A DISTRIBUTION-BASED STOCHASTIC MODEL OF COHORT LIFE EXPECTANCY

David McCarthy

Imperial College Business School

July 2011

Abstract

Recent mortality models have focused on randomness arising in the year of death as being the primary cause of variation in lifespan. However, there is increasing and puzzling evidence that random effects arising early in life may also have a substantial effect on long-life mortality. This paper presents a new class of mortality models based on this premise, which offer some substantial advantages over existing models, including the existence of easily computable solutions for the distribution of annuity prices and insurance contracts in the presence of stochastic mortality. Rather than estimating the distribution of the probability of death at each age, as in the standard approach, this technique is based on fitting the statistical distribution of the age at death of a cohort, conditional on reaching a certain age. The truncated normal distribution is found to fit well, which facilitates easy computations using the standard statistical toolbox for sampling from normal distributions.

David McCarthy
Imperial College Business School
Exhibition Road
South Kensington
London
SW7 2AZ
+44-207-594-9130
dg.mccarthy@imperial.ac.uk

Acknowledgements: The author would like to thank John Piggott and Filip Sykes for very useful discussions about the paper.

Introduction

There has been substantial recent progress in modelling uncertainty in lifespan. This has been largely driven by the need for financial intermediaries, such as pension funds and annuity providers, to understand the extent to which variation in the mortality underlying their liabilities affects their financial solvency.

The canonical stochastic mortality model, by Lee and Carter (1992), writes central rates of death at each age as an exponential function of a random mortality index, which is assumed to be contemporaneous with the age of death. Most extensions of this model, have followed the same basic approach, although some researchers have incorporated a cohort effect into the Lee-Carter equation.

The practice of modelling age-related mortality probabilities directly, while based on fundamentals, has some disadvantages. In particular, using this class of models produces mortality probabilities which have non-standard distributions, and which are serially correlated. This makes it difficult to compute statistics of direct interest to financial intermediaries, such as the distribution of annuity prices or insurance contracts, without intensive numerical computation. Also, using this class of stochastic mortality model to estimate the distribution of future mortality probabilities, and then pricing annuities or other liabilities off these, ignores idiosyncratic variation in mortality, which may be a significant source of uncertainty for smaller pension funds and insurance companies. Standard stochastic mortality models also make the assumption that all effects governing mortality arise in the year of death. There is increasing evidence that early life experiences have a significant influence on long-run mortality, especially that related to the incidence of chronic diseases. This has been identified in long-run mortality data as a "cohort effect".

This paper presents a different approach to modelling stochastic old-age mortality, based on fitting the distribution of the age of death of a cohort of individuals, conditional on their reaching a certain age. This new approach comes with some advantages. While the probabilities of death at each age are not explicitly modelled and need to be directly computed, semi-closed form solutions of the statistical distribution of quantities of direct interest to financial intermediaries, such as the present value of future mortality-contingent liabilities, can be computed. This allows easy calculation of risk measures related to the portfolio of mortality-related liabilities of financial intermediaries which incorporate both idiosyncratic and systematic mortality risk.

The approach we adopt does have some drawbacks. Firstly, there are difficulties associated with estimation, since the early mortality experience of recent cohorts cannot be included in the calibration reliably since not much is known about them. We use economic and demographic data on each cohort to try and forecast these. Secondly, the model makes the implicit assumption that *all* uncertainty related to lifespan is determined in the year of birth. This is admittedly unreasonable – although possibly no more unreasonable than the opposite assumption generally adopted by the Lee-Carter class of models.

The first section describes the approach. The second section applies it to the United Kingdom, using population data for males and females aged older than 60 from the year 1922 to the present, obtained from the Human Mortality Database, and uses the approach to estimate Value-at-Risk measures for different portfolios of life annuities.

Description of the approach

Standard mortality models, such as those based on Lee and Carter (1992), model directly the probability of death at each age *x* in a given year, *t*.[1] For instance, in the usual formulation of the Lee-Carter model, $m_{x,t}$, the central rate of death of an individual aged *x* in year *t*, which is directly related to $q_{x,t}$, the one-year probability of death, is given by:

$$m_{x,t} = \exp(a_x + b_x k_t),$$

where a_x and b_x are a set of age-specific constants representing the average mortality at each age and the average rate of improvement in mortality relative to a random mortality index, k_t. In the standard formulation, k_t is assumed to follow a random walk with normal innovations. This has several consequences for the resulting distribution of probabilities of death. Firstly, the central rate of death $m_{x,t}$ at each age has a log-normal distribution, but the initial probability of death, $q_{x,t}$, has a non-standard distribution. Secondly, $m_{x,t}$ is perfectly correlated across all ages due to the dependence on the single mortality factor k_t, meaning that the series of cohort-specific survival probabilities $\{{}_s p_{x,t}\}_{s=0,1,\ldots}$, essential for computing annuity prices, has a highly non-standard joint distribution with awkward auto-correlation properties. This makes it impossible to compute the distribution of annuity or insurance prices without using sampling techniques.

[1] In the discussion that follows, we focus on the Lee-Carter model because it is the canonical model of stochastic mortality, but all the issues raised apply equally to all other standard mortality models except where indicated.

There is a second disadvantage to the standard Lee-Carter class of models. All innovations in mortality are implicitly assumed to arise at time *t*, through the operation of the mortality index k_t. This makes the assumption that all uncertainty in lifespan is related to conditions in the calendar year of death. However, as discussed by Willetts (2003), there is increasing awareness that economic and other factors early in life may have substantial effects on long-life mortality.[2]

One way of dealing with this problem is to incorporate cohort effects into Lee-Carter-type models, such as in Haberman and Renshaw (2000), and others. We follow a different approach. By focusing on modelling the distribution of the age of death of a cohort of individuals, rather than on individual mortality probabilities, we obtain a simple, flexible and practical formulation of stochastic mortality which has substantial computational advantages.

Modelling the distribution of the age at death

Rather than modelling fundamentals, we start by modelling the distribution age at death of a cohort of individuals, first unconditionally for ease of explication, and then conditional on reaching a certain age. Our focus here is deaths at older ages; an empirical investigation revealed that the conditional normal distribution gives an excellent fit to the age of death at ages over 60. We should not be too surprised by this: lifespan, like height and other "natural" quantities, is a function of a whole number of random variables, and by the central limit theorem, the aggregate effect on lifespan of all of them, conditional on each factor being sufficiently independent, should result in something close to a normal distribution. While the normal distribution may get probabilities of death in any given year of age wrong, especially for the very old, the financial significance of these deviations for insurance companies and pension funds, as we shall see, is quite small. (We leave the empirical work to the next section).

[2] The relationship between low birthweight and the later development of coronary heart disease is well established (see, for example, Lawlor *et al* (2004)), while Factor-Litvak, P. and Susser, E. (2004), Barker *et al* (1989), Barker (1990, 1995) and Barker and Lackland (2002) show that early-life and even pre-natal experience seem to have a significant explanatory effect on the development and progression of other diseases of old age, such as cardiovascular disease, stroke and chronic bronchitis. Other researchers have explored the relationship between economic variables at birth and long-run mortality. For instance, van den Berg *et al* (2006) show that economic conditions early in life influence long-run life expectancy in a Dutch sample of lives, and van den Berg *et al* (2010) even show that economic conditions early in life seem to affect the recovery of individuals from adverse events – such as the death of a spouse - later in life.

So we model the age at death of a specific cohort, X, as having a normal distribution[3] with the following probability distribution function (PDF), conditional on the underlying parameters μ and σ^2:

$$f(X \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}(X-\mu)^2). \quad (1)$$

We then introduce further uncertainty into the distribution by allowing μ and σ^2 to have their own joint distribution. While we have a wide choice of possible distributions for these quantities, we focus on two choices.

Normal-scaled inverse-gamma (NSIG) distribution

Bayesian theory suggests that a convenient choice for the distribution of μ and σ^2 is the conjugate prior of the normal distribution with unknown mean and variance, which is a four-parameter distribution called the normal-scaled inverse-gamma (NSIG) distribution with the following PDF:

$$f(\mu, \sigma^2 \mid \lambda, \nu, \alpha, \beta) = \frac{\sqrt{\nu}}{\sigma\sqrt{2\pi}} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp(-\frac{1}{2\sigma^2}(2\beta + \nu(\mu-\lambda)^2)). \quad (2)$$

Here, the marginal distribution of the parameter μ follows a normal distribution, and that of the parameter σ^2 is an inverse-gamma distribution. The appendix gives the first two central moments of μ and σ^2 in terms of the meta-parameters λ, ν, α and β. Unfortunately, while analytically convenient, this distributional choice imposes the restriction that the covariance between μ and σ^2 is zero, which, as we shall see, is violated empirically.

