APPLYING SURVIVAL MODELS TO PENSIONER MORTALITY DATA

By S. J. Richards

[Updated on 10 March 2008]

ABSTRACT

Data from insurance portfolios and pension schemes lend themselves particularly well to the application of survival models. In addition to the traditional actuarial risk-rating factors of age, gender and policy size, we find that using geodemographic profiles based on postcode provides a major boost in explaining risk variation. Geodemographic models can be better than models based on pension size in explaining socio-economic variation, but a model using both is usually better still. Models acknowledging heterogeneity tend to fit better than models which do not. Finally, bootstrapping techniques can be used to test the financial applicability of a model, while weighting the model fit can be used to address concentration risk.

KEYWORDS

Postcode; zip code; geodemographic models; survival models; mortality laws; frailty; ageing; reliability theory; late-life mortality deceleration; concentration risk.

CONTACT ADDRESS

Stephen Richards, 4 Caledonian Place, Edinburgh, EH11 2AS, United Kingdom; Telephone: +44 (0)131 315 4470; Email: stephen@richardsconsulting.co.uk; Web: www.richardsconsulting.co.uk

1. Introduction

1.1 Actuaries have been long used to weighting their calculations by policy size to take account of socio-economic differentials amongst policyholders. So-called "amounts-based" measures routinely produce lower mortality rates than their lives-based equivalents due to the tendency for wealthier policyholders to live longer. This paper describes three different approaches to allowing for socio-economic differentials within the structure of a statistical model. These three approaches can even be combined to provide a robust model of socio-economic differentials for financial applications.

1.2 Cox (1972) introduced the application of *survival models* to life-table problems. The *proportional hazards model* introduced the idea of a force of mortality which is a constant proportion of the *baseline hazard* for some reference population. Such models use data at the level of the individual, rather than modelling counts of events for grouped data. The original Cox model assumed a constant proportion, so changes in mortality differentials with age could not be modelled, nor did it estimate the baseline hazard. The original Cox model was primarily useful for testing hypotheses (e.g. 'does Group A have different risk from Group B?'), but it has been extended for actuarial use in a number of ways since.

1.3 At its core, the simplest model of constant hazard assumes that the lifetime of an individual is distributed exponentially. This does not allow for age-related increases. Using a transformation of the exponential distribution of future lifetime produces more usable models for actuaries. For example, Aitken *et al* (1989) show that the extreme-value distribution arises from taking the natural logarithm of a power transformation of an exponential variable. This equates to a Gompertz mortality hazard (see Appendix 4), and can be fitted simply using a package like R.

1.4 However, survival models based on transformations of the exponential distribution have two major drawbacks. The first is that the power transformation has to be applied equally to all lives, which is the same thing as saying all lives have to have the same age-related change in mortality. This is not the case for males and females, nor is it typically true for different socio-economic groups or most other risk factors. Indeed, Strehler and Mildvan (1960) showed that the stronger the initial mortality differential between two populations, the faster those differentials would narrow with age.

© Stephen Richards. All rights reserved.

This so-called *compensation law of mortality* means that a model with multiple rating factors needs different rates of ageing.

1.5 The second drawback of the transformation approach, at least as implemented in standard software like R, is that it needs to assume observation from birth in order to fit into a linear model-fitting algorithm. Data used by actuaries, whether in life insurance or pension work, is almost always *left-truncated*, i.e. the lives in question only become known to the actuary many years after birth when the individual first enters observation under a contract.

1.6 Here we will use an extension of these models which allows mortality differentials to change with age, and also permits the handling of left-truncated data. We will work with the log-likelihood function directly, thus liberating ourselves from the constraint of needing a linear model-fitting algorithm and opening up a wider choice of patterns for mortality.

2. Data, validation and preparation

"These budget numbers are [...] for the fiscal year that ended February the thirtieth."

US President George W. Bush, October 11, 2006

2.1 The data used in this paper comprise 777,111 distinct lives in a combined portfolio of life-office pensioners and members of defined-benefit pension schemes. The life-office pensioners are mainly purchasers of pension annuities arising from money-purchase pension arrangements and are in the slight majority. There were 118,494 deaths observed in the combined portfolio, and several million life-years of exposure overall from the late 1990s to mid-2007.

2.2 Data preparation is an essential part of any model-building process and no model is valid unless the data are reliable. There are four stages of data preparation described here: (i) data extraction, (ii) data validation, (iii) deduplication, and (iv) profiling.

2.3 We prefer direct extraction from the administration or payments system. This is easy for companies to arrange, as it is usually a straightforward database query. Data which has been pre-processed for an actuarial valuation system is rarely suitable, as this tends to hide valuable data features for statistical modelling. We take direct extracts of data items without any kind of calculation being performed, such as exposed-to-risk or age rounding. In addition to being easier for the client to extract a date of birth instead of calculating an age, this also avoids mistakes in interpretation, such as 'age next' v. 'age last' or 'age nearest'. Perhaps best of all, this approach also does not tie us into a particular methodology, as it would if we asked for particular calculations for age or exposed-to-risk. Furthermore, the direct data contains more information: in the case of date of birth, this not only gives age, but also cohort and even season of birth.

2.4 Experience data from life insurers and pension funds is essentially a longitudinal study: at the start the exact date of policy commencement is known, as are the date of birth, the gender, the pension size and other features. The insurer or pension fund then makes regular payments throughout time, keeping address and other details reasonably up-to-date. Finally, the insurer receives timely notification of death, either from the bank when the account is frozen for probate, or else when reported by a surviving relative or partner. Such detailed, well-maintained data is therefore ideal for survival models. We find that data quality is typically best where some kind of regular payment is made, either paid as a pension or collected from policyholders in the form of a premium. We generally find that data quality is poor in other circumstances, e.g. pensions in deferment. Spouse mortality data is usually only reliable once a spouse's pension has commenced payment after the death of the main life.

2.5 The above comments about regular maintenance aside, it is nevertheless essential to check the validity of the basic data. Basic sense checks are always advisable, since even the best administration system has the occasional piece of nonsensical data. Dates need to be first checked for basic validity: no 30$^{\text{th}}$ February or 31$^{\text{st}}$ June, for example. The relationships of dates also need to checked: for example, date of birth $\leq$ commencement date $\leq$ date of death. Where a commencement date is not available, the date of first pension payment (or first premium collection) makes a good substitute. Other obvious checks apply: that there is a gender code, for example, and that the pension paid is positive.

2.6 Further data inspection is also essential, since poor data can often pass basic validity checks. A common example is 1st January 1901 — it is a perfectly valid date of birth to have, but if several hundred lives have it, it is more likely a false date of birth entered during a migration to a new computer system. The examination of the most frequently occurring data items is also instructive. For example, an excessively common date of death might indicate mass processing due to a certification exercise, suggesting that the date is that of processing, not death itself.

2.7 In this paper we will use mortality-experience data from pensions in payment from United Kingdom insurers and defined-benefit pension schemes. A feature of such data is the presence of duplicate records in the data set. By 'duplicate' we mean that the same person has more than one policy or benefit record. For life-company annuities it is common for people to have two or more annuities, and this is particularly common for wealthier policyholders. Multiple annuities can arise for customer-driven reasons, such as phased retirement, or for structural ones, such as different parts of the pension having different escalation rates and having to be managed separately. Pension schemes have similar duplication issues, although usually not to quite the same extent as annuity portfolios. A pensioner can have multiple records in the same scheme due to multiple periods of service, for example. Another possibility is having both a main pension (because the member worked for the employer) and a spouse's pension (because their deceased spouse worked for the same employer).

2.8 Duplicates are a major problem for statistical models due to the requirement for assumed independence between events being modelled. If we do not remove duplicate records then at the very least the standard errors on the parameter estimates will be incorrect. We thus need to turn the set of benefit records into a set of independent lives. One might hope that this could be done by means of a unique client identifier or National Insurance number. In practice, however, National Insurance numbers are not always routinely entered onto the administration system, or else dummy values are entered. Furthermore, servicing staff often find it easier to simply re-key a person onto a system than try to link a second benefit to the original client record. The major drawback of this is that the same person will then have two or more different client records, making the system client identifier an imperfect means of deduplication. To deduplicate the benefit records therefore requires a unique key which can be created from the ordinary data. A good first start is a combination of date of birth, gender, surname and postcode.

Table 1. Examples of matching surname fields using double metaphone (Phillips, 1990)

Record	Surname	Initial	Comment
1	Richie	G	
2	Ritchie	G	Metaphone match on surname in record 1.
3	Mohammed	A	
4	Muhammed	A	Metaphone match on surname in record 3.
5	Mohammad	A	Metaphone match on surname in record 3.
6	Mahamad	A	Metaphone match on surname in record 3.
7	Muammad	A	Metaphone match on surname in record 3.
8	O'HARE	M	
9	OHARE	M	Metaphone match on surname in record 8.
10	DE-SANTIS	J	
11	Desantis	J	Metaphone match on surname in record 10.
12	D'Santis	J	Metaphone match on surname in record 10.

2.9 To complicate things, however, re-keyed clients will often have subtle changes in name format or even mis-spellings. One way to handle variant spellings of surnames is to use the metaphone encoding system (Phillips, 1990). We have used an extension called double metaphone encoding, which can also handle non-Anglo-Saxon surnames. A further requirement is to intelligently strip out punctuation such as apostrophes and hyphens. Examples of this sort of approach are shown in Table 1.

Table 2. Examples of matching forename fields. The following five records are all identified as the same person according to the forename matching algorithm.

Surname	**Forename(s)**	**Comment**
Richards	Stephen	First initial only used.
Richards	Stephen J	First initial only used.
Richards	S	First initial used.
Richards	Mr S	Title skipped, first initial used.
Richards	Rev Stephen J	Title skipped, first initial only used.

2.10 We can make our deduplication key stronger still by including the first initial of the first forename. We do not usually use the full forename field, since this is not rigorously entered: some records have the forename, some have the forename and any middle names, some just have the first initial of the forename. It is also necessary to recognise and strip out any titles which have been included in the forename field. Table 2 shows examples of how this system will match a variety of records with differently formatted forename fields.

2.11 Deduplication must be done intelligently to build up the most complete profile of the life being modelled. This means adding benefit amounts together and picking a valid postcode from the two or more fields on offer. As with the validation stage, it is useful to tabulate the most frequently occurring data items. For example, in one portfolio we have seen (not included here) the most common surname turned out to be SPOUSE, which immediately alerted us to the fact that spouse records had been mistakenly included in the extract. The calculation of the largest numbers of duplicates eliminated can also be useful: in one portfolio we have seen, the repeated use of dummy data in the administration system was quickly identified due to very high duplicate merging.

3. Geodemographic profiling

3.1 Actuaries in the United Kingdom are increasingly making use of so-called geodemographic models of mortality, primarily driven by postcode. The market for bulk annuities has long been driven by postcode for rating socio-economic group, and now products marketed directly to individual consumers are priced using postcode (Legal and General, 2007).

3.2 Postcodes were introduced to the United Kingdom by the state-owned Royal Mail for the purpose of automating the sorting of mail. U.K. postcodes are alphanumeric and have covered the entire country since 1974. The full list is available electronically from the Royal Mail as the Postcode Address File, and U.K. postcodes are copyrighted. Postcodes have been widely adopted beyond their original mail-sorting purpose, including consumer profiling for marketing, and premium calculations for general insurance and bulk-annuity pricing.

3.3 There are around 1.8m postcodes in the U.K., covering around 27m postcode addresses, of which around 1.6m postcodes are residential. A postcode can cover a whole street, part of a street, or even a single building. Around 200,000 postcodes are for commercial addresses only, and some are non-geographic (such as mailbox addresses). In practice an average of around 15 residential households are covered by a single postcode, providing a high degree of granularity in determining where a person lives just from their postcode alone. In most cases a combination of a house number and a postcode is enough to deliver a letter to the correct address.

3.4 Postcodes in the U.K. usually take the form of one of the following patterns: A9 9AA, A99 9AA, A9A 9AA, AA9 9AA, AA99 9AA or AA9A 9AA, where A signifies a letter and 9 a digit. Unfortunately there is no check digit, so there is no way of knowing if a conformant postcode is actually valid short of looking up a database of current valid postcodes. The first one or two characters are called the *postcode area*, of which there are 124 in the U.K. This corresponds to a geographic region and can thus be used to determine the broad location of the address. This can be used for modelling regional variations in mortality, although in practice we usually find that there is little or no regional variation after allowing for socio-economic factors. The first half of the postcode is known as the *postal district* and it is a common mistake to think that this contains all the usable data for postcode profiling. It does not: with an average coverage of 8,800 households, the postal district is much less homogeneous than the full postcode and is far less useful for modelling mortality differentials as a result.

Table 3. Some geodemographic profilers in the U.K.

Name	**Provider**	**Type codes**	**Sample type code and description for EH4 2AB**
Acorn	CACI	57	13 "Prosperous Professionals"
CAMEO	Eurodirect	57	5B "Young & Older Single Mortgagees"
FSS	Experian	45	E13 "Fully committed funds"
Mosaic	Experian	61	A02 "Cultural Leadership"

3.5 A number of commercial profilers will map U.K. postcodes onto a smaller number of socio-demographic profiles, as listed in Table 3. Each system has descriptive names and profiles for each category: for example, the postcode EH4 2AB is Mosaic type 02 ("Cultural Leadership"). These socio-demographic profiles were developed primarily for direct marketing purposes, but, as we shall see, they are particularly effective at predicting mortality differentials. When dealing with real data, postcodes are sometimes missing or fragmented (which we will assign to type 98) or else valid but for a commercial address (which we will assign to type 99). Thus, when using the Mosaic system we will have 63 type codes (61 + 2), whereas we will have 59 type codes using the Acorn system (57 + 2) and 47 (45 + 2) with FSS. The Royal Mail typically recodes or reassigns postcodes continuously, so any geodemographic profiling needs to be updated annually (as do any models based on these profiles).

3.6 Similar postcode-driven systems apply in other countries, including the United States of America (zip code), Canada (postal code) and the Netherlands (postal code). As in the U.K., these countries use hierarchical systems, so a given postcode can be used to give both regional and socio-economic information. Similar modelling techniques can be applied to other countries, but the full address is usually required for socio-demographic modelling, not just the postcode. For example, the German *Postleitzahl* 89079 tells you the policyholder is in the area of Donaustetten in Baden-Württemberg, but this covers hundreds of households and cannot on its own be used for socio-economic profiling. The analogy would be just using the 'EH4 2' part of the full postcode 'EH4 2AB': the former is called the *postcode sector* and covers hundreds of households. This would be of limited use for socio-demographic profiling, although postcode sectors form the basis of the simpler Carstairs scores (McLoone, 2000) for assessing deprivation, and they work best when the sector is relatively homogeneous.

4. Limitations and pitfalls of geodemographic profiling

4.1 It is important to note some pitfalls in using geodemographic profiles. A particular issue is where a block of business has missing postcodes for a particular reason. One example is where pensions are paid to a trustee for onward forwarding to the pensioner and so the insurer holds no addresses for the pensioners. This gives us the situation where a specific and distinct class of business is not profiled for a systematic reason, rather than having profiles missing at random. In this case people with missing postcodes could have markedly higher mortality because a missing postcode was simply a marker for bulk-annuity pension-scheme business instead of true individual money-purchase annuities. Another example is a life office where annuitants were given the head-office address upon death. The vast majority of deaths thus all had a non-residential postcode and therefore ended up in category 99, which (unsurprisingly!) proved to be a category of very high mortality. There may also be a connection with lower mortality: foreign and overseas addresses will not be profiled, and so end up coded 98. If wealthier annuitants are disproportionately likely to live overseas, or if death reporting is less prompt than in the United Kingdom, then code 98 will be predictive of low mortality. Commercial services are available which can reformat address databases to Postcode Address File (PAF) format, which corrects and updates postcodes as well as filling in missing ones where the address is recognised. Such address "cleaning"is inexpensive and is a cost-effective way of boosting the power of a geodemograhic model of mortality.

