 and 18.

Of course, no actuary would use equivalent percentages like those for April and May 2020 in Figure 22. However, the importance of allowing for the mortality shock can be seen from

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

imagining what the results would be like without it: without the flexibility of the Schoenberg splines in time, the shock points would be moved down and all the other percentages in Figure 22 would be shifted up; any resulting basis would thus be imprudent for pricing a bulk annuity or longevity swap. The value of the methodology lies in accommodating the mortality spikes so that they cannot drive bias at other points in time.

13. Conclusions

For continuous variables where mortality varies either monotonically or with simple shape variation, a basis of Hermite splines usually provides all the flexibility that is needed. Examples include mortality by age and annuity amount. For actuarial purposes, quality of fit for the smallest annuity amounts is less important than the fit for the largest amounts.

In contrast, continuous variables where mortality fluctuates a lot, or where fluctuations are sharp and extreme, are better handled by a basis of local splines. Using cubic Schoenberg splines (*B*-splines), we can model mortality levels across time for annuity portfolios and pension funds. With two knot points per year, we can identify seasonal variation in mortality, and we can estimate portfolio-specific mortality improvements from the change in mortality levels between summer troughs. Where the number of observations is large, we find that the BIC is materially better for selecting the number of knot points than the AIC.

To handle mortality shocks like COVID-19, we add additional knot points from our a priori knowledge of the timing of mortality spikes in the population. This mixed approach of regular half-yearly knots and hand-placed additional knots allows the salient mortality features to be identified, that is, both seasonal variation and mortality shocks. A benchmark time point can be selected to use in setting a basis, safe in the knowledge that the model's other parameters are not unduly biased due to the presence of shocks because their effects are explicitly modelled.

The flexibility of the local Schoenberg splines further allows the modelling of the impact of late-reported deaths. This removes the need to discard the most recent experience data and thus permits the use of all available data for analysis and basis setting.

Acknowledgements. The author thanks Anil Gandhi and Caroline Roberts for their support in providing data. The author also thanks Gavin Ritchie, Torsten Kleinow, Stefan Ramonat, Prof. Angus Macdonald and two anonymous reviewers for helpful comments on earlier drafts of this paper. Any errors or omissions remain the responsibility of the author. Calculations were performed using the Longevitas survival modelling system. Graphs were produced in tikz, pgfplots and R.

Disclaimer: The views expressed in this publication are those of invited contributors and not necessarily those of the Institute and Faculty of Actuaries. The Institute and Faculty of Actuaries do not endorse any of the views stated, nor any claims or representations made in this publication and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this publication. The information and expressions of opinion contained in this publication are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this publication be reproduced without the written permission of the Institute and Faculty of Actuaries.

References

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

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

CMI Ltd. (2014). *Graduations of the CMI SAPS 2004–2011 Mortality Experience based on Data Collected by 30 June 2012 | Final "S2" Series of Mortality Tables*. CMI Ltd.

Collett, D. (2003). *Modelling Survival Data in Medical Research* (2nd ed.). Boca Raton, Chapman and Hall. ISBN 1-58488-325-1.

Craufurd Dunlop, J.C. & Watt, A. (1915). *Fifty-ninth annual report of the Registrar General for Scotland*, **volume 59**. Glasgow, H.M. Stationery Office.

Craufurd Dunlop, J.C. & Watt, A. (1916a). *Sixtieth Annual Report of the Registrar General for Scotland*, **volume 60**. Glasgow, H.M. Stationery Office.

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

Craufurd Dunlop, J.C. & Watt, A. (1916b). *Sixty-First Annual Report of the Registrar General for Scotland*, **volume 61**. Glasgow, H.M. Stationery Office.

Craufurd Dunlop, J.C. & Watt, A. (1918). *Sixty-Second Annual Report of the Registrar General for Scotland*, **volume 62**. Edinburgh, H.M. Stationery Office.

Craufurd Dunlop, J.C. & Watt, A. (1919). *Sixty-Third Annual Report of the Registrar General for Scotland*, **volume 63**. Edinburgh, H.M. Stationery Office.

Craufurd Dunlop, J.C. & Watt, A. (1920a). *Sixty-Fourth Annual Report of the Registrar General for Scotland*, **volume 64**. Edinburgh, H.M. Stationery Office.

Craufurd Dunlop, J.C. & Watt, A. (1920b). *Sixty-Fifth Annual Report of the Registrar General for Scotland,* **volume 65**. Edinburgh, H.M. Stationery Office.

Currie, I.D., Durban, M. & Eilers, P.H.C. (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society, Series B*, **68**, 259–280.

de Boor, C. (2001). *A Practical Guide To Splines, volume 27 of Applied Mathematical Sciences*. New York, Springer-Verlag. ISBN 978-0-387-95366-3.