However, using this distribution, the distribution of the age at death X, conditional on the underlying meta-parameters can be calculated analytically as:

$$f(X \mid \lambda, \nu, \alpha, \beta) = \iint_{\mu,\sigma^2} f(X \mid \mu, \sigma^2) f(\mu, \sigma^2 \mid \lambda, \nu, \alpha, \beta) d\mu d\sigma^2 \quad (3)$$

$$= \frac{\sqrt{\nu}}{\sqrt{2\pi}\sqrt{1+\nu}} \frac{\beta^{-\frac{1}{2}}\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha)} (1 + \frac{\nu(X-\lambda)^2}{2\beta(1+\nu)})^{-\alpha-1/2}. \quad (4)$$

[3] We ignore conditioning on survival to a particular age in this section for ease of explication. The extreme negative tail of the normal distribution will be removed by this.

This is recognisable as the PDF of a scaled and shifted Student's-t distribution, so we can write

$$X = \sqrt{\frac{\beta(1+\nu)}{\alpha\nu}}Y + \lambda\,, \tag{5}$$

where Y has a Student's t distribution with 2α degrees of freedom.

This means that we can calculate the mean and variance of X incorporating uncertainty in the underlying parameters very easily in terms of the meta-parameters as:

$$E(X \mid \lambda,\nu,\alpha,\beta) = E(\sqrt{\frac{\beta(1+\nu)}{\alpha\nu}}y + \lambda) = \lambda\,, \text{ and} \tag{6}$$

$$Var(X \mid \lambda,\nu,\alpha,\beta) = Var(\sqrt{\frac{\beta(1+\nu)}{\alpha\nu}}Y + \lambda) = \frac{\beta(1+\nu)}{\alpha\nu}Var(Y) = \frac{\beta(1+\nu)}{(\alpha-1)\nu}. \tag{7}$$

Using results in the appendix, we can express the central moments of the unconditional age at death X in terms of the moments of the underlying meta-parameters, and obtain that

$$E(X \mid \lambda,\nu,\alpha,\beta) = E(\mu)\,, \text{ and} \tag{8}$$

$$Var(X \mid \lambda,\nu,\alpha,\beta) = \text{var}(\mu) + \text{var}(\sigma) + E(\sigma)^2\,. \tag{9}$$

The addition of parameter uncertainty increases the variance of the age at death both through uncertainty in the mean age at death μ and the uncertainty in the range of variation of the age at death, σ.

Provided that the normal distribution (or the conditional normal distribution) fits the age at death of a cohort of individuals reasonably well (we shall see that it does), and that the underlying parameters follow an NSIG distribution or something reasonably close - these results give an easy way of calculating the percentiles of the distribution of the age at death of an individual incorporating mortality uncertainty. If α is small, we can use (4) and the numerical functions of the Student's t-distribution implemented in most statistical packages, such as Excel, and if α is large, it is even easier because we can use (8) and (9) and the fact that for large α, the Student's t distribution is very close to a normal distribution.

Bivariate normal distribution

One disadvantage of using the NISG distribution is the fact that the correlation between μ and σ is constrained to be zero. In empirical work, we shall see that the estimated variance-covariance matrix implies a correlation of around 60% between the estimates of these two parameters, at least for UK data.

We deal with this problem by also considering a bivariate normal distribution for the mean and standard deviation. This is a five-parameter distribution, with the fifth parameter allowing for the non-zero covariance between the mean and standard deviation. In this case, there are no convenient empirical results; the distribution

$$f(X \mid \lambda, \nu, \alpha, \beta, \gamma) = \iint_{\mu,\sigma} f(X \mid \mu, \sigma) f(\mu, \sigma \mid \lambda, \nu, \alpha, \beta, \gamma) d\mu d\sigma$$

must be solved numerically. We use the following definitions for the parameters:

$$E(\mu) = \lambda,\ E(\sigma) = \nu,\ Var(\mu) = \alpha,\ Var(\sigma) = \beta, \text{ and } Cov(\mu, \sigma) = \gamma.$$

Besides analytical intractability, another disadvantage of using the bivariate normal distribution is that it implies that there is a non-zero probability that $\sigma < 0$, although in practice this probability is extremely small and can be disregarded for our parameterisations.[4]

Modelling the distribution of single annuities and assurances

Once we have obtained the distribution of the age at death of a cohort of individuals using the results in the previous section, it is a relatively easy exercise to obtain the distribution of the present value of individual annuities and assurances. This is the great advantage of this method of modelling stochastic mortality over alternative methods.

An assurance which pays 1 on the death of an individual, which occurs at age *X*, has a present value equal to

$$L = \exp(-rX), \tag{10}$$

[4] The more theoretically pleasing – but more complex - alternative of modelling μ and σ^2 as (X, Y^2) where (X, Y) is a bivariate normal distribution brings few practical advantages.

if we assume a constant interest rate *r* and ignore conditioning on survival (which will come later). If *X* has the distribution in (4), then *L* has a log Student's t distribution, and for large α, a log-normal distribution.

Similarly, the present value at a constant rate of interest *r* of an annuity paid at a constant annual rate 1 from birth (again, ignoring conditioning), is given by

$$A = (1 - \exp(-rX)) / r . \qquad (11)$$

Once again, this has either a scaled log-normal or log-Student's t distribution, depending on the value of α.

In the case of a bivariate normal distribution for (μ, σ), the theoretical results are no longer so elegant, but we can easily derive the distributions for *A* and *L* numerically.

In either case, we can then use the derived distribution to calculate confidence intervals for the value of a single annuity or assurance directly, although deriving confidence intervals for the range of future mortality probabilities directly is difficult.[5]

Modelling the distribution of a portfolio of annuities and assurances

Insurance companies and pension funds typically have portfolios of large numbers of annuities or assurances. The standard approach using stochastic mortality models of the Lee-Carter type is to derive distributions of future mortality probabilities, and then to take the present value of portfolios of annuities off the percentile values of this distribution in order to obtain the percentile values of the portfolio. This implicitly makes the assumption that there is no idiosyncratic risk in the portfolio – so that the portfolio is effectively infinitely sized.

In our approach, this technique does not work naturally, because we do not directly derive distributions of future mortality probabilities, although our approach gives an analytical expression in semi-closed form for the distribution of the value of a portfolio of annuities or assurances which incorporates both idiosyncratic and systematic longevity risk.

[5] This is because, in our formulation, there are two parameters which control the shape of future mortality probabilities. There will be a range of possible combinations of these parameters which will generate any particular annuity value – including those on the boundaries of a confidence region.

However, modelling the distribution of the value of a portfolio of annuities or assurances requires some care. The reason for this is that the ages of death of each individual in a cohort are not independent because of the characteristics shared by the mortality of the cohort. We allow for this dependence by calculating the distribution of the portfolio value conditional on the value of μ and σ, and then by allowing for uncertainty in those values to reflect differences in cohort mortality.

So imagine, in the simplest case, that we have a portfolio of *n* assurances, indexed from *1* to *n*, which pay 1 on the death of an individual *i*. Once again ignoring conditioning on age in the explication, by (10) and (1), we know that

$$L_i \mid \mu, \sigma \sim LN(-r\mu, r^2\sigma^2). \tag{12}$$

Conditional on the values of μ and σ^2, these distributions are independent by assumption. Therefore, provided that we have a sufficiently large number of these, and writing

$$L = \sum_{i=1}^{n} L_i, \tag{13}$$

we can apply the standard version of the central limit theorem for uncorrelated random variables with finite standard deviation and have

$$L \mid \mu, \sigma^2 \dot{\sim} N(n\exp(-r\mu + \tfrac{1}{2}r^2\sigma^2), n(\exp(-2r\mu + r^2\sigma^2)(\exp(r^2\sigma^2) - 1))). \tag{14}$$

Then, the distribution of $L = \sum_{i=1}^{n} L_i$ conditional on λ, ν, α and β is given by the following integral if (μ, σ^2) follows a NSIG distribution:

$$f(L \mid \lambda, \nu, \alpha, \beta) = \iint_{\mu,\sigma^2} f(L \mid \mu, \sigma^2) f(\mu, \sigma^2 \mid \lambda, \nu, \alpha, \beta) d\mu d\sigma^2, \tag{15}$$

and by this integral if (μ, σ) follows a bivariate normal distribution:

$$f(L \mid \lambda, \nu, \alpha, \beta, \gamma) = \iint_{\mu,\sigma} f(L \mid \mu, \sigma) f(\mu, \sigma \mid \lambda, \nu, \alpha, \beta, \gamma) d\mu d\sigma. \tag{16}$$

The present value of a portfolio of annuities on the same set of lives is given by:

$$A = (n - L)/r. \tag{17}$$

While calculating the integrals in (15) and (16) is not trivial, they are straightforward to implement numerically. Confidence intervals or percentile values can then be read off the resulting distributions in the standard way. Conditioning on survival until a certain age simply changes the mean and variance of each individual annuity or assurance, but will not alter the underlying normality of the distribution of the sum of these liabilities, provided there is a sufficient number of them. (In the case of conditional distributions, more than about 50 annuities will be required to ensure that the portfolio value is well modelled by a normal distribution).

In the next section we illustrate this approach by applying it to the historical time series of the mortality of the male population of the United Kingdom.