4.2 One way to detect thess sorts of data problems is to calculate the Cramer's V statistic for all categorical variables. Cramer (1999) defines a statistic measuring the strength of association or dependency between two categorical variables. It takes the value 0 for no association, and the value 100 where two variables are perfectly associated and knowledge of one variable completely specifies the other. We include the death status as a categorical variable as well, which helps identify the sort of situation described in ¶4.1.

Table 4. Cramer's V statistic for life-office pensioner data set (all ages)

	Gender	**Region code**	**Size band**	**Status**	**Type**
Birth year	21.6	3.1	11.4	54.4	4.0
Gender		4.8	16.1	12.4	5.6
Region code			5.9	6.4	20.6
Size band				17.4	9.7
Status					10.4

Source: Own calculations using life-office annuitant data. The status variable takes the value 1 for a death, zero otherwise. The type variable is Experian's Postcode Mosaic Type (61 levels, plus two further levels for commercial addresses and unrecognised postcodes). The region code is the U.K. region extracted from the postcode (124 levels). The relatively high association between type and region code comes from the group of unrecognised postcodes, which are assigned a dummy type code of 98 and a dummy region code of XX.

4.3 Table 4 shows the results of all two-way associations between the categorical variables for the life-office data set. The Cramer's V statistic is symmetric, so only the values in the upper right of the matrix are shown. The diagonal is not shown, as all the values are 100 (a variable is always perfectly associated with itself). In a clean data set such as this one, the strongest association should usually be between year of birth and death status (people with earlier years of birth should be more likely to be dead). If the association between death status and type were larger, this would be evidence of one of the systematic issues in ¶4.1. Interestingly, we can see relatively weak association between pension size band and geodemographic type (9.7), which is perhaps surprising as we might expect them to both be proxies for socio-economic group. In general we find that a model with age, gender and geodemographic type can fit better than one driven by age, gender and pension size. However, for the sort of reasons outlined in ¶4.1, a model which incorporates both geodemographic type *and* pension size will usually be better than using either variable on its own. The association between gender and year of birth is relatively strong because of the tendency for males to be the first life and for females to be surviving spouses.

4.4 It is instructive to examine why geodemographic profiles might be more powerful predictors of pensioner mortality than pension size. In each case, the variable in question is merely acting as a *proxy* for the true underlying drivers of mortality differentials: smoking, diet, drinking and other health behaviours (or absence of them). In the presence of information on smoker status in a model, for example, we would expect a much-reduced impact of proxies for socio-economic group, such as pension size or geodemographic type. When the geodemographic profiles result in a better-fitting model than one based on pension size, this simply tells us that they are a better proxy for the underlying differentials. Indeed, we tend to find the importance of pension size reduces as new risk factors are added to a model. For example, the addition of a factor for birth cohort will often further reduce the role of pension size, since earlier birth cohorts tend to have smaller pensions than later generations. This reducing role of pension size is echoed by Richards and Jones (2004), who rated pension size as only the fifth most important rating factor in a model of annuitant mortality.

5. Methodology

5.1 Following Macdonald (1996a, b, c) we will be modelling using the instantaneous mortality hazard, known to actuaries as the force of mortality, as this makes better use of the data than q-type rates. To illustrate this, consider two groups each consisting of four lives alive at the start of the year. During the course of the year one life dies in each group, making the estimated mortality rate, $\hat{q}_A = \hat{q}_B = \frac{1}{4}$ in both cases. If the death in group A occurs at the end of January, the estimated force of mortality is $\hat{\mu}_A = \frac{1}{3\frac{1}{12}} = \frac{12}{37}$. If the death in group B occurs at the start of December the estimated force of mortality is $\hat{\mu}_B = \frac{1}{3\frac{11}{12}} = \frac{12}{47}$. As this simple example shows, working with the force of mortality means we can use all the information available, and will usually result in a better model. In contrast, working with q-type rates throws away the information on time of death and is therefore less sophisticated.

5.2 In this paper we will not model mortality of groups, however, since we have detailed information on each individual life. As we will see later, this gives us far greater power in modelling mortality than can be done using a GLM for Poisson counts. We therefore need to define some simple results at the level of the individual. We start with the hazard rate at age x, μ_x, which is given by:

$$\mu_x = \lim_{h \to 0^+} \frac{1}{h} \Pr\left(\text{death before age } x+h | \text{alive at age } x\right) \tag{1}$$

5.3 The probability of surviving from age x to age $x+t$, ${}_tp_x$, is given by:

$${}_tp_x = e^{-H_x(t)} \tag{2}$$

where $H_x(t)$ is the integrated hazard function:

$$H_x(t) = \int_0^t \mu_{x+s} ds \tag{3}$$

5.4 For each life i of n lives we have (i) an entry age, x_i, (ii) a time observed, t_i, and (iii) an indicator variable, d_i, for the state of the life at age $x_i + t_i$. The variable d_i takes the value 0 on survival and 1 on the event of interest. This event can be death (as in this paper) or any other decrement of interest, such as critical-illness claim, lapse or surrender. The likelihood function, L, is therefore given by:

$$L \propto \prod_{i=1}^{n} {}_{t_i}p_{x_i} \mu_{x_i+t_i}^{d_i} \tag{4}$$

and taking the natural logarithm of L gives us the log-likelihood function, ℓ:

$$\ell = \sum_{i=1}^{n} -H_{x_i}(t_i) + \sum_{i=1}^{n} d_i \log \mu_{x_i+t_i} \tag{5}$$

5.5 Thus, when applying survival models to individual data, it simply suffices to specify the structure of the hazard rate, μ_x, and subsequently derive $H_x(t)$. When fitting any model, we choose the parameter values to maximise the log-likelihood function in Equation 5. We have used Longevitas for all the models in this paper, which uses derivatives-based methods for speed and reliability in maximising log-likelihoods.

5.6 Many implementations of survival models at the individual level deal with age-varying mortality through a variable transformation. This demands that the lives be observed from outset, i.e. from birth if chronological age is to be used directly. In contrast, people tend only to start life-insurance contracts or pension benefits when they are well into adult life. The lifetimes observed are called *left-truncated*, since observation starts at age x_i and we have no data on deaths and exposure prior this age. Equally, when an extract of mortality data is taken not all lives will be dead at the extract date. Such data is called *right-censored*, since all that can be said of the mortality process is that it will occur some time after t_i years. Survival models based on transforming the exponential distribution can handle right-censorship easily enough, but left-truncation usually poses a problem. However, by dealing directly with the log-likelihood in Equation 5 we can automatically handle left-truncation.

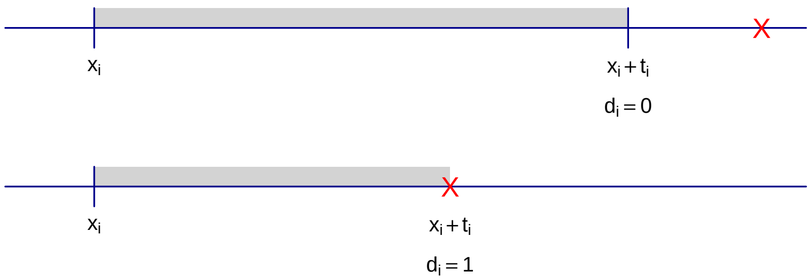

Figure 1. Diagram of survival-model setup. The time observed, t_i, is shown in grey, while deaths are marked with a cross, $\times$. Since people do not usually enter into life-insurance contracts at birth, observations are left-truncated, i.e. lives start being observed at age $x_i > 0$. The upper case is an example of right-censored data as death happens after the end of the observation period.

6. Mortality laws

6.1 A major advantage of fitting a statistical model is that smoothness is built-in and there is no need to separately graduate the resulting fitted rates. This is known as graduation by mathematical formula. One approach to defining μ_x is to use penalized splines, which can be done in one or more dimensions as described in Currie, Durban and Eilers (2004). This is a flexible means of capturing different shapes of mortality pattern. However, in this paper we will assume mortality follows some sort of law, which has the benefit of requiring fewer parameters.

Table 5. Some mortality laws and their corresponding integrated hazard functions, $H_x(t)$

Mortality law	μ_x	$H_x(t)$
Constant hazard	e^{α}	te^{α}
Gompertz (1825)	$e^{\alpha+\beta x}$	$\frac{(e^{\beta t}-1)}{\beta}e^{\alpha+\beta x}$
Makeham (1859)	$e^{\epsilon}+e^{\alpha+\beta x}$	$te^{\epsilon}+\frac{(e^{\beta t}-1)}{\beta}e^{\alpha+\beta x}$
Perks (1932)	$\frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$	$\frac{1}{\beta}\log\left(\frac{1+e^{\alpha+\beta(x+t)}}{1+e^{\alpha+\beta x}}\right)$
Beard (1959)	$\frac{e^{\alpha+\beta x}}{1+e^{\alpha+\rho+\beta x}}$	$\frac{e^{-\rho}}{\beta}\log\left(\frac{1+e^{\alpha+\rho+\beta(x+t)}}{1+e^{\alpha+\rho+\beta x}}\right)$
Makeham-Perks (1932)	$\frac{e^{\epsilon}+e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$	$te^{\epsilon}+\frac{(1-e^{\epsilon})}{\beta}\log\left(\frac{1+e^{\alpha+\beta(x+t)}}{1+e^{\alpha+\beta x}}\right)$
Makeham-Beard (1932)	$\frac{e^{\epsilon}+e^{\alpha+\beta x}}{1+e^{\alpha+\rho+\beta x}}$	$te^{\epsilon}+\frac{(e^{-\rho}-e^{\epsilon})}{\beta}\log\left(\frac{1+e^{\alpha+\rho+\beta(x+t)}}{1+e^{\alpha+\rho+\beta x}}\right)$

6.2 Some of the models in Table 5 are related to the Cox model. For example, the Gompertz model can be expressed as a proportion of a baseline hazard, albeit possibly as a time- or age-varying proportion. The Makeham model, however, cannot be expressed in terms of a baseline hazard due to the non-multiplicative e^ϵ term. The models in Table 5 are mainly non-linear in their nature, although this has not led to any real difficulties in fitting them. Here we have used derivatives-based methods for optimising the log-likelihood, with an explicit formulaic calculation of the information matrix for inversion to calculate the covariance matrix. For converting into mortality rates, q_x, for use in actuarial systems we use the exact formula:

$$q_x = 1 - e^{-H_x(1)} \tag{6}$$

6.3 Note that the naming convention in Table 5 is different from what might be seen elsewhere. For example, the model identified above as Makeham-Beard was proposed by Perks (1932). We have opted (i) to use the term Makeham wherever the constant e^ϵ appears, (ii) to name the simple logistic form $\frac{e^\bullet}{1+e^\bullet}$ after Perks, and (iii) to use the term Beard wherever the logistic form has a so-called heterogeneity parameter, ρ, whose role and derivation will be explained next.

6.4 One way to re-write the Gompertz law is $ze^{\beta x}$, where $z = e^\alpha$. If the members of the population are heterogeneous, then a so-called *frailty model* has each individual i with their own personal value of z_i. If the z_i are assumed drawn from a gamma distribution and fixed throughout life, then the population hazard rate is that of the Beard model even when the hazard of each individual is Gompertz. Similarly, heterogeneous individual Makeham mortality results in a population hazard rate following the Makeham-Beard law. These frailty-type arguments were first advanced by Beard (1959).

6.5 These results mean that the shape of the mortality law at the population level need not be the same as the mortality law acting on each individual. This phenomenon could be turned around and used as a test for the presence of unexplained heterogeneity: if individual mortality is assumed to follow the Gompertz law, for example, yet the Beard model fits the overall data set better, then one could conclude that there is further unexplained variation not covered by the model fitted. As a model is developed, the significance of the Beard parameter, ρ, could be used as a guide as to whether there are further risk factors to be found. Note that this is not a given: there are other structures which can lead to the Beard law, of which a gamma-distributed frailty is just one. For example, Vaupel and Yashin (1985) give a number of detailed examples where heterogeneity can cause unexpected observed effects.

6.6 Appendices 1 and 2 give proofs of the Beard-type laws arising from heterogeneity in α for both the Gompertz and Makeham laws. Appendix 3 shows how the Makeham-Beard law arises from a *cascade process*, where the force of mortality is related to the number of accumulated defects in an organism. Appendix 4 shows that a Gompertz force of mortality is equivalent to assuming that the future lifetime is a random variable from the extreme-value distribution, while the Beard law equates to a future lifetime from the logistic distribution.

7. Modelling mortality differentials

7.1 One of the key goals for actuaries in assessing risk is to group people into pools of similar risk. We are using here a model for the mortality of each individual, and we do this by assuming each life i has its own specific parameters which describe its own combination of risks. In the context of a Gompertz law for mortality, this means α_i and β_i for life i instead of a group value of α and β. Thus:

$$\begin{aligned} \alpha_i &= a_{baseline} + \sum_{j=1}^{m} z_{ij} a_j \\ \beta_i &= b_{baseline} + \sum_{j=1}^{m} z_{ij} b_j \end{aligned} \tag{7}$$

where there are m components (factors) to the overall risk, each a_j and b_j is a parameter for a particular risk (a_j) and its interaction with age (b_j), and z_{ij} is a binary indicator variable taking the value 1 when life i has risk factor j and the value 0 otherwise. Note that we have assumed an age interaction for each main effect in the model purely for simplicity. For example, in a model with risk factors for both gender and smoker status:

$$\begin{aligned}\alpha_i &= a_{baseline} + z_{i,male}a_{male} + z_{i,smoker}a_{smoker} \\ \beta_i &= b_{baseline} + z_{i,male}b_{male} + z_{i,smoker}b_{smoker}\end{aligned} \tag{8}$$

7.2 Note that the model is structured as measuring differences from a baseline profile. In the model specified in Equation 8, the baseline is a female non-smoker, while the model parameters measure male mortality as a departure from the female baseline, and smoker mortality is measured as a departure from the non-smoker baseline. The $z_{i,male}$ are zero-one indicator variables for whether life i is male, and the $z_{i,smoker}$ are similar zero-one indicators for whether a life is a smoker. The advantage of this structure is that there is no minimum group size required, which means there is no minimum number of lives required and no upper limit to the number of risk factors which can be investigated with this approach. This is particularly useful for portfolios which are by their very nature small, but where there is rich data available on each individual life. This approach also provides substantial benefits where the data set is rich in individual details, but the number of events is relatively small. A good example would be in term-assurance portfolios, where this approach could be useful for a reinsurer trying to make commercial pricing decisions.

7.3 However, many potentially useful rating factors are not easy to use directly. For example, there are 124 postcode regions in the U.K., so fitting this directly in a model would require 123 parameters (one is absorbed into the baseline). Similarly, the Mosaic type has 63 levels, so using this directly as a socio-economic factor would require 62 parameters. In each case we would have an unwieldy and over-parameterised model.

7.4 One solution is to group complex factors into simpler meta-factors, say to divide the postcode regions among three broad regional groupings, or to put the 63 Mosaic types into four or five lifestyle groups. There is a variety of ways of assigning these groups, but the one we will use in this paper is to find the optimal assignment by fitting hundreds (if not thousands) of alternative models and choosing the best-fitting one. This has the advantage of removing any human subjectivity from the groupings. The means whereby we assess the goodness of fit is the AIC (Akaike, 1987), which is defined as:

$$AIC = -2\ell + 2n \tag{9}$$

where n is the number of parameters used in fitting the model. A lower value of the AIC indicates a better-fitting model, as minimising the AIC is equivalent to maximizing the log-likelihood function for a given number of parameters. This does carry the risk of some sub-groups finding themselves in the "wrong" final group by random variation. If there is additional knowledge of which sub-groups belong together (or should be kept apart) then such constraints can be added to the search algorithm for minimising the AIC.