Eilers, P.H.C. & Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**(2), 89–121. ISSN 08834237, available at http://www.jstor.org/stable/2246049.

Eilers, P.H.C. & Marx, B.D. (2021). *Practical Smoothing: The Joys of P-Splines*. Cambridge University Press. ISBN 978-1-108-48295-0.

Eilers, P.H.C., Currie, I.D. & Durban, M. (2004). Fast and compact smoothing on large multidimensional grids. *Computational Statistics and Data Analysis*, **50**, 61–76.

Eilers, P.H.C., Gampe, J., Marx, B.D. & Rau, R. (2008). Modulation models for seasonal time series and incidence tables. *Statistics in Medicine*, **27**(17), 3430–3441. doi: 10.1002/sim.3188.

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

Heligman, L. & Pollard, J.H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, **107**(Part I, 434), 49–80.

Hurvich, C.M. & Tsai, C.L. (1989). Regression and time-series model selection in small samples. *Biometrika*, **76**(2), 297–307.

Kahan, W. (1965). Further remarks on reducing truncation errors. *Communications of the ACM*, **8**(I), 40.

Kaishev, V.K., Dimitrova, D.S., Haberman, S. & Verrall, R.J. (2016). Geometrically designed, variable knot regression splines. *Computational Statistics*, **31**(3), 1079–1105. doi: 10.1007/s00180-015-0621-7.

Kimber, S.A.J., Kreyssig, A., Zhang, Y.Z., Jeschke, H.O., Valenti, R., Yokaichiya, F., Colombier, E., Yan, J., Hansen, T.C., Chatterji, T., McQueeney, R.J., Canfield, P.C., Goldman, I.A. & Argyriou, D.N. (2009). Similarities between structural distortions under pressure and chemical doping in superconducting BaFe2As2. *Nature Materials*, **8**, 471–475.

Kreyszig, E. (1999). *Advanced Engineering Mathematics* (8th ed.). John Wiley and Sons. ISBN0-471-33328-X.

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

Marx, B.D., Eilers, P.H.C., Gampe, J. & Rau, R. (2010). Bilinear modulation models for seasonal tables of counts. *Statistics and Computing*, **20**(2), 191–202. doi: 10.1007/s11222-009-9144-9.

McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models*, (2nd ed.), **volume 37**. Monographs on Statistics and Applied Probability. London, Chapman and Hall. ISBN 0-412-31760-5.

McCutcheon, J.J. (1979). Some remarks on splines. *Transactions of the Faculty of Actuaries*, **37**, 421–438. doi: 10.1017/S0071368600008727.

ONS (2021). Daily deaths with covid-19 on the death certificate by date of death. Technical report, available at https://coronavirus.data.gov.uk/details/deaths (accessed 15 June 2021).

Owen, A. (1991). Discussion: multivaritate adaptive regression splines. *The Annals of Statistics*, **19**(1), 102–112. doi: 10.1214/aos/1176347969.

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

Public Health England (2020). *The Health Protection (Coronavirus, Restrictions) (England) Regulations*, available at https://www.legislation.gov.uk/uksi/2020/350/contents/made

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

Rau, R. (2007). *Seasonality in Human Mortality: A Demographic Approach*. Demographic Research Monographs, 03 **XVI**. Springer Verlag. ISBN 3-540-44900-0.

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

Richards, S.J. (2021a). Modelling mortality by continuous benefit amount. *Scandinavian Actuarial Journal*, page to appear, available at https://www.longevitas.co.uk/site/library/Amounts.pdf.

Richards, S.J. (2021b). *Mortality Shocks and Reporting Delays in Portfolio Data*. Longevitas Ltd, available at https://www.longevitas.co.uk/site/library/Monitoring_mortality.pdf.

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

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

Schoenberg, I.J. (1964). Spline functions and the problem of graduation. *Proceedings of the American Mathematical Society*, **52**, 947–950.

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

Spreeuwenberg, P., Kroneman, M. & Paget, J. (2018). Reassessing the global mortality burden of the 1918 inuenza pandemic. *American Journal of Epidemiology*, **187**(12), 2561–2567. ISSN 0002-9262. doi: 10.1093/aje/kwy191.

The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team (2020). *The Epidemiological Characteristics of An Outbreak of 2019 Novel Coronavirus Diseases (COVID-19)|China*. China CDC Weekly, 2:113, 2020. ISSN 2096-7071, available at http://weekly.chinacdc.cn//article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51.

van Berkum, F., Antonio, K. & Vellekoop, M. (2020). Quantifying longevity gaps using micro-level lifetime data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **184**(2), 548–570. doi: 10.1111/rssa.12631.

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

Ye, J. (1998). On measuring and correcting the effects of data mining and model selection. *Journal of the American Statistical Association*, **93**(441), 120–131. doi: 10.1080/01621459.1998.10474094.