Applying the model to UK cohort mortality

In this section, we apply the results of the preceding sections to the historical mortality of the United Kingdom. We obtained a set of age and time-specific population and death data for the UK male and female population older than 60 from the year 1922 until 2009 from the Human Mortality Database, maintained by the Department of Demography at the University of California at Berkeley. Our data has the male population and the female population of the United Kingdom on the 1st January each year, arranged by single year of age last birthday, and the number of deaths of individuals in each calendar year, again arranged by single year of age last birthday. We re-arranged the data into cohorts, starting with the cohort born in 1862, and ending with the cohort born in 1949. We assumed that births were uniformly distributed over the year, so calculated the number of deaths related to the population as the average of the number of deaths in that calendar year of people aged x and $x+1$.

Ideally, we would like to fit the model using a procedure which uses the mortality experience of adjacent cohorts to determine the correct parameters for each cohort. However, in order to understand the properties of each cohort's mortality, we first fit a distribution to each cohort independently, and then fit the entire sample jointly using maximum likelihood to estimate and project the trend in life expectancy.

Estimating the model: independent cohorts

When fitting each cohort independently, we need enough data on each cohort to fit reliable estimates. So, somewhat arbitrarily, we exclude all cohorts for whom more than

half of the cohort was still alive in 2009, which leaves all cohorts up to the cohort of 1929 in the sample.

For each of these cohorts, we fitted a left-truncated normal distribution to the observed age of death, conditional on survival to age 60. To handle entrants and exits from each cohort (caused by immigration or emigration and our smoothing procedure), we constructed a series of age and cohort-specific mortality probabilities and used these to construct survival curves.

Rather than following a maximum-likelihood approach, we chose that combination of parameters which minimised the absolute value of the largest deviation between the two cumulative probability distributions (which is the Kolmogorov-Smirnoff test statistic for the difference between two statistical distributions). Figure 1 shows the fitted and empirical distributions for selected cohorts. Unfortunately, the Kolmogorov-Smirnoff test itself is not valid, because we have estimated the parameters of the fitted distribution using the underlying data. However, in Figure 2, we show the K-S test statistic for each cohort up to the cohort of 1929 for males and females, and inspection reveals that the fit is extremely good in all cases, with the CDF's never differing by more than 3% in the entire sample. The normal distribution appears to fit male mortality much better than female mortality.

We note that using the normal distribution results in implied annual probabilities of death in individual years, especially at very old ages, which may be quite different from their empirical counterparts. However, the financial significance of these poor estimates is low, because very few people reach the ages in question. In fact, the difference in life annuity prices between using the fitted normal distribution for the age of death and the true empirical distribution of the age of death, calculated at an interest rate of 3% p.a., is always less than 2% of the annuity price for the completed cohorts 1862-1909, and usually considerably less than that (the average absolute difference over all cohorts is around 0.75% of the true annuity price). The results are shown in Figure 2. Once again, the errors from using the normal distribution appear to be much smaller for males than for females.

Figure 3 shows each parameter plotted across time, with the standard deviation on the left-hand axis and the mean on the right-hand axis. (Values for male cohorts are shown in solid lines; female cohorts are indicated by dotted lines). The values for σ are quite similar for male and female cohorts. In both cases, the value of σ is roughly constant until 1900 (although for males, there appears to be a slight positive gradient over this

period), while after 1900, it begins to increase, first quite slowly, and then more rapidly. The values of μ differ quite substantially between males and females. As expected, μ is much higher for females than for males, reflecting their longer life expectancy. But the pattern over time is also different. For females, μ increases in a roughly linear way over the entire sample period, while for males it is initially constant, and only begins to increase with the cohort born in 1900.

Projecting cohort life expectancy using maximum likelihood

The previous estimates show the trend in the parameters of the underlying distributions of age at death for male and female cohorts over time. Using these parameters, it would be possible to fit a bivariate random walk to the parameters and then use this to project them forward. However, this approach ignores the information contained in adjacent cohorts, which is especially useful for fitting to model later cohorts for which little data is available.

So we use a joint maximum likelihood procedure to fit the model to the entire dataset, including the most recent cohorts. This should allow a more accurate joint estimate of the trends in the mean and the variance of the age at death, and their joint asymptotic distribution.

For each cohort, we observe the dates of death of the entire cohort, starting at age 60. If we had no immigration or emigration, we could treat each cohort as a closed group and therefore have $N_{60,t}$ observations from a discretised truncated normal distribution, where *t* is an index for each cohort. For each cohort t, we would observe $D_{x,t}$ deaths at age *x*. We then write out the likelihood function of each cohort's mortality experience as one draw from a multinomial distribution.

Conditional on a series of means, and standard deviations,

$$\Omega = \left\{ \begin{matrix} \mu_t \\ \sigma_t \end{matrix} \right\}_{t=1862...1929}, \tag{18}$$

the log-likelihood (after removing a constant term) of observing a particular draw $\{D_{x,t}\}_{x=1862...1929,t=1922...2009}$ is then given by:

$$\ell(\{D_{x,t}, N_{x,t}\} \mid \Omega) = 1_{x+t=2009}(N_{x,t} - D_{x,t}) \log(1 - \sum_{x} q_{x,t}) + \sum_{x,t} D_{x,t} \log(q_{x,t}), \tag{19}$$

where

$$q_{x,t} = \frac{\Phi(x+1, \mu_t, \sigma_t) - \Phi(x, \mu_t, \sigma_t)}{1 - \Phi(60, \mu_t, \sigma_t)}. \tag{20}$$

The first term in log-likelihood function adjusts for the fact that for later cohorts there are still individuals in the sample who are alive at the end of the observation period.

However, in our data there appears to be a small amount of movement in and out of each cohort, presumably caused by immigration and emigration, and by small errors in the smoothing procedure used to generate our underlying deaths data. We can therefore treat the number of deaths at each age as an independent draw from a Poisson distribution, so

$$D_{x,t} \,|\, \{\mu_t, \sigma_t\} \sim Pois(N_{x,t} q_{x,t}), \tag{21}$$

$$\text{with } q_{x,t} = \frac{\Phi(x+1, \mu_t, \sigma_t) - \Phi(x, \mu_t, \sigma_t)}{1 - \Phi(x, \mu_t, \sigma_t)}. \tag{22}$$

(Strictly speaking, the number of deaths in each cohort at each age is distributed with a Binomial distribution, but this is well approximated with a Poisson distribution). This log-likelihood function, again adjusting for a constant term, is given by:

$$\ell(\{D_{x,t}, N_{x,t}\} \,|\, \Omega) = \sum_{x,t} D_{x,t} \log(N_{x,t} q_{x,t}) - N_{x,t} q_{x,t}. \tag{23}$$

Ideally, we would like to impose a distributional assumption on the values of μ and σ - such as a bivariate random walk – and then estimate the parameters of this process in a one-step procedure from the underlying data. While it is a simple matter to write out the likelihood function for such a procedure, its evaluation requires the repeated evaluation of a 2*n*-dimensional integral, where *n* is the number of cohorts in the data sample, which in our case is not computationally feasible, as discussed in the appendix.

We therefore adopt a less general, but implementable, strategy. We ignore the random element in the series Ω and fit a parameterised linear function, so we set:

$$\Omega = \left\{ \begin{matrix} \mu_0 + a(t - YEAR) \\ \sigma_0 + b(t - YEAR) \end{matrix} \right\}_{t=YEAR...1949}, \tag{24}$$

where YEAR is the first year in the estimation dataset. Then, defining $q_{x,t}$ as in (22), since

$$D_{x,t} \mid \{\mu_0, \sigma_0, a, b\} \sim Pois(N_{x,t} q_{x,t}), \tag{25}$$

with each $\{D_{x,t}\}_{x=YEAR..1949, t=1922...2009}$ independent, the log-likelihood function of observing the sample $\{D_{x,t}\}_{x=YEAR...1949, t=1922...2009}$ conditional on $\{\mu_0, \sigma_0, a, b\}$ is given by

$$\ell(\{D_{x,t}\} \mid \{\mu_0, \sigma_0, a, b\}) = \sum_{x,t} D_{x,t} \log(N_{x,t} q_{x,t}) - N_{x,t} q_{x,t} . \tag{26}$$

Alternatively, ignoring movements into and out of the cohort, the multinomial log-likelihood would be given by:

$$\ell(\{D_{x,t}\} \mid \{\mu_0, \sigma_0, a, b\}) = 1_{x+t=2009}(N_{x,t} - D_{x,t}) \log(1 - \sum_{x} q_{x,t}) + \sum_{x,t} D_{x,t} \log(q_{x,t}), \tag{27}$$

where $q_{x,t}$ is defined as in (20).

In either case, we can maximise the likelihood *w.r.t.* the four parameters $\{\mu_0, \sigma_0, a, b\}$ and then estimate their joint asymptotic distribution using standard results from maximum likelihood theory. This gives us the variance of the trend in life expectancy, which we can use to calculate the distribution of future annuity prices incorporating error into our estimates of the drift.

Table 1 reports maximum likelihood estimates using the Poisson and the multinomial likelihood functions, based on various sub-samples, for male cohorts, and Table 2 for female cohorts. Figures 5 and 6 plot the resulting implied values for μ and σ up to the cohort of 1950.