7.5 This approach can be used both for categorical factors — such as region, socio-economic group, product type etc — and for ordinal factors such as pension size and year of birth. By splitting the pensioners into a large number of equal-sized bands, the same process of minimizing the AIC will give us the optimum break-points for pension-size categories. We impose an additional restriction on such ordinal factors, namely that the resulting groups must be contiguous ranges. Thus, for a categorical factor like Mosaic, if type codes 50 and 52 are in the same group, then type code 51 is free to be in a different group. In contrast, for an ordinal factor like year of birth, if the years 1920 and 1922 are in the same group, then 1921 must be as well. Treating a factor as categorical will always give at least a good a fit as treating it as ordinal, but it is important to respect the additional structure of ordinal variables.

8. Modelling considerations

8.1 The first consideration is the date from which to start the modelling. Where deaths data has been archived, for example, modelling should start at that date and not before. For example, if a pension had a commencement date in 1980, but all deaths were archived from the administration system in 2000, then modelling must start at the later date to avoid under-estimating mortality. Equally, one has to be careful about the choice of end date for modelling due to the tendency for delays in death reporting. For example, if an extract of data were taken in June 2007, one might only want to model mortality as far as end-2006 to ensure late-reported deaths were not a material issue.

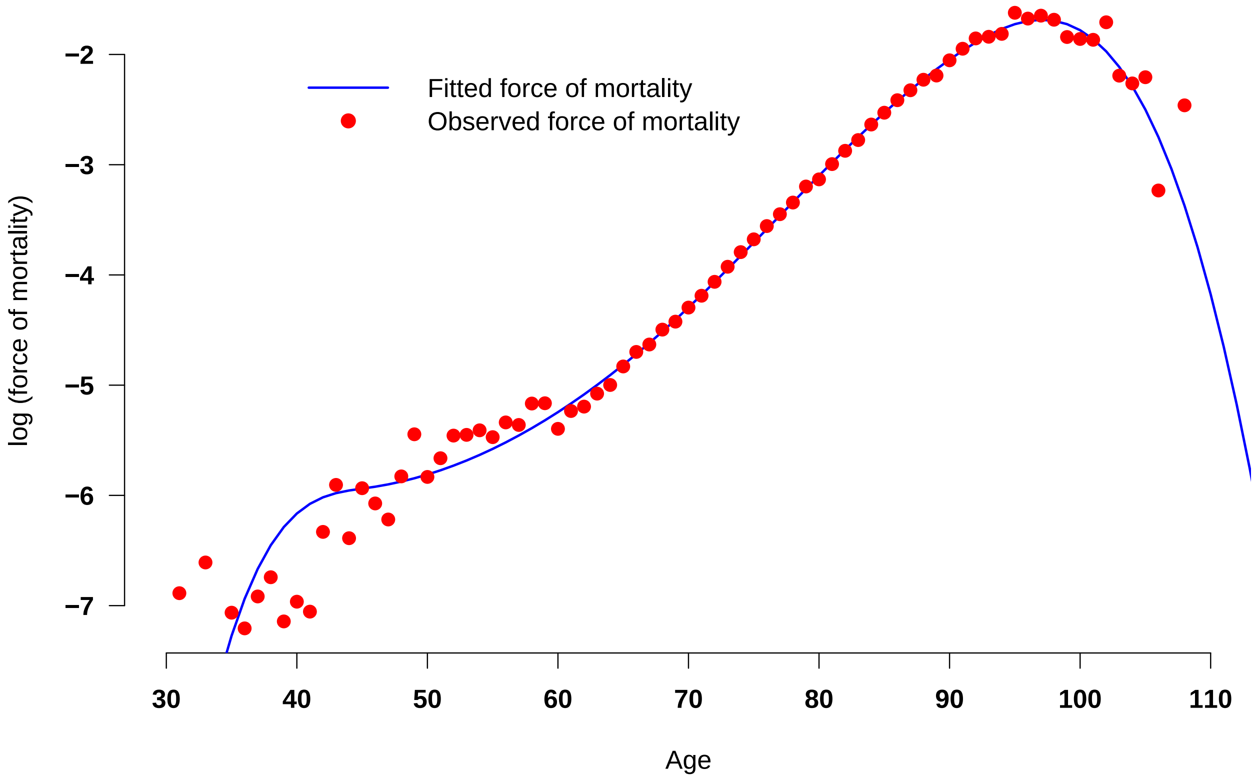

Figure 2. Force of mortality for pensioners between ages 30 and 110: observed crude force of mortality (•) together with fitted values from P-spline regression. Only the mortality between ages 60 and 95 shows regular behaviour suitable for a mortality law, with evidence of data-quality problems above age 95. Source: Own calculations using mortality experience of a portfolio of life-office pensioners.

8.2 Before fitting a model it is necessary to consider the nature of the data. For example, we find that the mortality of pensioners below age 60 does not exhibit the simple and straightforward patterns of mortality above age 60 — see Figure 2. For this reason we typically truncate exposure (and exclude deaths) prior to age 60 when modelling pensioner mortality. We also find that mortality data is often unreliable above age 95–100 – see again Figure 2— so we typically truncate exposure (and exclude deaths) after 95 for this data set. We also often find that some lives are suspiciously long-lived, such as when numerous pensioners are apparently older than 105 years. We remove such cases from the exposure calculation entirely, since their apparent high age is often an artefact of the way records have been entered or stored on the administration system. An example would be an orphan's or guaranteed pension which had been set up on the payment system with a date of birth of 01/01/01 simply in order to fill a required field. If such cases are not excluded, mortality at older ages will be under-stated.

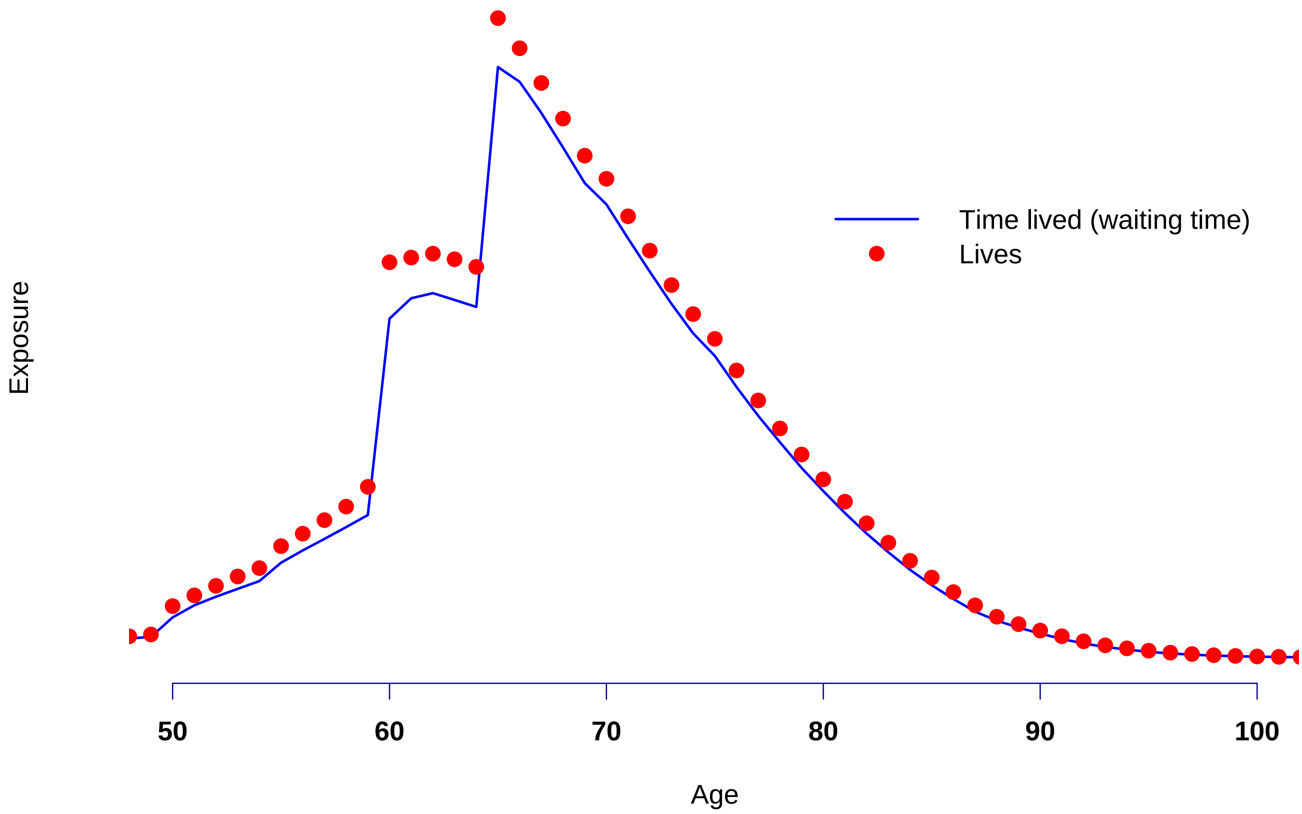

Figure 3. Distribution of lives and exposure. The discontinuities at ages 60 and 65 mark the two most common ages at retirement. The exposure line is always less than the number of lives: the exposure is the time lived (or waiting time) during the year of age and is therefore lower due to the fractional years of life lost due to death. Source: Own calculations using mortality experience of a portfolio of life-office pensioners.

9. Model results

9.1 We decided to model mortality from 2000 onwards in order to be confident about the quality of recent data. Although the data extract was taken in mid-2007, we stopped modelling at end-2006 as this was felt to be the latest date not materially impacted by delays in death reporting. We start by fitting a model with age and intercept only, i.e. no attempt to model sub-categories of risk. As a practical aside, we must remember that it is the *difference* in AIC which is important, not the absolute value: a difference of 2 (say) in the AIC counts as statistically significant.

Table 6. Initial model with age only

Mortality law	**AIC**	**AIC relative to Gompertz**	**Parameters**
Gompertz	386742	0	2
Makeham	386744	2	3
Perks	386618	-124	2
Beard	386560	-182	3
Makeham-Perks	386620	-122	3
Makeham-Beard	386559	-183	4

Source: Own calculations using mortality experience of life-office pensioners aged between 60 and 95 between 2000-2006.

9.2 Table 6 shows the results for sitting a simple model with age only. One of the first features of interest is that adding a parameter does not always improve the model fit: the Makeham model has a *higher* AIC than the Gompertz model. Thus, just comparing these two models there seems little support for a material Makeham constant component to mortality. The exception lies with the Makeham-Beard model, which has a better AIC compared to the Beard model. Since the only difference between the two lies in the Makeham parameter, this might suggest that a constant component to mortality can best be identified in the presence of the heterogeneity parameter, ρ.

9.3 Another interesting feature is how much better the logistic forms — Perks, Beard, Makeham-Perks and Makeham-Beard — compare to the others. The Perks law is a simple twist on the Gompertz law, and both involve only two parameters, yet the Perks model has an AIC which is 124 lower, a significant improvement merely from restructuring the same number of parameters into a different functional form.

9.4 The final feature of note in Table 6 is that the Beard model has an AIC 182 lower than the Gompertz model, from which it is derived using a frailty approach (see ¶6.4 and Appendix 1). Similarly, the Makeham-Beard model has an AIC 185 lower than the Makeham model from which it can be derived using the same approach (see ¶6.4 and Appendix 2). If we accept the interpretation of the Beard ρ parameter as an indicator of unexplained variation, this suggests that there is significant unexplained heterogeneity present. Of course, we know what one of these sources is — gender — since we have deliberately excluded it from the model.

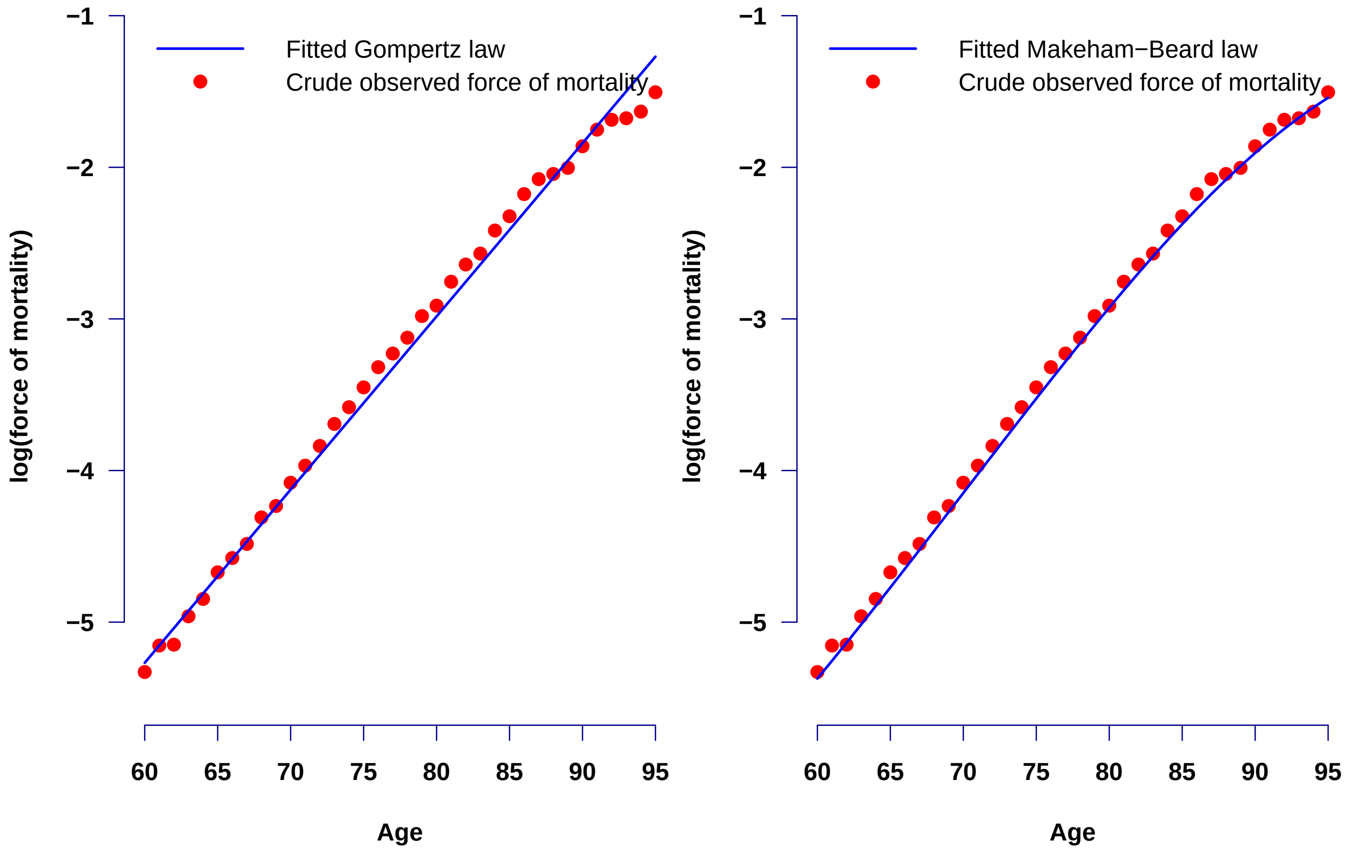

Figure 4. Comparison of Gompertz and Makeham-Beard models for age only. Exposure and deaths above age 95 have been excluded as evidence in Figure 2 suggested that such data was unreliable. Delays in death reporting are felt to be minimal, since the modelling was done up to end-December 2006, at least six months before the extract of data was taken. Source: Models fitted in Table 6.