Appendices

A. Implementation

A general approach to fitting left-truncated survival models is presented in Richards (2020, Appendix B). The starting point is the log-likelihood, ℓ, defined in equation (8). The contribution of a single life i to the log-likelihood is given in equation (15):

$$\ell_i = -\left[\int_0^{t_i} \mu_{x_i+s,y_i+s} ds\right] + d_i \log \mu_{x_i+t_i,y_i+t_i} \tag{15}$$

To maximise the log-likelihood function, we require the gradient with respect to each parameter, and to estimate standard errors we require the second pure and cross-derivatives for all parameters. General formulae are given in Richards *et al.* (2020, Appendix A).

The nature of the mortality hazards in equations (1), (2), (4), (7) and (11) means that closed-form expressions for the integrated hazard in equation (15) do not exist. For the results in this paper, we have used the Romberg method of numerical integration, but Clenshaw–Curtis integration is also an option. Richards *et al.* (2020, Appendix A) discuss some practical tests for ensuring that numerical integration has taken place sufficiently accurately. Due to the computationally intensive nature of the calculations, we process work in parallel across 63 threads (Butenhof, 1997). For the summation of lots of potentially small contributions like equation (15), we use the floating point error correction algorithm of Kahan (1965).

Where a material OBNR effect exists, as in Figure 18, the value of $\hat{\kappa}_j$ can be large and negative for the rightmost spline; in the model behind Figure 18, for example, $\hat{\kappa}_j \in (-5, 4), \forall j \in \{1, \ldots, 22\}$, but $\hat{\kappa}_{23} = -66.5$. This can lead to a large number of iterations where the algorithm for optimising the log-likelihood places a limit on the size of the change in parameter estimates; examples include the parscale control parameter in R Core Team (2017, optim function) and the STPMX variable in Press *et al.* (2005, dfpmin function). To reduce the number of iterations, the limit on the step size between successive parameters can be increased, or an adaptive step size used when the number of iterations proves large. In our implementation, we double the maximum step size every 50 iterations – this has little impact on most models with parameters in the range $(-20, +20)$ but adapts to cope with models with extreme parameter values caused by OBNR.

B. Model Parameters

There are two groups of model parameters to be set for the mortality models in this paper: (i) configuration settings, whose values are decided in advance by the analyst and (ii) parameters whose values are estimated from the data.

B.1. Model Settings Decided by the Analyst

Table B.1 lists the configuration settings that need to be decided in advance by the analyst, that is, they are not estimated from the data. Where appropriate, the values used in the main body of the paper are given.

Other analyst decisions include the identifiability constraint on the $\{\kappa_{0,j}\}$ parameters (we use $\kappa_{0,0} = 0$) and the knot spacing.

https://doi.org/10.1017/S1357321721000180 Published online by Cambridge University Press

Table B.1. Configuration Parameters

Parameter	Value	Description and role
x_0	50	Age below which $\log \mu_x$ is unmodelled. See equation (1).
x_1	110	Age above which $\log \mu_x$ is unmodelled. See equation (1).
$\tau(a, \lambda_0)$	$\frac{ae^{\lambda_0}}{1+ae^{\lambda_0}}$	Transform function mapping annuity amount $a \geq 0$ onto $[0, 1]$. See equation (6) and Richards (2021a) for other transforms. The value of λ_0 is estimated from the data.
bdeg	3	Order of Schoenberg splines used in equation (7).

B.2 Parameters Estimated from the Data

Table B.2 sets out the parameters whose values are estimated from the data by maximising the log-likelihood in equation (3) or (8).

Table B.2. Overview of Parameters

Parameter	Name	Description and role of parameter
α_0	Intercept	$\log \mu_{x_0}$.
α_{female}	Gender. F	Addition to α_0 for females.
m_0	Age Gradient Youngest	Initial rate of change of log(mortality) to the right of x_0.
m_1	Age Gradient Oldest	Rate of change of log(mortality) approaching age x_1 from below.
ω_0	Oldest	$\log \mu_{x_1}$, the logarithm of the limiting hazard above x_1.
ω_{female}	Gender. F: Oldest	Addition to ω for females.
ω_{amount}	Amount Ultimate	Maximum mortality reduction on a logarithmic scale from having a near-infinite annuity amount; see equation (6).
λ_0	Amount Transform Parameter	Parameter used in mapping annuity amount from $[0, \infty)$ to $[0, 1]$; see equation (6).
$\kappa_{0,j}$	Time Spline. *j*	Coefficient of time spline *j*. Splines are numbered from 0, but the 0th spline is absorbed into the baseline hazard.

Cite this article: Richards SJ. (2022). Allowing for shocks in portfolio mortality models. *British Actuarial Journal*. https://doi.org/10.1017/S1357321721000180