Depending on the underlying dataset, the implied drift in the mean varies from 0.1 to 0.3, reflecting the rapid acceleration in the improvement in the life expectancy of recent cohorts. Despite the fact that the drift in female life expectancy is much more constant than for males, there is still substantial variation in the estimates for females than for males.

Improving the projection by conditioning on demographic variables

We therefore investigate whether conditioning on other known information makes our estimates more stable or not. We introduce a vector *Z* of cohort-specific information known to affect mortality, and write:

$$\Omega = \left\{ \begin{matrix} \mu_0 + a(x - YEAR) + \beta' Z \\ \sigma_0 + b(x - YEAR) + \gamma' Z \end{matrix} \right\}_{x=YEAR...1949} \tag{28}$$

In choosing variables to include in *Z* we need to take care to include cohort-specific variables rather than variables which apply to the whole population at a particular time point. We also wish to avoid simultaneity bias, and so only use variables which can be observed before the cohort reaches its sixtieth birthday. For estimation purposes, we would also prefer variables which are stationary, since this will avoid the problem of spurious regression coefficients caused by trends in the underlying variables.

We collected information on the proportion of 60-year old males and females who smoked in each year from 1928-2009 from Forey *et al* (2009).[6] In including this variable, we are implicitly making the assumption that smoking behaviour from age 60 onwards is similar in all of the cohorts. Most cohorts do appear to show a decline in smoking prevalence with age, at least partly the result of selection, as smokers tend to die earlier, and presumably partly explained by individuals giving up smoking as they age.

Our next variable aims to capture some of the lifetime economic and lifestyle-related characteristics of the cohort. We chose to use data on the proportion of each cohort that was in social classes I and II on a 7-point Goldthorpe occupation-based social class scale after the age of 35 from Heath and Payne (1999).[7] While according to the Goldthorpe scale, classes I, II and III represent non-manual workers, we chose to include only classes I and II since this maximised the difference between the experience of men and women in the available data, important for estimation purposes, as will be explained later.

Finally, to condition on early life experience, we also collected information on per-capita GDP growth in the year before the cohort was born from Maddison (2006). We included

---

[6] We chose the narrowest data interval nearest the age of 60 in each survey year, and used linear interpolation to impute missing data. In every case, we used surveys which recorded the proportion of people who reported using any tobacco product, and included occasional smokers (as opposed to limiting our counting to regular smokers).

[7] The Goldthorpe scale allocates individuals to social classes based on their occupation. Heath and Payne (1999) report social class separately for males and females for all individuals in a cohort above the age of 35, therefore giving a reasonable estimate of the professional experience of each cohort over their adult lives. Results are reported by decade of birth year. We interpolated between these decadal values using linear interpolation to get estimates for individual cohorts.

this variable since van den Berg *et al* (2006) show that economic conditions early in life appear to exert some influence on late-life mortality and morbidity.

Our data series are shown for males and females as Figure 7.[8] As can be seen in the figure, smoking and social class data are strongly trended. To minimise the difficulty of spurious regression estimates caused by regressing on trended variables, we fit a model to male and female cohort life expectancy jointly for these variables. We assume that males and females have separate values of μ_0, σ_0, a and b, but joint values of γ and β in equation (28). Since GDP growth is not trended, we permitted males and females to have separate coefficients for this variable. Because there appeared to be little difference between the Poisson and multinomial model estimates in Table 2, we fit only the multinomial model in this section. Results are shown in Table 3 for different sub-periods.

We focus first on the estimates using data from 1900 onwards. The smoking variable is highly economically significant, reducing lifespan by approximately 5 years for both males and females. Social class, on the other hand, appears to increase lifespan dramatically for both males and females, although since there is not much variation between males and females in the social class variable, this estimate in particular should be treated with caution. An extra percentage point change in GDP in the year before birth increases lifespan by around 5 days for men, and around half that for women, so this variable, while statistically significant, does not appear to be economically important.

The results for the sample with data from 1915 onwards illustrate well the difficulties of regressions on trended data. The coefficient estimates are highly unstable, with some reversals of sign and large changes in magnitude for the time trend and demographic estimates. The reason is clear once the data in Figure 6 is re-examined. In the period after 1915, there is very little difference in changes in smoking patterns or social class make-up between men and women. In both sexes, smoking patterns decline dramatically, and the proportion of the population in the top two social classes increases in a roughly linear way. Our strategy of using variation between men and women to identify these effects will therefore not succeed in these data, and we must unfortunately treat the estimates with caution.

In Figures 7 and 8 we report the fitted values for μ and σ for men and women implied by the two sets of regression estimates, along with the estimated values for μ and

[8] The data as well as an MS Excel routine to calculate the likelihood function are available at www.dgmccarthy.com.

σ obtained by fitting each cohort independently. The values are remarkably close in-sample, especially for women, indicating that deviations from a linear trend in mortality improvement for men and women are quite well modelled by the demographic variables we have chosen. However, while for women the projected values for μ and σ are quite similar for the two sub-samples, for men they differ quite substantially, illustrating the problems of out-of-sample projections of mortality improvements. Part of the difficulty here may well be the problems we have in estimating suitable values for trended variables, as discussed. In what follows, we use the values based on the sample estimated using data from 1900.

Estimating the distribution of the present value of a portfolio of annuities

We now use the model to estimate the distribution of the present value of a portfolio of annuities. We make the assumption that any factors affecting mortality not included in the model – so anything other than a basic time trend and the effects of smoking and social class – can be treated as a random walk without drift. We therefore calculate the difference between the model-implied values of μ and σ using the estimates from Table 3 based on the data sample from 1900-1949, and the values of μ and σ fitted to each individual cohort from 1900-1929. We then fit a bivariate random walk with zero drift to these 30 residuals. Somewhat arbitrarily, we treat our starting point as the individual estimates for the cohort of 1929, and then use the implied distributions of μ and σ with a mean value equal to the value implied by our model estimates in the first two columns of Table 3 and with a variance-covariance matrix given by the random walk. Resulting meta-parameter values are shown in Table 4.

We then use these meta-parameters to obtain estimates of the 99th percentile of the distribution of the present value of various portfolios of life with and without mortality uncertainty. The table shows the expected value of the present value of one annuity in the first line, and then the difference between the 99th percentile of the distribution and the mean value, expressed as a proportion of the present value of the portfolio. Results are shown in Table 5 for different portfolio sizes, ranging from 100 to 100,000 annuities to illustrate the effects of the diversification of idiosyncratic risk on the distribution.

In the case of fixed life expectancy, the 99th percentile of the distribution of the present value is approximately 9.8% above the mean for a portfolio of 100 annuities written on males born in 1940; this falls slightly for males born in 1950 and 1960. As the portfolio size increases by a factor of 10, the difference between the 99th percentile and the mean

falls by a factor of the sqrt of 10, so by the time the portfolio is 100,000 annuities, the 99th percentile is 0.4% above the mean.

In the case of random life expectancy, though, the 99th percentile is larger than in the case of fixed life expectancy – 11.9% above the mean value for males born in 1940, increasing to 13.4% above the mean value for males born in 1960, but it does not fall as rapidly as the sample size increases. This is caused by the correlation between the life expectancy of the members of the cohort. By the time there are 100,000 annuities in the portfolio, the 99th percentile is 6.7% of the mean value, increasing to 9.3% of the value for cohorts born in 1960. For females, the values with mortality uncertainty are somewhat smaller, reflecting the more accurate in-sample projections based on our demographic data.

Approximation

Although the results in the previous section require only the evaluation of a single integral, they are quite computationally intensive because the probability density function function needs to be integrated over its range to obtain the distribution function necessary to calculate percentile values.

However, since the distribution of the present value of the liabilities, conditional on the values of μ and σ, is approximately normal, if we ignore any correlation between the mean and the variance of the distribution, we can use the NSIG distribution to obtain an easy approximation to the distribution of the portfolio. We numerically calculate the expected value and variance of the present value and variance of a single annuity by integrating over the underlying distribution of μ and σ. Call these four values $E[\mu^*]$, $Var[\mu^*]$, $E[\sigma^*]$ and $Var[\sigma^*]$. If we assume that μ^* and $\sigma^{*\,2}$ follow an NSIG distribution, then we then know from (8) and (9) that the present value of a portfolio of *n* of these annuities will follow an approximate normal distribution (actually, a Student's t distribution) with mean

$$M = nE(\mu^*), \tag{29}$$

and variance

$$\Sigma = n^2 \operatorname{var}(\mu^*) + n \operatorname{var}(\sigma^*) + nE(\sigma^*)^2. \tag{30}$$

This formula allows for the correlation between the present value of the different annuities caused by the shared draw of meta-parameters, but incorporates the fact that

they are independent conditional on a realisation of the meta-parameters. The formula is an approximation, though, because it ignores the correlation between M and Σ caused by the fact that both are based on a realisation of μ and σ, which are themselves correlated, and because μ^* and σ^{*2} will not follow an exact NSIG distribution if μ and σ are jointly normal.