9.5 Figure 4 shows the fitted force of mortality and the crude observed forces of mortality for two of the models in Table 6. The crude rates show a near-linear increase in log(mortality) — i.e. exponential increase in mortality — over a wide age range, followed by a decelerating rate of increase in mortality. This deceleration is a feature of many populations, and is referred to as *late-life deceleration* by Gavrilov and Gavrilova (2001). Models which reproduce this feature — Perks, Beard, Makeham-Perks and Makeham-Beard — tend to fit better than models which do not. Note that neither of the illustrated models in Figure 4 would be regarded as “good”— the Gompertz model has long runs of over- and under-fitting, while the Makeham-Beard model shows evidence of consistent bias towards under-stating the force of mortality. However, the deviations between the observed rates and the fitted curves are smaller for the Makeham-Beard model, which is what gives it the lower AIC value in Table 6. We can now turn to a more realistic model incorporating different values of α and β according to gender.

Table 7. Model with age and gender, i.e. Age*Gender

Mortality law	**AIC**	**AIC relative to Gompertz**	**Parameters**
Gompertz	384824	0	4
Makeham	384826	2	5
Perks	384765	-59	4
Beard	384761	-63	5
Makeham-Perks	384762	-62	5
Makeham-Beard	384728	-96	6

Source: Own calculations using mortality experience of life-office pensioners aged between 60 and 95 between 2000-2006. Note that the AICs can be compared directly with those in Table 6.

9.6 As expected, Table 7 shows that adding gender to the model has made a very large improvement: the AIC has dropped by between 1799 and 1918, depending on the model. Table 7 also shows many of the same features as Table 6. The one most of interest to us is that the Beard model has an AIC 63 lower than the Gompertz model, from which it is derived using a frailty approach. Similarly, the Makeham-Beard model has an AIC 96 lower than the Gompertz model (or 98 lower than Makeham model from which it can be derived using the same frailty approach which links the Gompertz and Beard models). The gap between the frailty models and their antecedents has shrunk between Table 6 and Table 7, reflecting the fact that including gender in the model has substantially reduced the unexplained heterogeneity. If we accept the interpretation of the Beard ρ parameter as an indicator of unexplained variation, the fact that there are still significant gaps between the two suggests that there is still significant unexplained heterogeneity present.

9.7 According to Richards and Jones (2004), one major source of further heterogeneity is likely to be socio-economic group, which we can investigate using the traditional actuarial proxy of pension size. We split the population into fifty equal-sized groups of pensioners and look for the optimal breakpoints giving us three size bands. There are 1,176 unique combinations of assigning fifty groups to three size-bands, while both preserving the ordinal structure of the original variable and having at least one group in each band. The optimal breakpoints are determined by minimizing the AIC.

Table 8. Model with age, gender and pension size-band, i.e. Age*(Gender+SizeBand)

Mortality law	**AIC**	**AIC relative to Gompertz**	**Parameters**
Gompertz	383562	0	8
Makeham	383564	2	9
Perks	383515	-47	8
Beard	383513	-49	9
Makeham-Perks	383510	-52	9
Makeham-Beard	383486	-76	10

Source: Own calculations using mortality experience of life-office pensioners aged between 60 and 95 between 2000-2006. The SizeBand variable is derived from fifty bands of equal numbers of annuitants, optimised into a three-level ordinal factor. Note that the AICs can be compared directly with those in Tables 6 and 7.

9.8 The results of using a three-level grouping based on pension size are shown in Table 8. In all cases using pension size has made a very material improvement in the model fit, as evidenced by drops in the AIC of between 1242 and 1262, depending on the model used. Again we see that the Makeham model fits less well than the Gompertz, and that the logistic models are generally better than the others. As before, the Makeham-Beard model has performed best, with an AIC 76 units lower than the Gompertz model. If we accept the role of the Beard parameter, ρ, the gap between the Gompertz and Beard AICs (as well as the gap between the Makeham and Makeham-Beard AICs) suggests that there is still significant unexplained variation. However, the fact that the gap in AIC due to the Beard parameter has narrowed from 63 to 49 suggests that at least some of this heterogeneity has been accounted for. While the optimisations were free to set different break-points for the pension size-bands, in each case the optimum was achieved by setting the first break-point between the 14^{th} and 15^{th} fiftieths and by setting the second between the 41^{st} and 42^{nd}. This means that the first group consists of the 28% pensioners with the smallest pensions, the second group with the next 54% and the third group with the 18% of pensioners with the largest pensions.

9.9 The alternative to using pension size is to profile the population according to geodemographic type. This we do using Experian's postcode Mosaic profiles (we could also use CACI's Acorn system, Eurodirect's CAMEO code, or Experian's alternative FSS classification). We then optimise the assignment of the geodemographic types to three lifestyle groups. The optimisation is determined as before by minimizing the AIC, but here the types are treated as categorical, i.e. a type code is free to belong to a different lifestyle group than its immediately adjacent neighbours.

Table 9. Model with age, gender and lifestyle group derived from postcode via Mosaic type, i.e. Age*(Gender+Lifestyle)

Mortality law	**AIC**	**AIC relative to Gompertz**	**Parameters**
Gompertz	383537	0	8
Makeham	383539	2	9
Perks	383518	-19	8
Beard	383520	-17	9
Makeham-Perks	383513	-24	9
Makeham-Beard	383509	-28	10

Source: Own calculations using mortality experience of life-office pensioners aged between 60 and 95 between 2000-2006. The Lifestyle variable is derived by finding the mapping of the Postcode Mosaic Type which minimises the AIC for the number of lifestyle categories. Note that the AICs can be compared directly with those in Tables 6, 7 and 8.

9.10 The results of using a geodemographic type based on postcode are shown in Table 9. Comparing the figures with the equivalents in Table 8 we can see that models based on geodemographic type can have lower AICs than models based on pension size, depending on the choice of mortality law, although they are better described as similarly significant overall.

9.11 Another interesting feature of Table 9 is that the Makeham-Beard model is still the best-fitting. Both of the laws with the Beard frailty parameter, ρ, fit better than their precursor models. If we accept the interpretation of the Beard parameter as an indicator of further unexplained variation, this would suggest that, despite the large improvement, there is still some remaining unexplained heterogeneity amongst the pensioners.

9.12 One concern about relying on geodemographic codes might be some kind of systematic data error which gives the illusion of predictive power. For example, we have encountered life offices where address data is deleted upon death. In this case, the geodemographic profile would be 98 (unknown) and this code would only be "predictive" of death merely because of the treatment of addresses on death. We have also encountered life offices where the pensioner's address is changed to be that of the head office. In this case, the geodemographic profile would be 99 (non-residential postcode), and, again, this code would only be "predictive" due to the link to death processing. For this data set, however, we see from Table 4 that the association between death status and geodemographic type (10.2) is actually much less than the association between death status and pension size (17.4). This suggests that this particular data set does not suffer from any obvious corrupting linkage between postcode and death processing, so the geodemographic type is a genuinely powerful predictor of mortality.

9.13 The primary purpose of both geodemographic type and pension size lies in acting as a proxy for socio-economic differentials. However, while we have shown that both geodemographic type and pension size can do this, the Cramer V statistic in Table 4 suggests relatively little linkage between type code and pension size, so we might expect some benefit from using both variables in the model. In this data set 10.5% of pensioners do not have a geodemographic type, i.e. those of type 98 and 99, and for them pension size is the only available proxy for socio-economic status. We would therefore expect the predictive power of pension size to be linked to the number of lives without geodemographic profiles. Furthermore, one can imagine that two people of identical geodemographic type will have different standards of living if one has several times the pension income of the other. The converse is also true: two people with equal pension sizes might be distinguished by their postcode-identified lifestyle. Other reasons why pension size is useful in addition to postcode might be where a wealthier person chooses to live in a less salubrious postcode — perhaps due to family or other ties — or has multiple addresses and has registered the policy to an address which is not representative of their overall lifestyle. In such cases, pension size will reveal additional insights into a pensioner's lifestyle which a postcode cannot achieve on its own. Thus, we might expect a model using both postcode and pension size to perform better because each can compensate for the other's failings.

Table 10. Model with age, gender, lifestyle and pension size, i.e. Age*(Gender+SizeBand+Lifestyle)

Mortality law	AIC	AIC relative to Gompertz	Parameters
Gompertz	382597	0	12
Makeham	382599	2	13
Perks	382583	-14	12
Beard	382583	-14	13
Makeham-Perks	382575	-22	13
Makeham-Beard	382576	-21	14

Source: Own calculations using mortality experience of life-office pensioners aged between 60 and 95 between 2000-2006. The Lifestyle variable is derived from optimising a mapping of the Postcode Mosaic Type. The SizeBand variable is derived from fifty bands of equal numbers of annuitants, optimised into a three-level ordinal factor. Both Lifestyle and SizeBand variables were re-optimised in the presence of each other to check the original mappings still hold true in this more-complicated model. In both cases they produced the same definitions of Lifestyle and SizeBand as previously. Note that the AICs can be compared directly with those in Tables 6, 7, 8 and 9.

9.14 In Table 10 the gap between the AICs of the Gompertz and Beard models has shrunk further from 17 to 14. Using the interpretation of the Beard ρ parameter as a marker for unexplained heterogeneity, this suggests that a model using both geodemographic type and pension size is better at explaining heterogeneity than either variable on its own. This is reflected in the large falls in AIC between Table 9 and Table 10. Comparing just the results for the Makeham-Beard model, using pension size on its own improved the AIC by 1242 ($= 384728 - 383486$), whereas using postcode-driven lifestyle groups improved the AIC by 1219 ($= 384728 - 383509$). However, using both factors together improved the AIC by 2152 ($= 384728 - 382576$). This improvement is far larger than would be expected from pension size acting as a proxy for the unprofiled cases, so there seems to be a genuine pension-size effect independent of the geodemographic type. Part of this may also be the cohort effect (Richards, Kirkby and Currie, 2005), since older and long-retired people will tend to have smaller pensions.

Table 11. Impact of pension size and lifestyle

Gender	Pension size	Lifestyle	e_{65}	$\bar{a}_{65}^{5\%}$	$\bar{a}_{65}^{2.5\%}$	Change in $\bar{a}_{65}^{5\%}$	Change in $\bar{a}_{65}^{2.5\%}$
Female	Highest	Upper	22.88	13.26	17.05	n/a	n/a
Male	Highest	Upper	20.23	12.23	15.43	-7.8%	-9.5%
Male	Highest	Lower	18.56	11.50	14.34	-6.0%	-7.1%
Male	Middle	Lower	17.06	10.83	13.36	-5.8%	-6.8%
Male	Lowest	Lower	15.62	10.12	12.37	-6.6%	-7.4%
Overall						-23.7%	-27.4%

Source: Own calculations using mortality experience of life-office pensioners aged between 60 and 95 between 2000-2006. Annuity reserves and life expectancies at age 65 calculated assuming continuous payment of 1 per annum, continuous interest at 5% or 2.5% per annum, and mortality according to the Makeham-Beard model and parameters from Table 10. Although three levels were fitted for lifestyle, only the upper and lower types are shown here: 88.0% of the other level had missing postcodes, so this cannot be said to be a true lifestyle classification. Note that these calculations have no allowance for future improvements.

9.15 Table 11 shows the impact of pension size and lifestyle group on life expectancy and annuity value. These figures are calculated on as close a basis as possible to the equivalent tables in Richards and Jones (2004) to facilitate comparison. The overall change of 23.7% in Table 11 is less than the 33.6% in Richards and Jones (2004) as the latter used a larger number of levels for lifestyle. This enabled the identification of a greater range of lifestyle sub-groups, which results in a greater range of life expectancies. To put the values in Table 11 in perspective, the equivalent complete life expectancy using the PCA00 table is 18.4 years for males and 20.9 for females, and a typical pricing margin for an annuity at the time of writing is around 4–5% of the best-estimate liability. Since these mortality differentials can change the annuity factor by more than the pricing margin, accurate modelling of mortality is very important to the profitability of a life office.

10. Checking the model fit

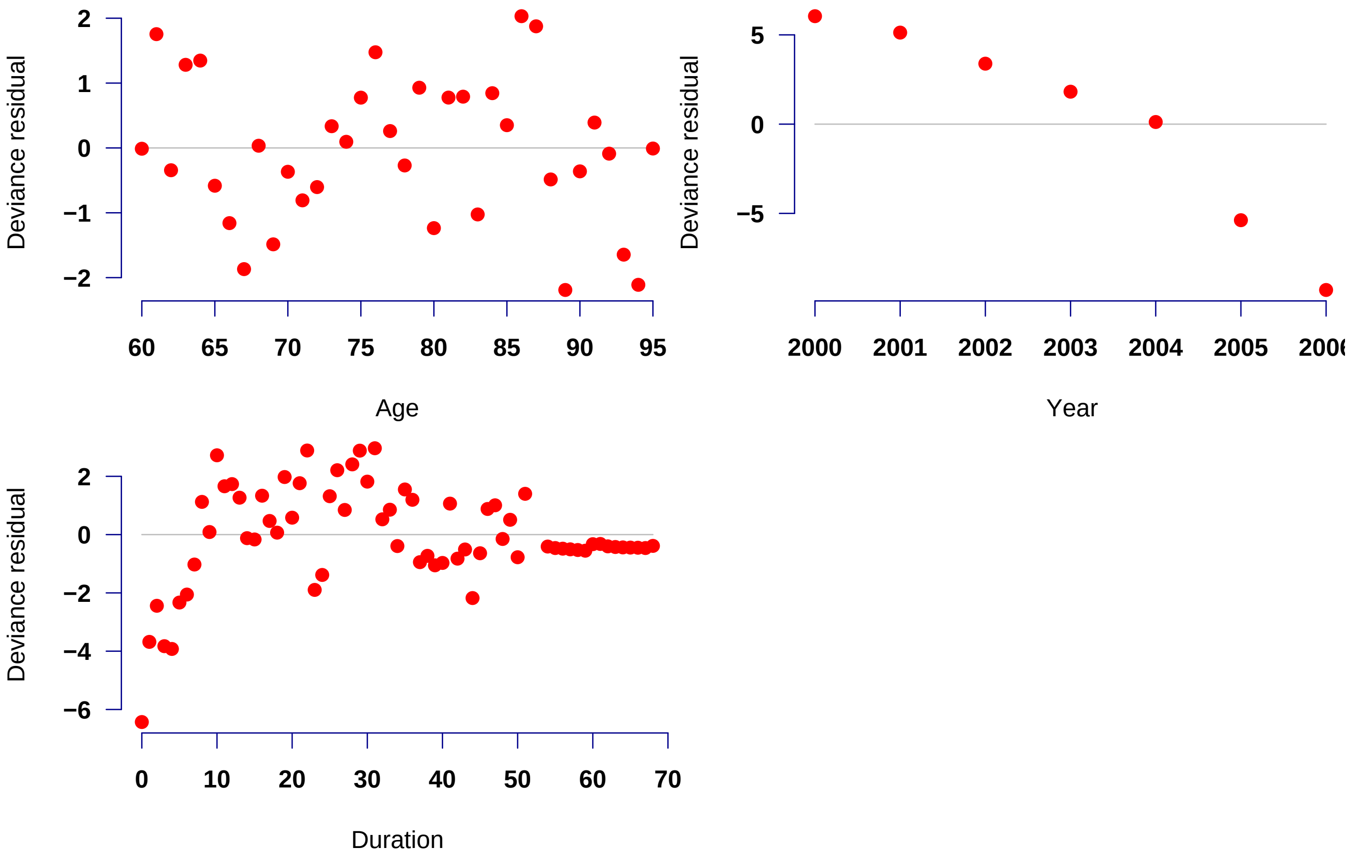

Figure 5. Deviance residuals plotted against (i) age, (ii) calendar time, and (iii) duration since pension commencement. Source: Makeham-Beard model from Table 10 with AIC of 382576. The deviance residuals are calculated assuming a Poisson distribution for the number of deaths within each age range ($x \to x+1$), time period ($y \to y+1$) or duration. The Poisson parameter is the sum of the integrated hazard functions over each sub-range over all individuals who have an exposure in the sub-range.