Table 6 contains estimates of the first two moments of μ^* and σ^* for one annuity in each cohort for males and females and the resulting estimates of M and Σ, as well as the percentage difference between the 99th percentile values calculated using (29) and (30) and the accurate values in Table 5. In no case does the approximation differ by more than 0.4%, and it is usually much less than this.

This suggests a method by which the reserve requirements of insurance companies and annuity providers can be accurately calculated. Firstly, write a routine to calculate the mean and standard deviation of the mean and standard deviation of the present value of an annuity or insurance portfolio. This routine can easily incorporate annuities in the same portfolio which have different sizes and which are written on lives in different cohorts. Then, use (29) and (30) to calculate the mean and variance of the present value of the distribution, and use simple normal tables to calculate the desired percentile values.

Conclusion

In this paper we have illustrated the potential of a type of mortality model based on fitting the distribution of future life conditional on reaching a particular age. We have shown that the conditional normal distribution fits the distribution of future lifespan well in historical data, and can easily be used to price annuities with only small errors. We have developed a methodology for projecting the parameters of the distribution based on demographic data. The model fits UK population mortality well in historical data for both males and females, but some work remains to obtain reliable out-of-sample estimates, particularly for males.

We have shown the potential of the model in estimating annuity prices and the distribution of their present value for future cohorts. Although accurate results are computationally intensive, we have developed an approximation based on Bayesian theory which works very well in certain circumstances. The model may provide an alternative pricing mechanism for insurers and annuity providers who wish to check the

projection methodologies currently in use, and it provides an easy and practical way in which insurers and annuity providers can calculate the Value-at-Risk of their annuity portfolios.

References

Barker, D.J.P., Osmond, , C. & Law, C.M. (1989). "The intrauterine and early postnatal origins of cardiovascular disease and chronic bronchitis". *Journal of Epidemiology and Community Health*, 43, 237-40.

Barker, D.J.P. (1990). "The fetal and infant origins of adult disease." B.M.J. 301, 1111.

Barker, D.J.P. (1995). "The fetal origins of coronary heart disease." B.M.J. 311, 171-4.

Barker, D.J.P. & Lackland, D.T. (2002). "Prenatal influences on stroke mortality in England & Wales." *Stroke*, 34(7), 1602-3.

Cairns, Andrew, David Blake and Kevin Dowd, (2006), "A Two-Factor Model for Stochastic Mortality: Theory and Calibration.", *Journal of Risk and Insurance,* 73, pp. 687-718.

Factor-Litvak, P. and Susser, E. (2004). "A life course approach to neuropsychiatric outcomes", in (D. Kuh and Y. Ben-Shlomo, eds), *A Life Course Approach to Chronic Disease Epidemiology*, 2nd edition, pp. 324–43, Oxford: Oxford University Press

Forey, Barbara, Jan Hamling, John Hamling and Peter Lee, (2002), *International Smoking Statistics*, Chapter 27, 2nd edition, Wolfson Institute of Preventive Medicine and Oxford University Press, Oxford, UK.

Friedman, Howard S. and Leslie R. Martin, (2011), *The Longevity Project*, Hudson Street Press, New York, NY.

Haberman, Stephen and Renshaw A.E., "Lee-Carter mortality forecasting with age specific enhancement", *Insurance: Mathematics and Economics*, 33(2), pp. 255 - 272

Heath, Anthony and Clive Payne, (1999), "Twentieth Century Trend in Social Mobility in Britain", WP 70, June 1999, Centre for Research into Elections and Social Trends.

Lawlor, D.A., Ben-Shlomo, Y. and Leon, D.A. (2004). "Pre-adult influences on cardiovascular disease", in (D. Kuh and Y. Ben-Shlomo, eds), *A Life Course Approach to Chronic Disease Epidemiology*, pp. 41–76, Oxford: Oxford University Press.

Lee, Ronald D. and Lawrence R. Carter, (1992), "Modelling and Forecasting U.S. Mortality.", *Journal of the American Statistical Association*, 87: 659–671.

Maddison, Angus, (2006), *The World Economy, Volume 1: A Millenial Perspective and Volume 2: Historical Statistics*, OECD Publishing, Paris France.

Van den Berg, Gerhard, Maarten Lindeboom and France Portrait (2006) "Economic Conditions Early in Life and Individual Mortality", *American Economic Review,* Vol 96, No 1, pp. 260-302.

Willetts, Richard, (2003), *The cohort effect: insights and explanations*, Richard Willetts Consulting, http://www.willets.co.uk/downloads/The%20cohort%20effect.pdf.

Tables

Start date		1862	1862	1900	1900	1915	1915
		Poisson	Multinom	Poisson	Multinom	Poisson	Multinom
mu		70.629	70.857	71.818	71.890	73.629	73.757
		(8.60E-04)	(8.13E-04)	(1.31E-03)	(1.26E-03)	(2.22E-03)	(2.15E-03)
sig		9.500	9.376	9.990	9.912	11.309	11.254
		(6.74E-04)	(6.24E-04)	(1.06E-03)	(9.95E-04)	(1.95E-03)	(1.86E-03)
a		0.096	0.093	0.216	0.214	0.327	0.319
		(1.72E-05)	(1.65E-05)	(5.60E-05)	(5.46E-05)	(1.51E-04)	(1.48E-04)
b		0.027	0.028	0.095	0.097	0.180	0.182
		(1.55E-05)	(1.46E-05)	(6.12E-05)	(5.83E-05)	(2.04E-04)	(1.98E-04)
LL		-233953	-256400	-59786	-64094	-16935	-18057
Implied 1950	m	79.118	79.024	82.637	82.603	85.064	84.919
	s	11.833	11.822	14.760	14.768	17.626	17.618

Table 1: Estimated parameters, fitted by maximum likelihood, to male cohorts from start date to 1949. Second panel shows model-implied estimates for mu and sigma for cohort of 1950.

Start date		1862	1862	1900	1900	1915	1915
		Poisson	Multinom	Poisson	Multinom	Poisson	Multinom
mu		74.742	74.953	78.679	78.694	79.948	80.007
		(5.91E-04)	(5.66E-04)	(9.42E-04)	(4.98E-05)	(1.70E-03)	(1.65E-03)
sig		9.238	9.140	9.720	9.636	11.057	10.976
		(4.85E-04)	(4.54E-04)	(8.00E-04)	(5.17E-05)	(1.67E-03)	(1.61E-03)
a		0.118	0.114	0.163	0.163	0.302	0.299
		(1.34E-05)	(1.29E-05)	(5.11E-05)	(0.00E+00)	(1.67E-04)	(1.64E-04)
b		0.036	0.036	0.109	0.110	0.223	0.226
		(1.20E-05)	(1.14E-05)	(5.41E-05)	(0.00E+00)	(2.13E-04)	(2.08E-04)
LL		-296448	-328600	-140588	-153167	-34426	-36098
Implied 1950	m	85.130	85.006	86.840	86.830	90.513	90.456
	s	12.401	12.308	15.151	15.112	18.848	18.875

Table 2: Estimated parameters, fitted by maximum likelihood, to female cohorts from start date to 1949. Second panel shows model-implied estimates for mu and sigma for cohort of 1950.

Start year		1900		1915	
		M	F	M	F
Mu(0)		69.47	75.80	66.60	75.20
Sigma(0)		16.98	13.93	17.12	14.76
A		0.01	0.12	0.29	0.29
B		0.00	0.11	0.26	0.28
Beta (GDP)		1.25	0.778	1.340	2.081
Gamma (GDP)		-0.96	-0.70	0.07	2.27
Beta(SM)			-4.63		4.62
Gamma(SM)			-9.98		-2.25
Beta(SC)			34.11		18.32
Gamma(SC)			1.11		-19.41
LL		-194211.95		-51753.5	
Implied 1950	Mu	82.69	91.29	85.40	92.20
	Sigma	15.16	17.35	17.72	17.97

Table 3: Estimated parameters, fitted by maximum likelihood, to male and female cohorts from start date to 1949. Second panel shows model-implied estimates for mu and sigma for cohort of 1950. All variables are highly significant so standard errors are not reported.

	Males				
Cohort	lambda	nu	alpha	beta	gamma
1940	80.52	14.07	0.66	2.32	0.46
1950	82.69	15.16	1.25	4.43	0.88
1960	82.86	15.19	1.85	6.54	1.29

	Females				
Cohort	lambda	nu	alpha	beta	gamma
1940	87.24	15.71	0.44	1.71	0.10
1950	91.29	17.35	0.84	3.26	0.20
1960	92.44	18.33	1.25	4.82	0.29

Table 4: Projected meta-parameter values for future cohorts, with statistical distributions, multivariable normal model. Based on estimates for sample from 1900 onwards.

Males				
		1940	1950	1960
Random LE	100	0.119	0.124	0.134
	1000	0.074	0.084	0.097
	10000	0.067	0.078	0.093
	100000	0.067	0.077	0.093
Fixed LE	100	0.098	0.094	0.094
	1000	0.032	0.032	0.032
	10000	0.012	0.012	0.012
	100000	0.004	0.004	0.004

Females				
		1940	1950	1960
Random LE	100	0.089	0.085	0.087
	1000	0.042	0.044	0.048
	10000	0.033	0.037	0.042
	100000	0.033	0.036	0.041
Fixed LE	100	0.084	0.08	0.08
	1000	0.028	0.026	0.026
	10000	0.01	0.01	0.01
	100000	0.004	0.004	0.004

Table 5: 99th Percentile of annuity distribution for a portfolio of annuities, depending on the size of the portfolio and the year of birth of the cohort, trend estimates based on sample 1900-1949. Values expressed as the difference between the 99th percentile and the mean of the distribution, expressed as a proportion of the mean.

Males

		1940	1950	1960
	$E(\mu^*)$	15.270	16.249	16.327
	$\text{var}(\mu^*)$	0.449	0.561	0.674
	$E(\sigma^*)$	6.263	6.408	6.383
	$\text{var}(\sigma^*)$	0.330	0.436	0.543
M	n=100	1527	1625	1633
	1000	15270	16249	16327
	10000	152701	162493	163271
	100000	1527009	1624935	1632712
Σ	n=100	77	85	93
	1000	490	597	704
	10000	4529	5650	6770
	100000	44895	56173	67422
Δ	n=100	-0.001	-0.001	-0.001
	1000	0.001	0.002	0.003
	10000	0.002	0.003	0.004
	100000	0.002	0.003	0.003

Females

		1940	1950	1960
	$E(\mu^*)$	17.941	19.405	19.811
	$\text{var}(\mu^*)$	0.258	0.313	0.366
	$E(\sigma^*)$	6.335	6.406	6.508
	$\text{var}(\sigma^*)$	0.290	0.384	0.442
M	n=100	1794	1941	1981
	1000	17941	19405	19811
	10000	179409	194053	198110
	100000	1794092	1940533	1981099
Σ	n=100	68	71	75
	1000	327	373	420
	10000	2655	3196	3714
	100000	25860	31374	36623
Δ	n=100	0.000	0.001	0.001
	1000	0.001	0.001	0.001
	10000	0.001	0.001	0.002
	100000	0.001	0.001	0.002

Table 6: Values of $E(\mu^*)$, $\text{var}(\mu^*)$, $E(\sigma^*)$ and $\text{var}(\sigma^*)$ for a single annuity on a life in each cohort, calculated numerically using results from Table 5. Corresponding values of M and Σ are shown as well as the difference between the 99.5th percentile values calculated using the approximation and the true values shown in Table 6.

Figures

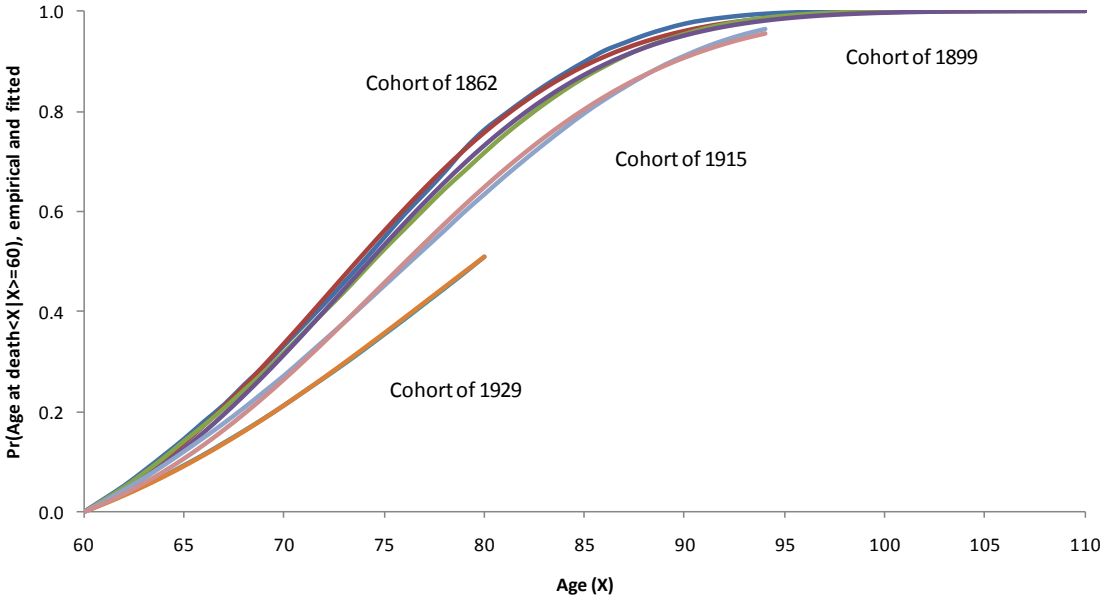

Figure 1: Fitted and empirical distributions of age at death for male cohorts, conditional on surviving to age 60.

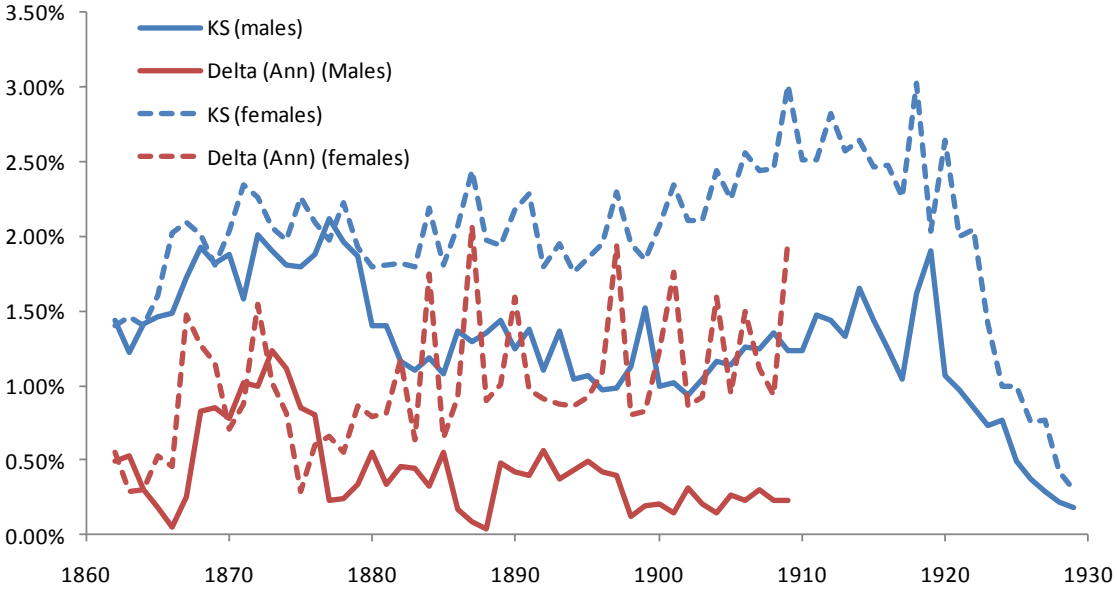

Figure 2: Kolmogorov-Smirnoff test statistic and error between true and approximate annuity prices, males and females.

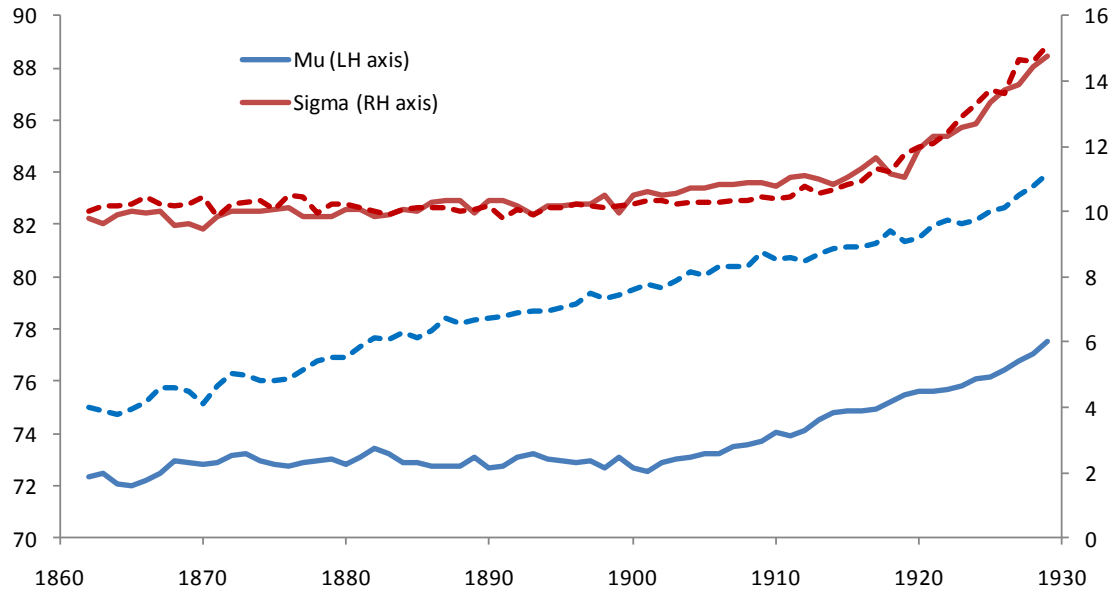

Figure 3: Fitted estimates for mu (blue line, right-hand axis) and sigma (red line, left-hand axis), male cohorts 1862-1929. Fitted using Kolmogorov-Smirnoff procedure, each cohort independent. Dashed lines indicate females.