10.1 It is not enough to select a model by targetting the smallest AIC. One must also check the residuals for evidence of non-random patterns, or for residuals too large to be plausibly normally distributed. Following McCullagh and Nelder (1989) we use deviance residuals, and Figure 5 shows that there are perhaps too many residuals around ± 2 for comfort. There are also possible cyclic patterns with age, which may be caused by the cohort effect — see Richards (2008). A strong pattern by calendar time is also evident, suggesting falling mortality rates over the seven-year period. Finally, the pattern by duration suggests initial selection in the first few years after the pension commences. We can extend the model to cope with all three: cohort effects, time trend and initial selection.

10.2 The cohort effect is a tendency for mortality to be lower for later years of birth, as documented recently by Willets (2004). This is a broadly continuous trend, as shown by Richards, Kirkby and Currie (2006). We can get a quick solution in the model by treating year of birth as an ordinal variable, however, and using the AIC to optimise the breakpoints for at least three (say) broad cohorts. Although this does not acknowledge the continuous nature of the cohort effect, it will suffice for our purposes here to detect the broad pattern.

10.3 We can fit a parameter for time trend by extending the models as follows for the Gompertz law:

$$\mu_{x,y} = e^{\alpha+\beta x+\delta y} \tag{10}$$

where y is the calendar time as a real variable, so we can track the time trend continuously over the period. In practice we need to keep the variables well-scaled for the fitting procedure, so we actually work with $y' = y - 2000$. Note that the time trend observed in Figure 5 is likely to be partly composed of the cohort effect, and partly of a genuine calendar-year effect as mortality improves due to medical treatments and public-health initiatives. However, if we fit a model with a combined cohort factor and a time-trend parameter, then we should be able to broadly separate the two. Such a model is essentially an Age-Period-Cohort (APC) model, a variant of which was used in Richards *et al* (2007) to explore time trends separately from cohort effects in population data.

10.4 Finally, Figure 5 suggests we consider a select period of at least three years, and possibly longer. This is likely to be an ordinary temporary initial selection, but made stronger still by the advent of the enhanced-annuity market: the portfolio here is of standard, non-underwritten annuities and so will contain the presumably healthier lives who did not qualify for enhanced annuities. We can allow for this in the model by adjusting each individual's value for α for the time in years since retirement. For simplicity we will assume that the selection effect is constant within a select period. As before, we will fit two-way interactions between age and the factors for gender, lifestyle and pension size, although we will not fit any other second- or higher-order interactions.

Table 12. Comparison of Makeham-Beard models

Model	**AIC**	**AIC relative to Age model**	**Parameters**
Age	386559	0	4
Age*Gender	384728	-1831	6
Age*(Gender+SizeBand)	383486	-3073	10
Age*(Gender+Lifestyle)	383509	-3050	10
Age*(Gender+SizeBand+Lifestyle)	382576	-3983	14
Age*(Gender+SizeBand+Lifestyle)+Time	382360	-4199	15
Age*(Gender+SizeBand+Lifestyle)+Time+SelectPeriod	382281	-4278	18
Age*(Gender+SizeBand+Lifestyle)+Time+SelectPeriod +Cohort	382254	-4305	20

Source: Own calculations using mortality experience of life-office pensioners aged between 60 and 95 between 2000-2006. SelectPeriod is a three-year select period. Note that the AICs can be compared directly with those in Tables 6, 7, 8, 9 and 10.

10.5 As Table 12 shows, there is a significant time trend, with an AIC difference of 216 between a model containing a time trend and model without it ($= 382576 - 382360$). There is also a significant selection effect, with an AIC difference of 79 between a model containing a select period and a model without it ($= 382360 - 382281$). There is a cohort effect, but it does not appear to be quite as strong as the time-trend and selection effects: the drop in AIC due to a three-band cohort is 27 units ($= 382281 - 382254$).

Table 13. Parameters from Age*(Gender+SizeBand+Lifestyle) + Time + SelectPeriod + Cohort model

Parameter name	Estimate	Std. error	Z value	p-level
Age ($\beta_{baseline}$)	0.10142	0.0025	39.93	0
Beard (ρ)	-12.0213	160.217	-0.08	0.9402
Cohort3.2	0.0986196	0.0226	4.36	0
Cohort3.3	0.0360475	0.0342	1.05	0.2925
Gender.M	2.49759	0.1401	17.83	0
Gender.M:Age	-0.0254885	0.0017	-14.94	0
Intercept ($\alpha_{baseline}$)	-11.6428	0.2257	-51.59	0
Lifestyle.2	-0.553361	0.1629	-3.4	0.0007
Lifestyle.3	-2.19545	0.1545	-14.21	0
Lifestyle.2:Age	0.0105833	0.002	5.21	0
Lifestyle.3:Age	0.0281301	0.0019	14.72	0
Makeham (ϵ)	-6.83223	0.1933	-35.35	0
SelectPeriod.1	0.0872149	0.0488	1.79	0.0736
SelectPeriod.2	0.135504	0.0479	2.83	0.0047
SelectPeriod.3	0.279667	0.0396	7.06	0
Size.2	-1.39607	0.113	-12.35	0
Size.3	-2.67862	0.1722	-15.56	0
Size.2:Age	0.0155757	0.0014	10.96	0
Size.3:Age	0.0291181	0.0021	13.56	0
Time (δ)	-0.0352319	0.0026	-13.4	0

Source: Own calculations using mortality experience of life-office pensioners aged between 60 and 95 between 2000-2006. SelectPeriod is a three-year select period. Parameters are from the final model in Table 12.

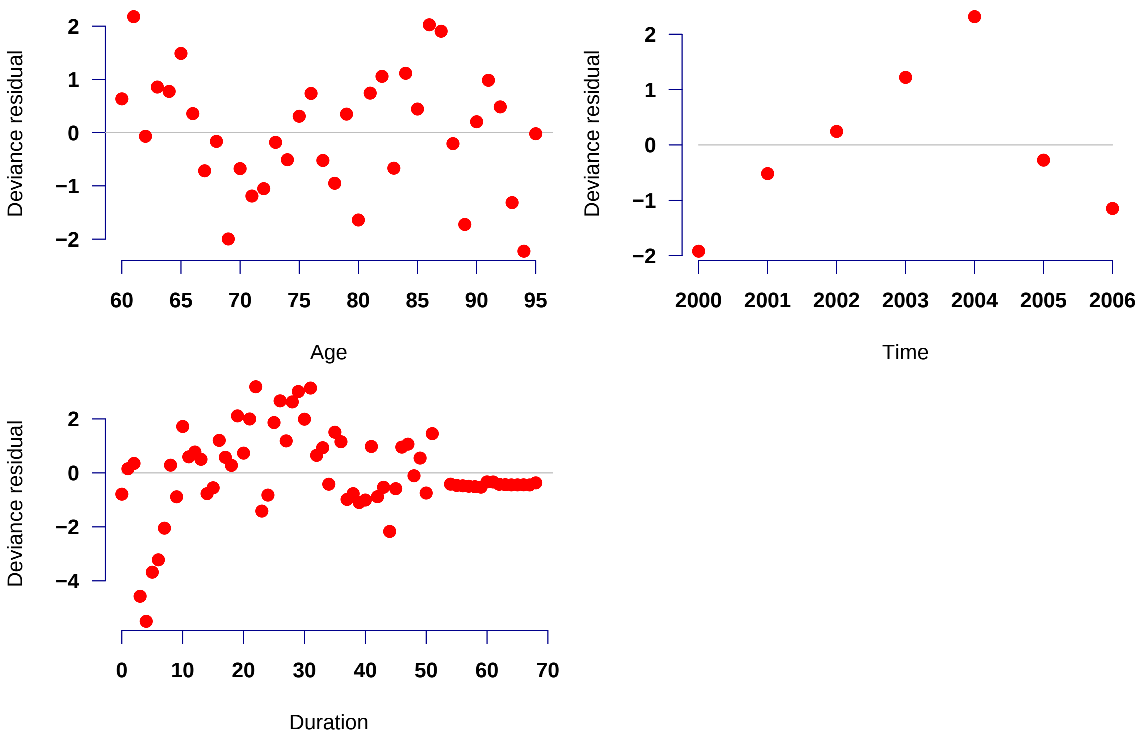

Figure 6. Deviance residuals plotted against (i) age, (ii) calendar time, and (iii) duration since pension commencement. Note that the vertical scale for residuals in (iii) has materially changed from Figure 5.. Source: Final Makeham-Beard model from Table 12 with AIC of 382254.

10.6 There is a substantial improvement in AIC of 322 over the model in Figure 5, but Figure 6 still shows some concerns remain over the non-random patterns of deviance residuals. This could be resolved by making further refinements to the model, or by looking further at the quality of the underlying data. One feature of survival models is that they mercilessly expose any flaws in the data set: this can be frustrating initially, but it eventually leads to better data and better modelling. As demonstrated in Table 11, the financial impact of mortality differentials is too important to gloss over unusual patterns (or, much worse, to not use survival models and never know those patterns were there in the first place).

10.7 One question about the conclusion about the relative strength of the time trend, cohort effect and selection is whether the order of fitting makes a difference. This would change the ascribed drop in the AICs, but the conclusion of order is unaffected. Table 13 shows the Z-values for the fitted parameters, and it appears that the time trend is strongest (Z value of -13.4), followed by the selection effect (Z value of 7.06 comparing initial selection against ultimate rates), then finally the cohort effect (Z value of 4.36 comparing the 1914–1929 generation with the pre-1914 one).

11. Concentration risk

11.1 A major issue in financial work is *concentration risk*, namely the tendency for a given proportion of the portfolio membership to have a much larger proportion of benefits (and therefore liabilities). A statistical model is democratic in that each life has equal weight, whereas both annuity portfolios and pension schemes are very unequal. There are a number of ways in which this concentration of financial risk can be illustrated. One way is to calculate the Gini coefficient, which is used widely in social statistics to measure income inequality. The Gini coefficient takes the value of 0% when everyone has the same income (equality), and 100% when one individual has everything (perfect inequality). The Gini coefficient for the U.K. as a whole was 36.8% in 2005 according to the CIA World Factbook, and we find that most pension schemes and annuity portfolios are generally much more unequal than society as a whole. Another way is to sort the membership by pension size and calculate the proportion of pension benefits paid to each decile (say), or to calculate what proportion of the membership receives half of all the benefits.

Table 14. Concentration of pension benefits by membership decile

Membership decile	**Percentage of portfolio pension: (i) Life office**	**(ii) Pension schemes**
1	54.3%	46.3%
2	15.2%	17.8%
3	9.4%	11.4%
4	6.6%	8.0%
5	4.9%	5.8%
6	3.6%	4.1%
7	2.7%	2.9%
8	1.8%	2.0%
9	1.1%	1.2%
10	0.4%	0.5%
Total	100.0%	100.0%
Gini coefficient	66.0%	60.9%

Source: Own calculations using data for life-office pensioners aged between 60 and 95 between 2000-2006, and pension-scheme members aged between 65 and 100 between 2000-2006. Half of all pensions are paid to 7.8% of policyholders in the life-office data, and 11.7% in the pension-scheme data.

11.2 Table 14 shows that the top decile of membership has around half of all annual pensions paid. We therefore have a group whose financial significance is five times what their headcount would suggest. Equally, the bottom decile of membership has just $\frac{1}{2}\%$ of all pensions paid, so their financial significance is twenty times *less* than what their numbers would suggest. The inequalities would seem even larger if we measured the proportion of liability for each of the membership deciles: in life-office portfolios, for example, where benefit type can be chosen at retirement, wealthier people tend to choose larger average escalation rates and are more likely to buy a surviving spouse's pension. Both of these choices depress the initial pension, and so concentration by reserves is likely to be at least as pronounced as concentration by pension size.

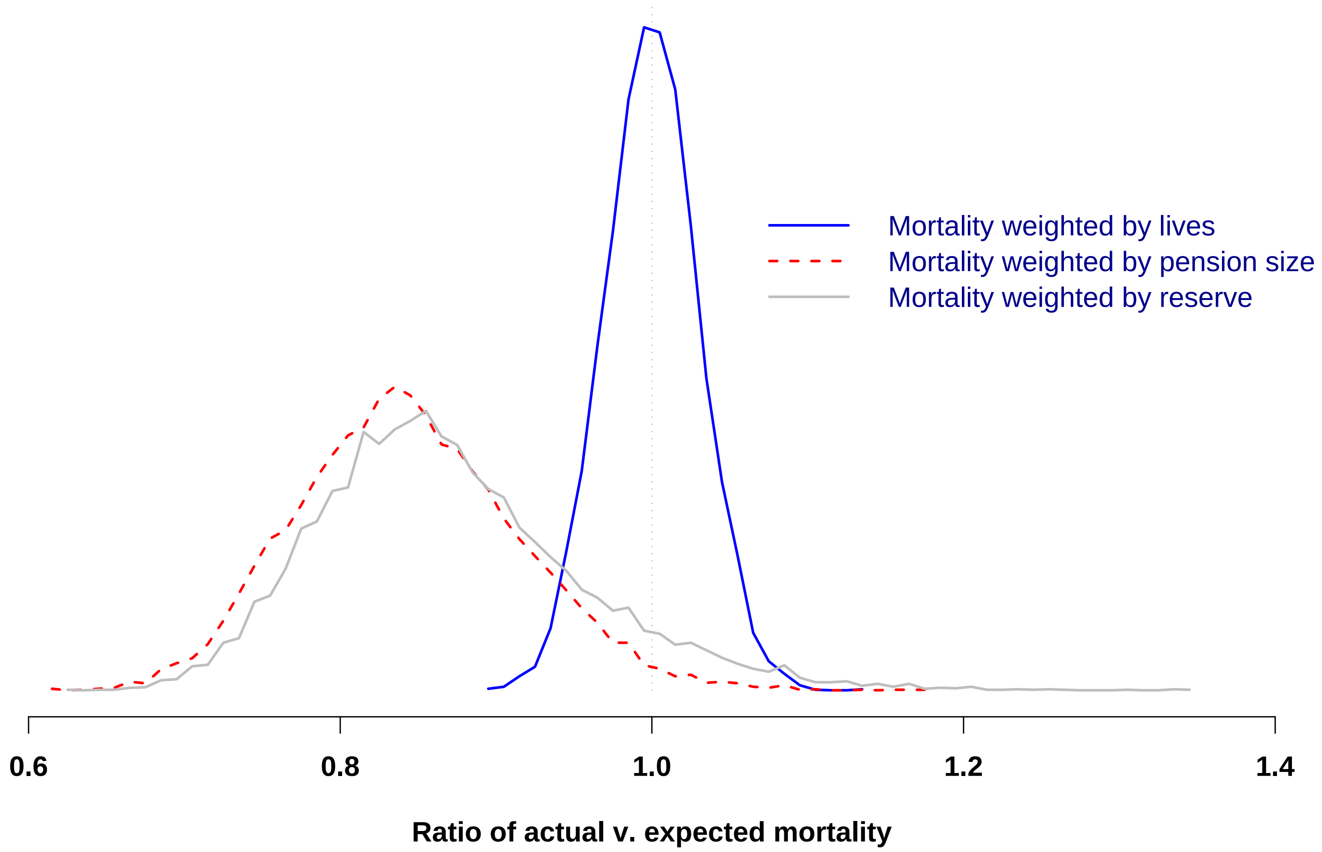

Figure 7. Frequency plot of ratio of actual v. expected mortality for 10,000 bootstrapped portfolios of 50,000 lives (lives-weighted fit). The model is accurate on a lives basis, but clearly overstates mortality weighted by pension size. The additional volatility of amounts-weighted mortality is also clear from the wider horizontal spread. Source: Sampling with replacement from 2006 mortality experience of a portfolio of several hundred thousand life-office pensioners, with expected mortality according to a Makeham-Beard model for age and gender only. The frequencies are scaled so that the area under each curve is 1, i.e. the curves show the empirical density function of the ratio of actual to expected mortality.