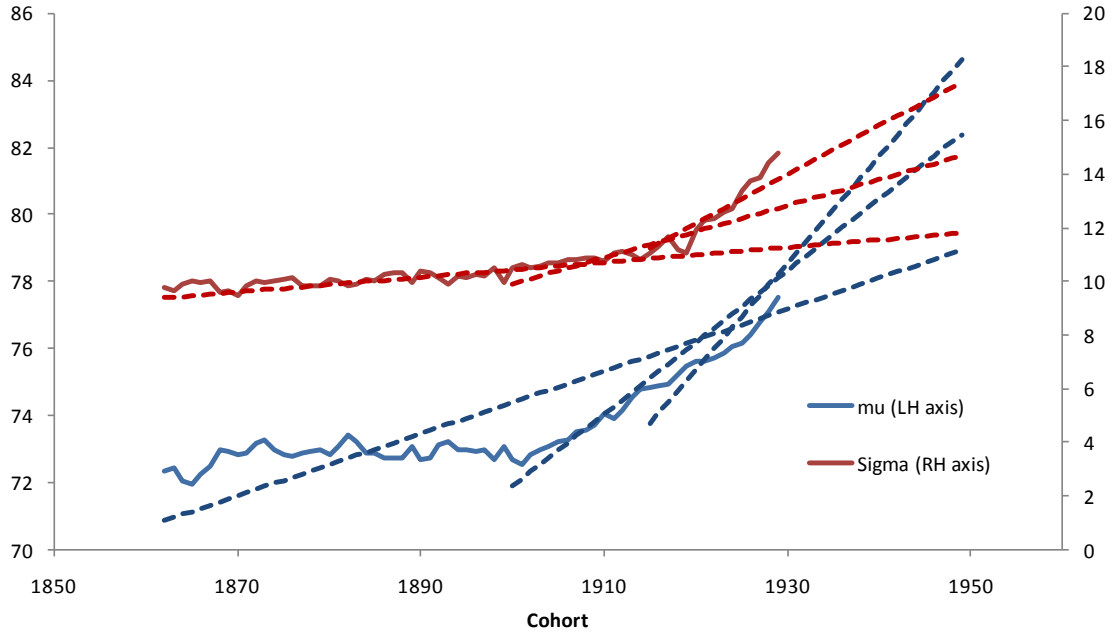

Figure 4: Parameter estimates fitted to each male cohort (solid lines) and joint estimates fitted using maximum likelihood (dotted lines), for sample periods 1862-1949, 1900-1949 and 1915-1949.

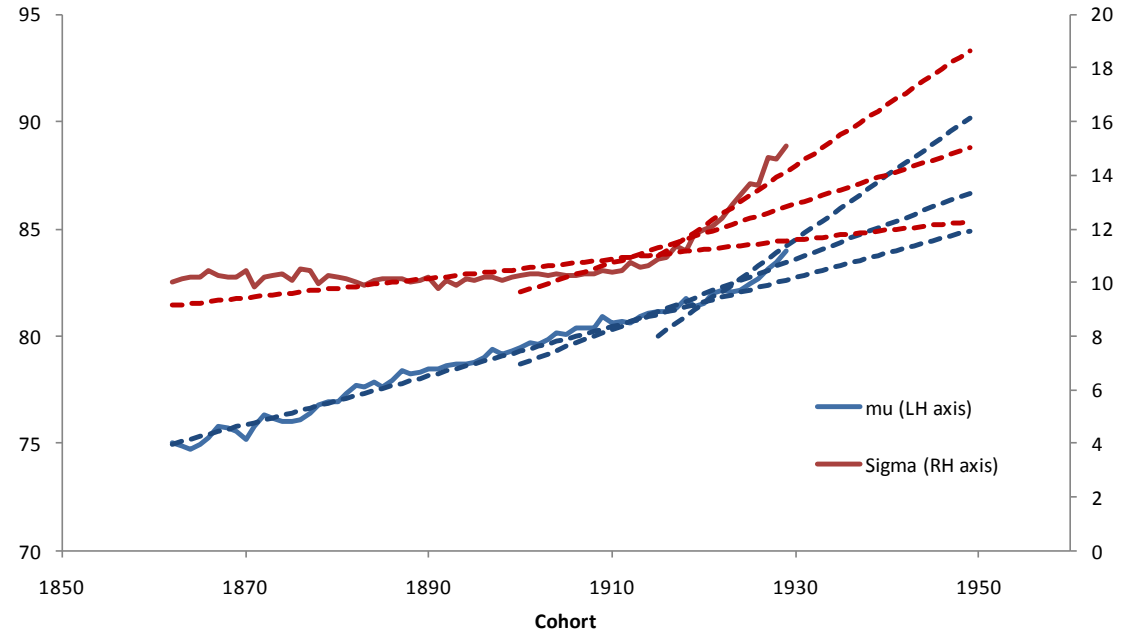

Figure 5: Parameter estimates fitted to each female cohort (solid lines) and joint estimates fitted using maximum likelihood (dotted lines), for sample periods 1862-1949, 1900-1949 and 1915-1949.

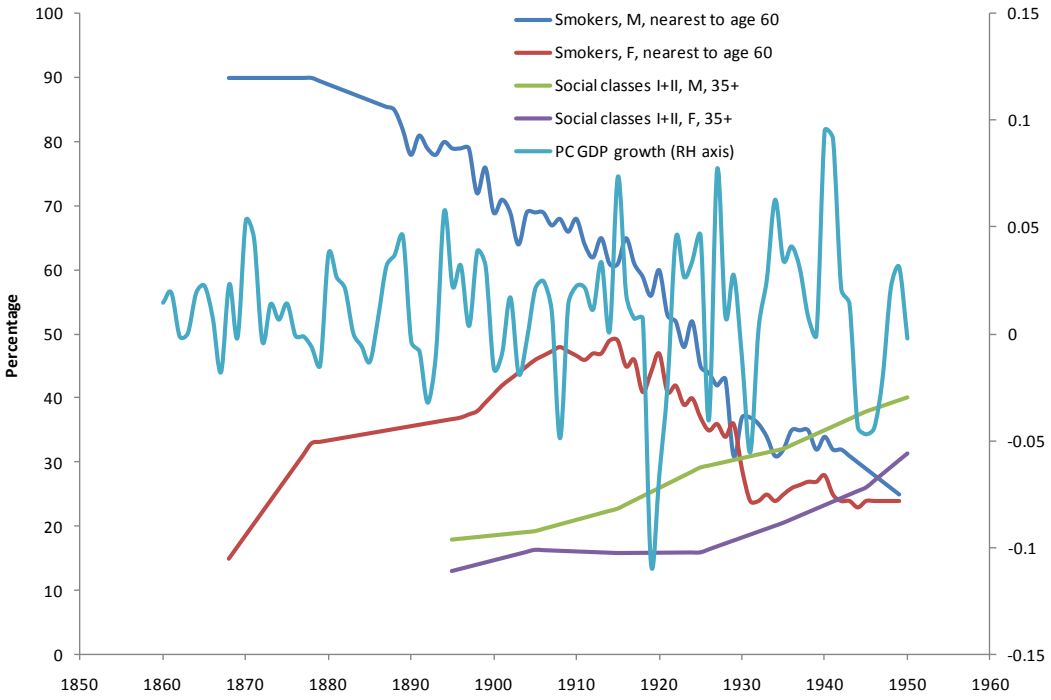

Figure 6: Proportion of each cohort who smoke at age 60, proportion in social classes I+II, age 35+, males and females and per capital GDP growth in the year before the cohort was born. From Forey et al (2009), Heath and Payne (1999), and Maddison (2009).