11.3 The central issue with this level of concentration risk is that the parameters in a statistical model are primarily driven by the part of the population which accounts for the least part of the financial risk. This might not be a problem if the fitted model is sufficiently rich and the data set is sufficiently large or homogeneous. In practice, however, there is usually a sub-group of very financially significant lives whose mortality is likely to be over-stated due to their relatively small size in the portfolio. A deliberately extreme version of this problem is shown in Figure 7, where an overly simple model based on just age and gender has been fitted and used to predict the mortality for a series of bootstrapped observations from a large portfolio. Sampling with replacement we choose 50,000 lives and calculate the ratio of the actual mortality outcome compared to the model's prediction, i.e. a classic "actual over expected" analysis. Unsurprisingly, the lives-based model is on average a good predictor of the actual outcome, as the ratio is distributed around 1, although there is a lot of variability in a single year's experience for even a portfolio of 50,000 lives. However, as every actuarial student knows, mortality weighted by amounts is a more-useful indicator of financial significance, and mortality weighted by pension size is much lighter than this lives-based model predicts: around 15% lighter on average, but with even more variation in a single year.

11.4 The traditional actuarial approach to mortality analysis is to weight each death or life in the exposed to risk by its pension size and this, too, can be accommodated within a survival-model framework. Equation 5 gives the general formula for the log-likelihood function under a survival model for a single decrement (death), and the contribution of a single life, i, is given by:

$$\ell_i = -H_{x_i}(t_i) + d_i \log \mu_{x_i+t_i} \tag{11}$$

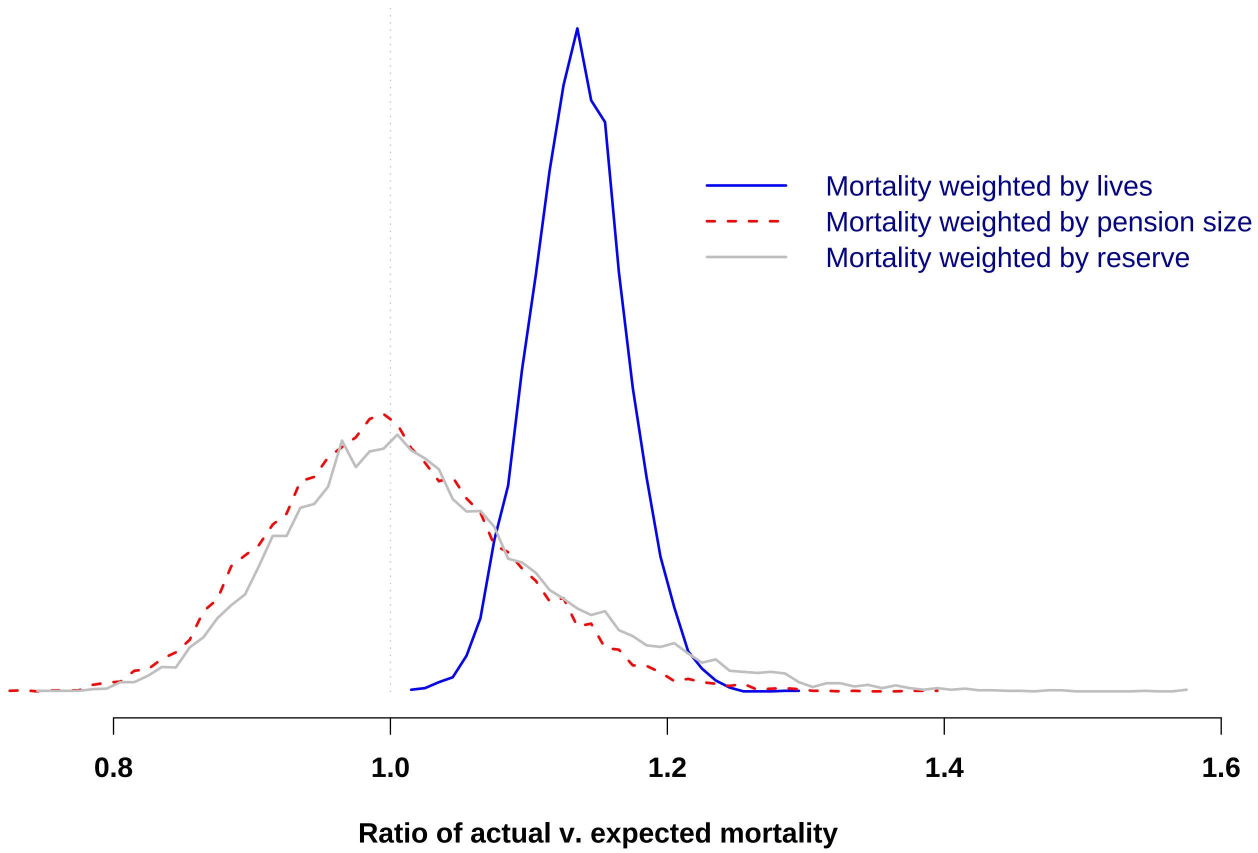

Figure 8. Frequency plot of ratio of actual v. expected mortality for 10,000 bootstrapped portfolios of 50,000 lives (amounts-weighted fit). The basis of calculation is the same as in Figure 7, but the contribution to the log-likelihood is weighted by pension size. The model is now accurate on an amounts basis on average, albeit with quite high variability. The model now understates mortality on a lives basis. Source: Sampling with replacement from 2006 mortality experience of a portfolio of several hundred thousand life-office pensioners, with expected mortality according to a Makeham-Beard model for age and gender only. The frequencies are scaled so that the area under each curve is 1, i.e. the curves show the empirical density function of the ratio of actual to expected mortality.

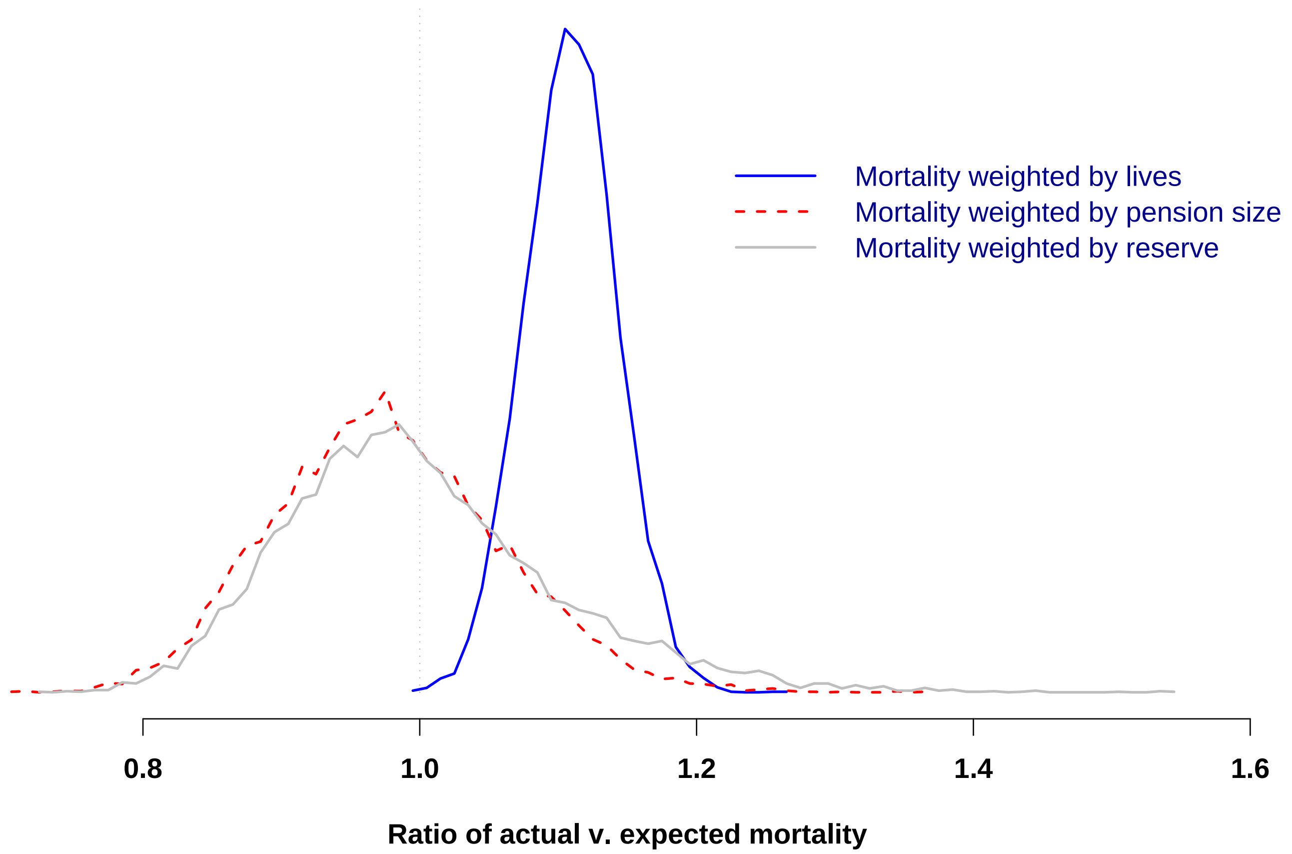

Figure 9. Frequency plot of ratio of actual v. expected mortality for 10,000 bootstrapped portfolios of 50,000 lives (reserve-weighted fit). The basis of calculation is the same as in Figure 7, but the contribution to the log-likelihood is weighted by the estimated reserve, calculated using 5.5% interest and the PCA00 mortality table. The model is now accurate on an reserve basis on average, albeit with quite high variability. As with the amounts-weighted model, mortality is understated on a lives basis. Source: Sampling with replacement from 2006 mortality experience of a portfolio of several hundred thousand life-office pensioners, with expected mortality according to a Makeham-Beard model for age and gender only. The frequencies are scaled so that the area under each curve is 1, i.e. the curves show the empirical density function of the ratio of actual to expected mortality.

11.5 Implicit in Equation 11 is that each life has equal weight. One means of taking financial significance into account is to weight each life's contribution to the log-likelihood. This is an *ad hoc* adjustment to a well-established statistical methodology, but it can be justified if a small number of lives have disproportionate financial significance. If each life has weight w_i, the individual contribution to the weighted log-likelihood, ℓ_i^w, is then:

$$\ell_i^w = -w_i H_{x_i}(t_i) + w_i d_i \log \mu_{x_i+t_i} \tag{12}$$

where w_i could be the reserve, the pension size, or some function of the pension size such as the logarithm or square root. The w_i are scaled so that $\sum_i w_i = n$, which means the weighted log-likelihood can still be broadly compared to the unweighted one. Figure 8 shows how weighting the likelihood in this manner can transform even the simplest model into one which can take better account of financial significance in mortality modelling. However, it is potentially dangerous to rely on amounts-weighting for overly simplistic models, since mortality differentials change with age. It is better to include as many statistically significant risk factors as can be found in a lives-based model, say by reference to the AIC, and thus account for the incidence and timing of such differentials. If found necessary from a bootstrapping check, the final fitted values for financial purposes can be obtained by a refit with some kind of weighting.

11.6 Not all schemes can have their mortality predicted by a model parameterised using the experience from another portfolio, however. Defined-benefit pension schemes often contain individuals concentrated in a particular geographic region and in a particular industry or occupation. If that occupation is particularly hazardous, then a model derived from general pensioner mortality — whether using postcode or pension size — will under-state that scheme's mortality. Such schemes can either have their mortality modelled on a stand-alone basis, or else in a meta-model of multiple schemes where membership of a particular scheme is treated as an additional risk factor.

12. Conclusions

12.1 Life-insurance and pension-scheme data is a form of longitudinal study, and therefore lends itself particularly well to the application of survival models. The assumption of a suitable law for the force of mortality removes the need to separately graduate or smooth the rates.

12.2 Geodemographic models of mortality can fit better than ones purely based on pension size, but a model which combines both will fit better than using either one in isolation. The use of geodemographic profiles in a statistical model also enables the discovery of data issues which remain hidden from the traditional comparisons of actual mortality against a standard table. The mortality differentials identified by such models are highly financially significant and their impact can easily exceed the pricing margin on annuity business at the time of writing.

12.3 Simple process models show how a system composed of non-ageing elements can nevertheless show age-related increases in mortality. These models also yield a slowing down in the rate of increase at advanced ages, known as *late-life mortality deceleration.* This deceleration is observed in human populations, and can be shown to arise from heterogeneity amongst lives. Mortality laws which incorporate an explicit heterogeneity or *frailty* parameter fit better than those which do not.

12.4 The use of bootstrapping can determine if there might be further financially important variation not accounted for in a mortality model, while the use of weights in model-fitting can help limit any mis-statement of financial risk.

Disclosure of interests

The author is a director of Longevitas Ltd and has a financial interest in the Longevitas modelling software used in all the work in this paper. The author also has a financial interest in Longevity Risk Segments, a data product for pension schemes developed jointly with Experian Ltd. The author has a financial interest in mortalityrating.com, a service which rates portfolios of pensions in payment using U.K. postcode.

Acknowledgements

The author thanks Gavin Ritchie for invaluable help in implementing the data validation and deduplication processes. The author also thanks Professor David Wilkie, Professor Steven Haberman, Andrew Howe, Stephen Makin and David McGuinness, together with two anonymous referees, for helpful comments. Any errors or omissions remain the sole responsibility of the author.