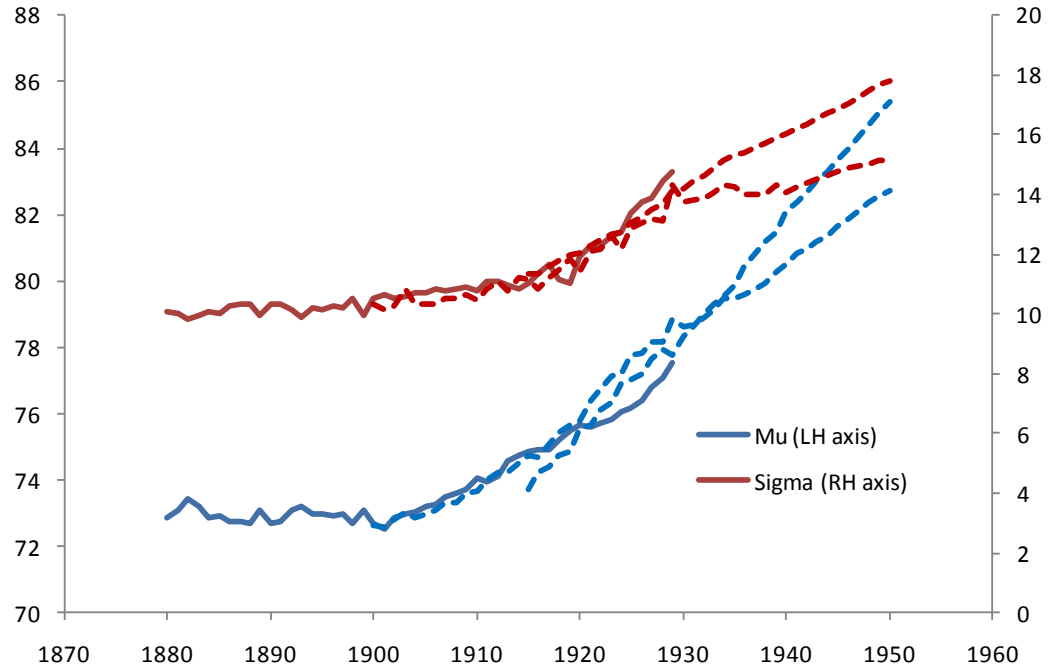

Figure 7: Parameter estimates fitted to each male cohort (solid lines) and joint estimates fitted using maximum likelihood (dotted lines), for sample periods 1880-1949, 1900-1949 and 1915-1949, including effect of smoking, social class and GDP growth.

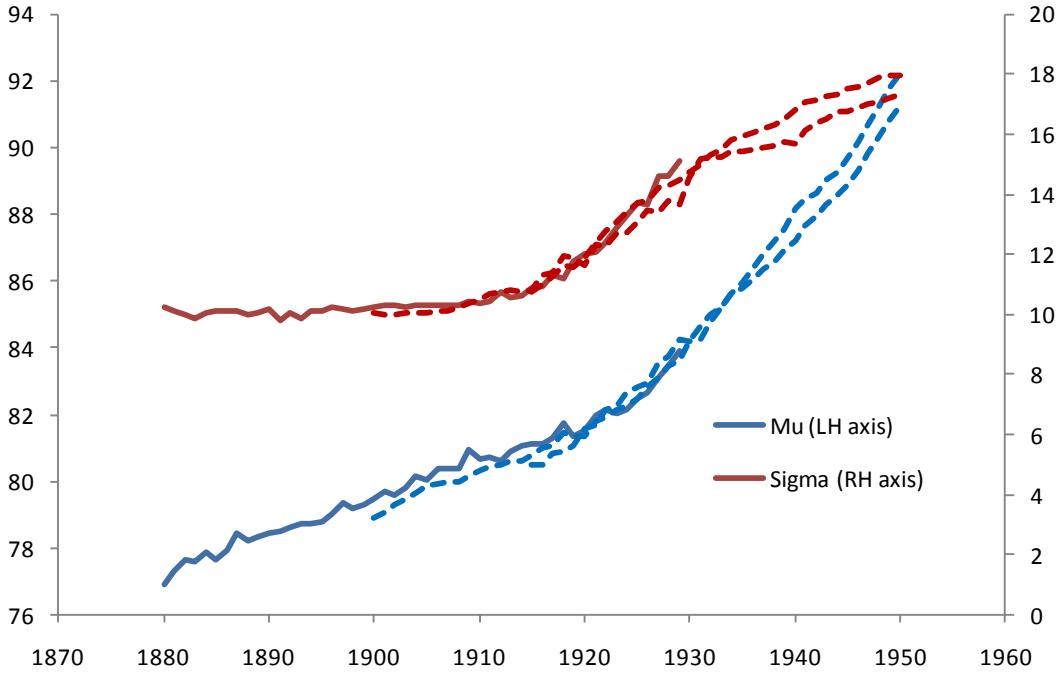

Figure 8: Parameter estimates fitted to each female cohort (solid lines) and joint estimates fitted using maximum likelihood (dotted lines), for sample periods 1880-1949, 1900-1949 and 1915-1949, including effect of smoking, social class and GDP growth.

Appendix

The normal-scaled inverse gamma distribution

Further, we can calculate the unconditional moments of the parameters μ and σ^2 in terms of the parameters of the NSIG distribution as follows:

$$E[\mu^t] = \iint\limits_{\mu,\sigma^2} \frac{\sqrt{\nu}}{\sigma\sqrt{2\pi}} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \mu^t \exp(-\frac{1}{2\sigma^2}(2\beta + \nu(\mu-\lambda)^2)) d\sigma^2 d\mu$$

If $t = 1$,

$$E[\mu] = \lambda \text{, and} \tag{A1}$$

if $t = 2$,

$$E[\mu^2] = \lambda^2 + \frac{1}{\nu}\frac{\beta}{\alpha-1}, \tag{A2}$$

so

$$Var(\mu) = \frac{1}{\nu}\frac{\beta}{\alpha-1}. \tag{A3}$$

Similarly,

$$E[\sigma^{2t}] = \iint\limits_{\mu,\sigma^2} \frac{\sqrt{\nu}}{\sigma\sqrt{2\pi}} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \sigma^{2t} \exp(-\frac{1}{2\sigma^2}(2\beta + \nu(\mu-\lambda)^2)) d\sigma^2 d\mu$$

$$= \beta^t \frac{\Gamma(\alpha-t)}{\Gamma(\alpha)}. \tag{A4}$$

So,

$$E[\sigma] = \beta^{1/2} \frac{\Gamma(\alpha-1/2)}{\Gamma(\alpha)}, \tag{A5}$$

$$E[\sigma^2] = \frac{\beta}{\alpha-1} \text{, and} \tag{A6}$$

$$E[\sigma^4] = \frac{\beta^2}{(\alpha-1)(\alpha-2)}. \tag{A7}$$

For convenience, we calculate the variance of both σ^2 and σ, giving:

$$Var[\sigma] = \beta\left(\frac{1}{\alpha-1} - \frac{\Gamma(\alpha-\frac{1}{2})^2}{\Gamma(\alpha)^2}\right), \text{ and} \tag{A8}$$

$$Var[\sigma^2] = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}. \tag{A9}$$

We also note that

$$\begin{aligned} E[\mu\sigma] &= \iint_{\mu,\sigma^2} \frac{\sqrt{\nu}}{\sigma\sqrt{2\pi}} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \mu\sigma \exp\left(-\frac{1}{2\sigma^2}(2\beta + \nu(\mu-\lambda)^2)\right) d\sigma^2 d\mu \\ &= \lambda\beta^{\frac{1}{2}} \frac{\Gamma(\alpha-\frac{1}{2})}{\Gamma(\alpha)}, \end{aligned} \tag{A10}$$

so $\text{cov}(\mu,\sigma) = 0$.

The one-stage maximum likelihood estimation of the model

If we assume that the parameters μ and σ follow a random walk with drift, the series

$$\Omega = \left\{ \begin{matrix} \mu_x \\ \sigma_x \end{matrix} \right\}_{x=1862...1929}$$

itself has a joint normal distribution with mean M and variance-covariance matrix Σ, with

$$M = \begin{bmatrix} \mu_0 \\ \sigma_0 \\ \mu_0 + x \\ \sigma_0 + y \\ \dots \\ \dots \\ \mu_0 + nx \\ \sigma_0 + ny \end{bmatrix}, \text{ and} \tag{A11}$$

$$\Sigma = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_M^2 & \rho\sigma_M\sigma_S & \sigma_M^2 & \rho\sigma_M\sigma_S & 0 & 0 \\ 0 & 0 & \rho\sigma_M\sigma_S & \sigma_S^2 & \rho\sigma_M\sigma_S & \sigma_S^2 & 0 & 0 \\ 0 & 0 & \sigma_M^2 & \rho\sigma_M\sigma_S & 2\sigma_M^2 & 2\rho\sigma_M\sigma_S & 0 & 0 \\ 0 & 0 & \rho\sigma_M\sigma_S & \sigma_S^2 & 2\rho\sigma_M\sigma_S & 2\sigma_S^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & ... & \\ 0 & 0 & 0 & 0 & 0 & 0 & & ... \end{bmatrix}, \qquad \text{(A12)}$$

and for either definition of $\ell(\{D_{x,t}\}|\Omega)$ in the text, the likelihood function of $\{D_{x,t}\}_{x=1862...1929,t=1922...2009}$ conditional on $\tilde{\theta} = \{\mu_0, \sigma_0, x, y, \rho, \sigma_M, \sigma_S\}$ is given by

$$L(\{D_{x,t}\}|\tilde{\theta}) = \int_{\Omega} (2\pi)^{-n} |\Sigma|^{-1/2} \exp((\Omega - M)'\Sigma^{-1}(\Omega - M)) \exp(\ell(\{D_{x,t}\}|\Omega) d\Omega . \qquad \text{(A13)}$$

This is a 2*n*-dimensional integral which is not currently computationally feasible for our sample size (*n*=30 for the reduced sample and 78 for the full sample, even if cohorts born after 1929 are ignored).