Data validation and preparation was done using Longevitas, which was also used to fit all the models. Graphs were done in R and typesetting was done in pdfTEX. The R source for the graphs, together with some C programs for performing calculations in the appendices, can be found at http://www.richardsconsulting.co.uk/laws.html

References

Aitken, M, Anderson, D, Francis, B. and Hinde, J. (1989). Statistical modelling in GLIM, *Oxford University press, 283–285.*

Akaike, H. (1987). Factor analysis and AIC, *Psychometrica,* **52***, 317–333.*

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

Boole, G. and Moulton, J. F. (1960). A Treatise on the Calculus of Finite Differences, *2nd revised edition, New York, Dover, 1960.*

CACI Ltd (2007). , *URL http://www.caci.co.uk.*

Cox, D. R. (1972). Regression models and life tables, *Journal of the Royal Statistical Society, Series B,* **24***, 187–220 (with discussion).*

Cramer, H. (1999). Mathematical Methods of Statistics, *Princeton University Press, ISBN13: 978-0-691-00547-8.*

Currie, I. D., Durban, M. and Eilers, P. H. C. (2004). Smoothing and forecasting mortality rates, *Statistical Modelling,* **4***, 279–298.*

Dickson, D. C. M. and Waters, H. R (2002). The distribution of the time to ruin in the classical risk model, *ASTIN Bulletin, Vol. 32, No. 2, 2002, 299–313.*

EuroDirect Database Marketing Ltd (2007). , *URL http://www.eurodirect.co.uk.*

Experian Ltd (2007). , *URL http://www.experianim.com.*

Gavrilov, L. A. and Gavrilova, N. S. (2001). The Reliability Theory of Aging and Longevity, *Journal of Theoretical Biology (2001)* **213***, 527–545.*

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

Horiuchi, S. and Coale, A. J. (1990). Age patterns of mortality for older women: an analysis using the age-specific rate of mortality change with age, *Mathematical Population Studies,* **2***(4), 245–267.*

Izsak, J. and Gavrilov, L. A. (1995). A typical interdisciplinary topic: questions of the mortality dynamics, *Archives of Gerontology and geriatrics,* **20***(1995), 283–293.*

Larson, H. J. (1982). Introduction to Probability Theory and Statistical Inference, *John Wiley and Sons, Inc, New York, 318–319.*

Legal and General plc (2007). Legal and General links with Hargreaves Lansdown to pioneer postcode-rated annuities, *www.legalandgeneralmediacentre.com.*

Longevitas Development Team (2007). Longevitas v2.2, *Longevitas Ltd, Edinburgh, United Kingdom. URL http://www.longevitas.co.uk.*

Macdonald, A. S. (1996a). An actuarial survey of statistical models for decrement and transition data, I: multiple state, Poisson and Binomial models, *British Actuarial Journal,* **2** *(1) 129–155.*

Macdonald, A. S. (1996b). An actuarial survey of statistical models for decrement and transition data, II: competing risks, non-parametric and regression models, *British Actuarial Journal,* **2***(2) 429–448.*

Macdonald, A. S. (1996c). An actuarial survey of statistical models for decrement and transition data, III: counting process models, *British Actuarial Journal,* **2***(3) 703–726.*

Makeham, W. M. (1859). On the law of mortality and the construction of annuity tables, *Journal of the Institute of Actuaries,* **8***, 301–310.*

McCullagh, P. and Nelder, J. A. (1989). Generalized Linear models, *2nd ed. Chapman and Hall, London.*

McLoone, P. (2000). Carstairs Scores for Scottish Postcode Sectors from the 1991 Census, *Public Health Research Unit, University of Glasgow.*

Park, S. K. and Miller, K. W. (1988). , *Communications of the ACM, Col. 31 1192–1201.*

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

Philips, L. (1990). Hanging on the metaphone, *Computer Language, 1990,* **7** *(12), 39–43.*

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (2002). Numerical Recipes in C++: The Art of Scientific Computing, *Cambridge University Press.*

R Development Core Team (2004). R: a language and environment for statistical computing, *R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.r-project.org.*

Richards, S. J. and Jones, G. L. (2004). Financial aspects of longevity risk, *Staple Inn Actuarial Society, London.*

Richards, S. J., Kirkby, J. G. and Currie, I. D. (2006). The Importance of Year of Birth in Two-Dimensional Mortality Data, *British Actuarial Journal,* **12** *(I), 5–61.*

Richards, S. J., Ellam, J. R., Hubbard, J., Lu, J. L. C, Makin, S. J. and Miller, K. A. (2007). Two-dimensional mortality data: patterns and projections, *British Actuarial Journal (to appear).*

Richards, S. J. (2008). Detecting year-of-birth mortality patterns with limited data, *Journal of the Royal Statistical Society, Series A (2008)* **171***, Part 1, 279–298.*

Strehler, B. L. and Mildvan, A. S. (1960). General Theory of Mortality and Aging, *Science 1 July 1960: 14–21*.

Vaupel, J. W. and yashin, A. I. (1985). Heterogeneity's ruses: some surprising effects of selection on population dynamics, *The American Statistician,* **39***(3), 176–185*.

Willets, R. C. (2004). The Cohort Effect: Insights and Explanations, *British Actuarial Journal, Volume 10, Part IV, No. 48*.

Appendix 1: Derivation of the Beard model

A1.1 Suppose an individual has a Gompertz hazard, i.e. $\mu_x^z = ze^{\beta x}$, where $z = e^{\alpha}$. Suppose further that z is drawn from a gamma distribution at birth, i.e. the density function for z, $f(z)$, is:

$$f(z) = \frac{b^a z^{a-1}}{\Gamma(a)} e^{-zb} \tag{13}$$

where $\Gamma()$ is the gamma function and $a > 0$ and $b > 0$ are the gamma parameters.

A1.2 In general, the hazard rate of the population at age x is as follows:

$$\mu_x = \frac{\int_0^\infty f(z)_x p_0^z \mu_x^z dz}{\int_0^\infty f(z)_x p_0^z dz} \tag{14}$$

A1.3 Taking Equations 2 and 13 together with the appropriate integrated hazard from Table 5, Equation 14 becomes:

$$\begin{aligned}\mu_x &= e^{\beta x} \frac{\int_0^\infty z^{(a+1)-1} \exp\left(-z\left[b + \frac{(e^{\beta x}-1)}{\beta}\right]\right) dz}{\int_0^\infty z^{a-1} \exp\left(-z\left[b + \frac{(e^{\beta x}-1)}{\beta}\right]\right) dz} \\ &= \frac{a\beta e^{\beta x}}{(b\beta - 1) + e^{\beta x}} \\ &= \frac{e^{\alpha' + \beta x}}{1 + e^{\alpha' + \rho' + \beta x}}\end{aligned} \tag{15}$$

where $\rho' = -\log(a\beta)$ and $\alpha' = \log\left(\frac{a\beta}{b\beta - 1}\right)$.

Appendix 2: Derivation of the Makeham-Beard model

A2.1 Following Horiuchi and Coale (1990), suppose an individual has a Makeham hazard, i.e. $\mu_x^z = e^{\epsilon} + ze^{\beta x}$, where z is drawn from a gamma distribution at birth as in Appendix 1. Taking Equations 2 and 13 together with the appropriate integrated hazard from Table 5, Equation 14 becomes:

$$\begin{aligned}\mu_x &= \frac{\int_0^\infty \left(e^{\epsilon} + ze^{\beta x}\right) z^{a-1} \exp\left(-xe^{\epsilon}\right) \exp\left(-z\left[b + \frac{(e^{\beta x}-1)}{\beta}\right]\right) dz}{\int_0^\infty z^{a-1} \exp\left(-xe^{\epsilon}\right) \exp\left(-z\left[b + \frac{(e^{\beta x}-1)}{\beta}\right]\right) dz} \\ &= \frac{e^{\epsilon}(b\beta - 1) + (a\beta + e^{\epsilon}) e^{\beta x}}{(b\beta - 1) + e^{\beta x}} \\ &= \frac{e^{\epsilon} + e^{\alpha' + \beta x}}{1 + e^{\alpha' + \rho' + \beta x}}\end{aligned} \tag{16}$$

where $\rho' = -\log(a\beta + e^{\epsilon})$ and $\alpha' = \log\left(\frac{a\beta + e^{\epsilon}}{b\beta - 1}\right)$. Horiuchi and Coale (1990) further showed that the frailty z remains gamma-distributed for all ages.

Appendix 3: Mortality as a cascade process

A3.1 The following builds on results from Izsak and Gavrilov (1995). Let n denote the integer number of defects of an individual, $n \geq 0$. Denote by t_n the state of an individual having n defects, and let s_n be the number of individuals in a population with n defects. In a small interval of time, t, an individual can either die or accumulate another defect. The force of mortality for an individual in state t_n is $\mu_0 + n\mu$, where $\mu_0 > 0$ is the constant background rate of mortality and $\mu > 0$ is a constant. The instantaneous transition rate for $t_n \to t_{n+1}$ is $\lambda_0 + n\lambda$, where λ_0 and λ are both greater than zero. The instantaneous transition rate is therefore a linearly increasing function of the number of existing defects, which makes this a *cascade process.* These assumptions lead to the linear differential equation:

$$s'_n(x) = \big[\lambda_0 + (n-1)\lambda\big] s_{n-1}(x) - \big[\lambda_0 + \mu_0 + n(\mu + \lambda)\big] s_n(x) \tag{17}$$

A3.2 Solving the initial value $s_0(0) = s_0, s_n(0) = 0$, we get the $s_n(x)$ functions. The number of individuals living at age x is:

$$\sum_{n=0}^{\infty} s_n(x) = \left(\frac{\mu + \lambda}{\mu + \lambda e^{-(\mu+\lambda)x}}\right)^{\frac{\lambda_0}{\lambda}} e^{-(\mu_0+\lambda_0)x} \tag{18}$$

and so the force of mortality at age x, $\mu(x)$, is given by:

$$\mu(x) = \mu_0 + \frac{\mu\lambda_0\left(1 - e^{-(\mu+\lambda)x}\right)}{\mu + \lambda e^{-(\mu+\lambda)x}} \tag{19}$$

A3.3 At this point Izsak and Gavrilov (1995) note that this approximates a Makeham law over a wide age range, with a decelerating increase in the force of mortality at advanced ages and an ultimate value of $\lambda_0 + \mu_0$ for very advanced ages. In fact no approximations are necessary as Equation 19 can be re-written as:

$$\mu(x) = \frac{\left(\mu_0 - \frac{\mu\lambda_0}{\lambda}\right) + \frac{\mu}{\lambda}(\mu_0 + \lambda_0)e^{(\mu+\lambda)x}}{1 + \frac{\mu}{\lambda}e^{(\mu+\lambda)x}} \tag{20}$$

A3.4 If we assume $\mu_0 > \frac{\mu\lambda_0}{\lambda}$, then we can set $\epsilon = \log\left(\mu_0 - \frac{\mu\lambda_0}{\lambda}\right)$, $\beta = \mu + \lambda$, $\alpha = \log\mu + \log(\mu_0 + \lambda_0) - \log\lambda$, and $\rho = -\log(\mu_0 + \lambda_0)$, Equation 20 can then be re-written as:

$$\mu(x) = \frac{e^{\epsilon} + e^{\alpha+\beta x}}{1 + e^{\alpha+\rho+\beta x}} \tag{21}$$

A3.5 Thus, our cascade process for mortality has resulted in a Makeham-Beard force of mortality. This fits neatly with the frailty derivation of the same mortality law, since at each age x there are individuals with varying numbers of defects and thus heterogeneity in the population at a given age. Iszak and Gavrilov (1995) pointed out that letting $x \to \infty$ in Equation 19 yielded a limit to the force of mortality of $\mu_0 + \lambda_0$, and therefore an upper limit to the mortality rate, q_x, which was less than 1.

A3.6 Gavrilov and Gavrilova (2001) developed the idea of an *initial virtual age*, i.e. a life can start out with a non-zero number of defects. This sort of approach explains why the actuarial practice of rating ages up or down according to a standard table works well: if the main difference between two populations is their average initial damage or initial virtual age, then the mortality of one population will be concisely expressed in terms of an age rating against the other.

Appendix 4: Mortality laws and future lifetime as a random variable

A4.1 A model for the force of mortality is equivalent to assuming that the future lifetime is a continuous random variable, T_x, say. The probability density function of T_0, the lifetime from birth, is given by:

$$f(t) = {}_tp_0\mu_t \qquad t > 0 \tag{22}$$

According to Aitken *et al* (1989) the extreme-value distribution for T_0 has probability density:

$$f(t) = \frac{1}{\sigma}\exp\left(\frac{t-\theta}{\sigma} - \exp\left(\frac{t-\theta}{\sigma}\right)\right) \tag{23}$$

and hazard function:

$$\mu_t = \frac{1}{\sigma}\exp\left(\frac{t-\theta}{\sigma}\right) \tag{24}$$

A4.2 Setting $\alpha = -\log\sigma - \dfrac{\theta}{\sigma}$ and $\beta = \dfrac{1}{\sigma}$, Equation 24 can be re-written as:

$$\mu_t = e^{\alpha+\beta t} \tag{25}$$

which we recognise as the Gompertz force of mortality. Equally, assuming a Gompertz force of mortality throughout life is the same as assuming the total future lifetime is a random variable drawn from the extreme-value distribution.

A4.3 Aitken *et al* (1989) also give the probability density for the logistic distribution for T_0 as:

$$f(t) = \frac{\frac{1}{\sigma}\exp\left(\frac{t-\theta}{\sigma}\right)}{\left[1+\exp\left(\frac{t-\theta}{\sigma}\right)\right]^2} \tag{26}$$

which has hazard function:

$$\mu_t = \frac{\frac{1}{\sigma}\exp\left(\frac{t-\theta}{\sigma}\right)}{1+\exp\left(\frac{t-\theta}{\sigma}\right)} \tag{27}$$

A4.4 Setting $\alpha = -\log\sigma - \dfrac{\theta}{\sigma}$, $\beta = \dfrac{1}{\sigma}$ and $\rho = \log\sigma$. Equation 27 can be re-written as:

$$\mu_t = \frac{e^{\alpha+\beta t}}{1+e^{\alpha+\rho+\beta t}} \tag{28}$$

which we recognise as the Beard force of mortality. Equally, assuming a Beard force of mortality throughout life is the same as assuming the total future lifetime is a random variable drawn from the logistic distribution.

Appendix 5: Mortality laws and reliability theory

A5.1 We begin with some results for order statistics. If a random variable X has probability density function f and distribution function F, we can define two new random variables for the maximum and minimum of a set of n independent, identically distributed random variables $\{X_1, X_2, \ldots, X_n\}$:

$$\begin{aligned} X_{max} &= \max\{X_1, X_2, \ldots, X_n\} \\ X_{min} &= \min\{X_1, X_2, \ldots, X_n\} \end{aligned} \tag{29}$$

A5.2 Larson (1982) gives the probability density and distribution functions for X_{max} and X_{min} as follows:

$$\begin{aligned} f_{max}(t) &= n[F(t)]^{n-1} f(t) \\ f_{min}(t) &= n[1 - F(t)]^{n-1} f(t) \\ F_{max}(t) &= [F(t)]^n \\ F_{min}(t) &= 1 - [1 - F(t)]^n \end{aligned} \tag{30}$$

from which we can derive the hazard function, $\mu(t)$, as follows:

$$\mu(t) = \frac{f(t)}{1 - F(t)} \tag{31}$$

A5.3 The remainder of this appendix draws heavily from Gavrilov and Gavrilova (2001). We assume that an *element* has a constant hazard rate of failure, λ, and so the hazard function, $\mu_t^{element}$, is given by:

$$\mu_t^{element} = \lambda \qquad \lambda > 0, t > 0 \tag{32}$$

The time to failure of an element therefore has exponential distribution with probability density function, $f_t^{element}$, and cumulative distribution function, $F_t^{element}$:

$$\begin{aligned} f_t^{element} &= \lambda e^{-\lambda t} \\ F_t^{element} &= 1 - e^{-\lambda t} \end{aligned} \tag{33}$$

A5.4 A *block* is composed of n elements working in parallel, so the failure of all elements is required for the block to stop working. The time to failure of a block is therefore the maximum of failure times of the n elements. Using Equations 30 and 31, the hazard for a block with n working elements is then:

$$\mu_t^{block,n} = \frac{\lambda n e^{-\lambda t} \left(1 - e^{-\lambda t}\right)^{n-1}}{1 - \left(1 - e^{-\lambda t}\right)^n} \qquad n \geq 1 \tag{34}$$

A5.5 Finally, a *system* is composed of m blocks operating serially, i.e. the failure of any one of the blocks results in failure of the system (death). The time to failure of a system is therefore the minimum of the failure times of the m blocks. We add a further detail, namely the probability, p, that any given element is actually working at outset. The resulting hazard for the system as a whole is now given by:

$$\mu_t^{system,m} = mnpc\lambda e^{-np} e^{-\lambda t} \sum_{i=1}^{n} \frac{(np)^{i-1}}{(i-1)!} \cdot \frac{(1 - e^{-\lambda t})^{i-1}}{(1 - (1 - e^{-\lambda t})^i)} \qquad 0 < p \leq 1, m \geq 1, c > 0 \tag{35}$$

A5.6 The normalising constant, c, allows for the fact that if $p < 1$ there is a non-zero probability that all elements of at least one block are non-functioning at outset, and thus that the system is "stillborn". We do not need to know the precise value of c, as it is a constant applies across all times (or ages), and so does not change the basic shape of the force of mortality. In example calculations in this section we set $c = 1$ for simplicity — further details of the derivation of c and the above formulae can be found in Gavrilov and Gavrilova (2001).

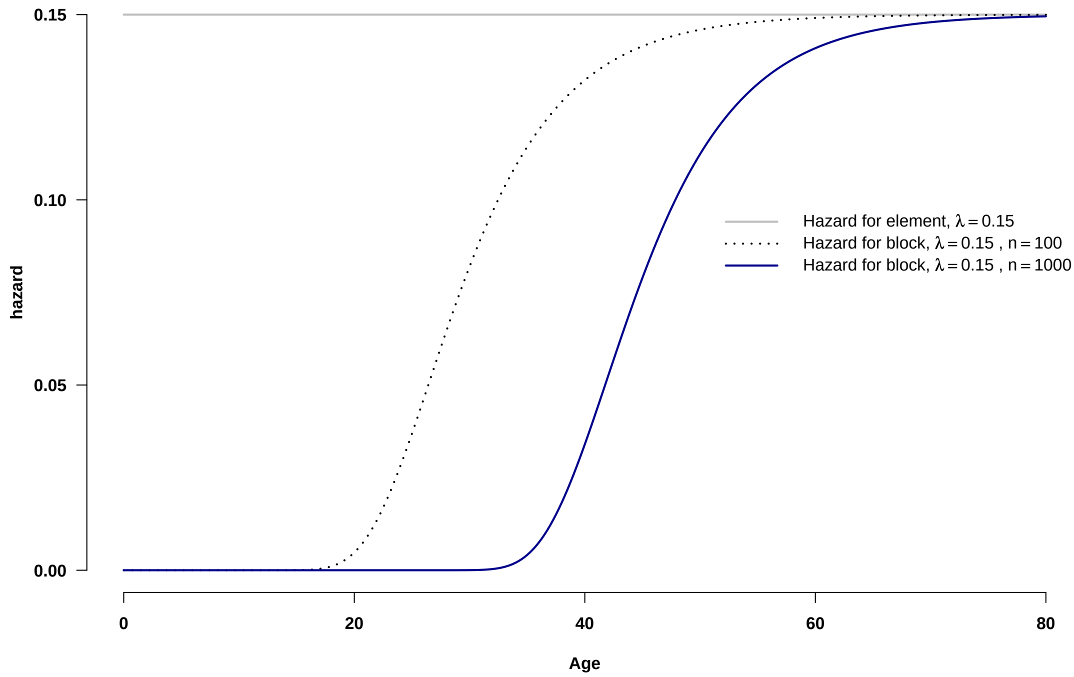

Figure 10. Diagram illustrating hazard functions for elements and blocks. An element has constant hazard, but the hazard functions of blocks exhibit a period of exponential growth in hazard (mortality), followed by a period of deceleration of the rate of increase. Using only the simplest of elements with non-ageing mortality, we have created blocks which exhibit both age-related increase mortality and the later deceleration of those age-related increases.

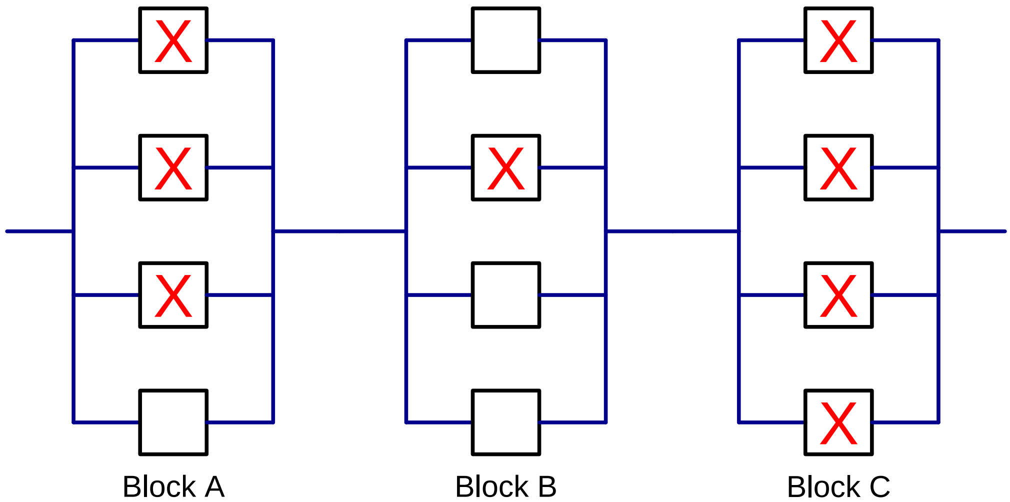

Figure 11. Diagram illustrating a system with redundancy. The system is composed of three blocks linked serially (A, B and C), where the failure of any one of the blocks will result in failure of the system as a whole (death). Each block is composed of four elements which work in parallel, i.e. in order for a block to fail all elements must fail. Failed elements are marked with a cross ($\times$). Despite three failed elements, Block A is still functioning as it has one working element left. Block C has failed completely because all its elements have failed, and the failure of this block results in death of the whole system.

A5.7 An illustration of this system structure is given in Figure 11. The interesting thing about this system is that ageing — i.e. increasing mortality with age or time — has arisen from a simple combination of elements which themselves do not have age-related mortality. Furthermore, this same structure which gives rise to ageing also gives rise to a decelerating rate of increase at advanced ages. Gavrilov and Gavrilova (2001) describe how this type of structure applies to many biological organisms, i.e. self-assembled systems made from small elements with extensive redundancy compensating for some initially non-functioning elements.

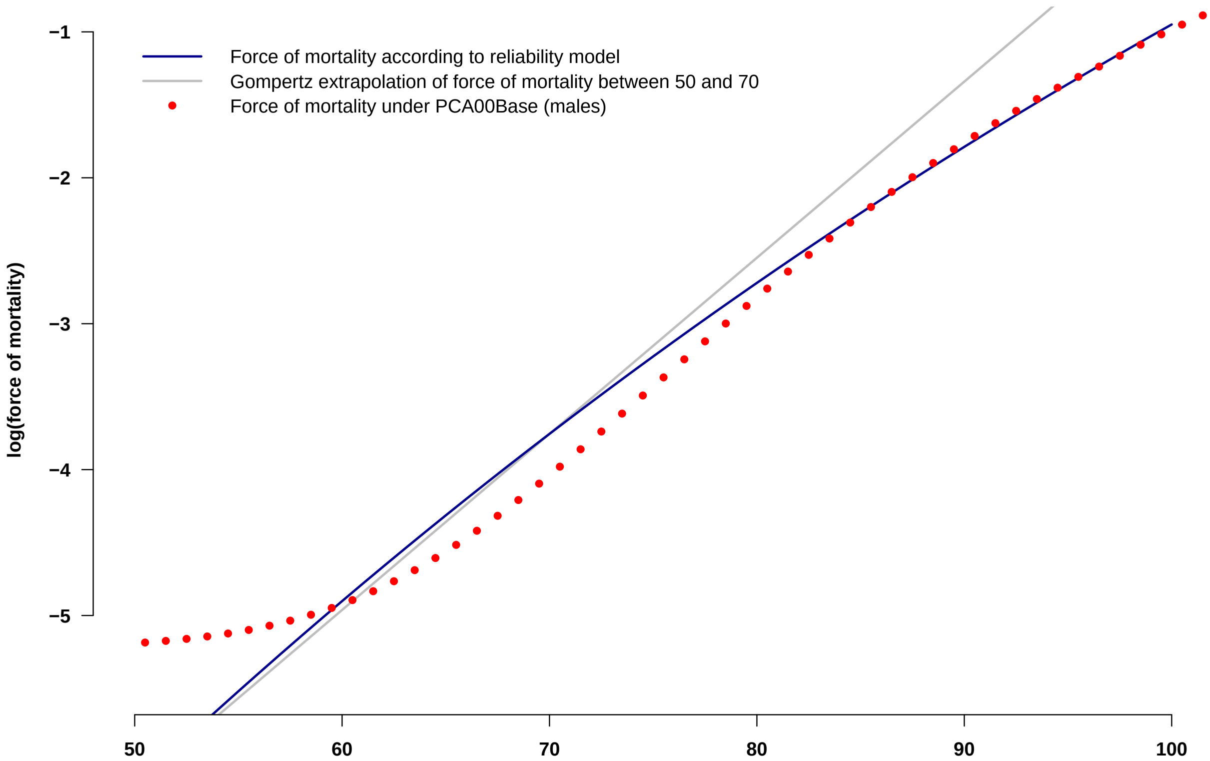

Figure 12. Demonstration of force of mortality under a reliability model. The reliability model has arbitrary specimen parameters $\lambda = 0.01$, $n = 22$, $m = 5850$, and $p = 0.999$, while the Gompertz "equivalent" was obtained by a straight-line extrapolation of the log mortality between ages 50 and 70. The force of mortality under PCA00 (males) has also been plotted to show the same underlying pattern as the reliability model, namely near-linear increase over a large part of the age range with mortality deceleration at higher ages. The parameters for the reliability model were chosen to match PCA00 (males) at age 60 and 100, and some more experimentation is required to get a better match between ages 60 and 80. Nevertheless, it is clear that reliability theory is a plausible basis both for the phenomenon of ageing and late-life mortality deceleration.

Appendix 6: Mortality laws and ruin theory

A6.1 We explore here the idea of a life subject to a damage process following a compound Poisson distribution. This is a continuous-time alternative to the reliability-theory approach adopted by Iszak and Gavrilov (1995). An organism has $x > 0$ functioning elements, but is subject to random damage at rate λ. When an event occurs the extent of the damage is drawn from an exponential distribution with parameter μ. The organism has a repair rate of $r \geq 0$ which it can use to repair existing damage up to the original level of x. This situation is directly analogous to the typical ruin-theory set-up described in Dickson and Waters (2002): x is the surplus level, λ is the claim rate, μ describes the exponential claim size and r is the continuous rate of premium income. The only difference is that we do not allow damage to be repaired above x, i.e. we do not allow growth. This is analogous to a dividend strategy which pays out immediately above a certain capital level.

A6.2 Denote by $\phi(x, t)$ the probability that an organism with x initial functioning elements will survive for time $t \geq 0$. ϕ is therefore a survivor function with $0 \leq \phi \leq 1$. The following boundary conditions are satisfied:

$$\begin{aligned} \phi(0, t) &= 0, \qquad \forall t \geq 0 \\ \phi(x, 0) &= 1, \qquad \forall x > 0 \end{aligned} \tag{36}$$

A6.3 The survival function can be conditioned on whether or not damage occurs in a small interval dt and, if it does, whether the damage kills the organism. Denoting the damage probability density function by f and the distribution function by F, for small interval of time dt the following applies:

$$\begin{aligned} \phi(x, dt) &= (1 - \lambda dt).\phi(x, 0) + \lambda dt F(x) + o(dt) \\ &= (1 - \lambda dt).1 + \lambda dt(1 - e^{-\mu x}) + o(dt) \\ &= 1 - \lambda dt e^{-\mu x} + o(dt) \end{aligned} \tag{37}$$

since with probability $\lambda dt(1 - F(x))$ there is a damage event which is large enough to kill the organism outright. From this we can calculate the force of mortality at time zero with initial value x from:

$$\begin{aligned} \mu(x, 0) &= \lim_{dt \to 0^+} \frac{1 - \phi(x, dt)}{dt} \\ &= \lim_{dt \to 0^+} \frac{\lambda dt e^{-\mu x} + o(dt)}{dt} \\ &= \lambda e^{-\mu x} \end{aligned} \tag{38}$$

A6.4 Equation 38 makes intuitive sense: the force of mortality at time 0 is the event intensity, λ, multiplied by the probability of the damage being enough to kill the organism outright. Extending the derivation of the force of mortality to any future time t is tricky and it is easier to explore $\mu(x, t)$ by Monte-Carlo means. The figure below shows that the choice of some combinations of parameters leads to age-related increases in mortality, as well as a deceleration at the most advanced ages.

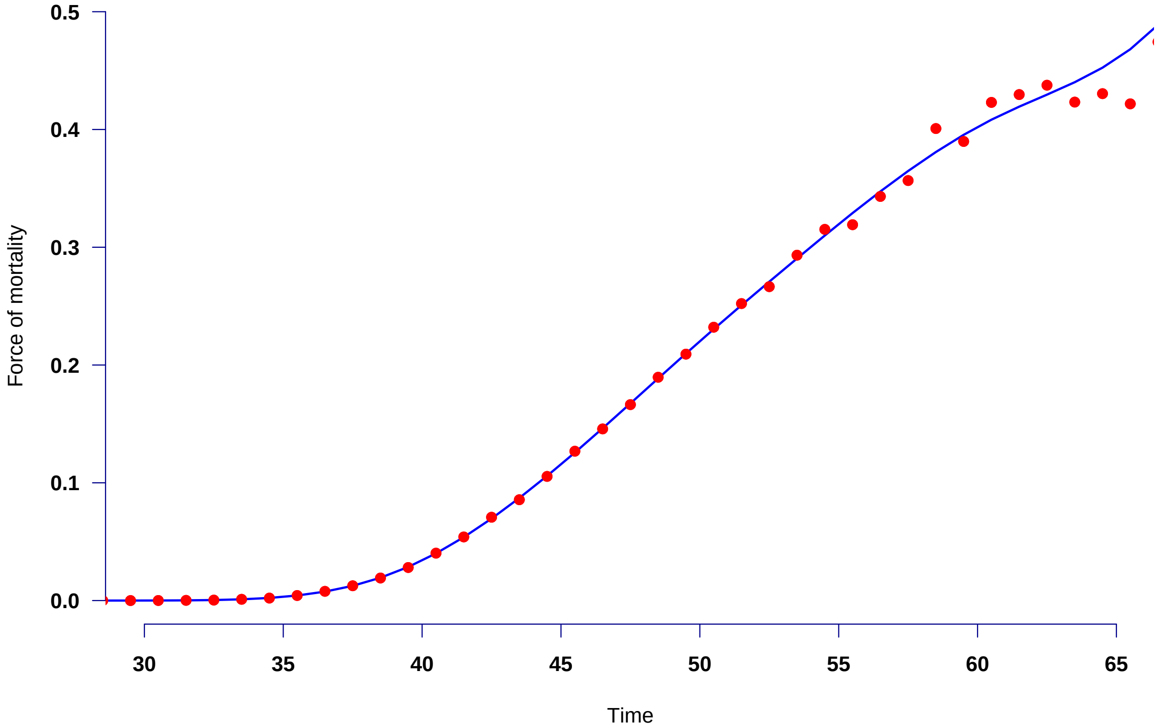

Figure 13. Force of mortality in one million simulations of ruin. Crude forces of mortality (•) and the P-spline fitted values (solid line). The parameters for the simulations were $\lambda = 100$, $\mu = 90$, $r = 0.9$ and $x = 10$. Although the model is very simple, it appears to produce both age-related increases in mortality and an apparent late-life deceleration in mortality increases at the oldest ages. This would need to be confirmed with more analytic work, however, as there is always the risk that the empirical hazard here is shaped by flaws in the random-number generator: we have used the Minimal Standard generator discussed by Park and Miller (1988), with implemementation taken from Press *et al* (2002). The program which generated this data is available for download at http://www.richardsconsulting.co.uk/laws.html