Modelling Socio-Economic Differences in the Mortality of Danish Males Using a New Affluence Index

Andrew J.G. Cairns,[a] Malene Kallestrup-Lamb,[b] Carsten P.T. Rosenskjold,[b] David Blake,[c] and Kevin Dowd[d]

Abstract

We investigate and model how the mortality of Danish males aged 55-94 has changed over the period 1985-2012. We divide the population into ten socio-economic sub-groups using a new measure of affluence that combines wealth and income reported on the Statistics Denmark national register database. The affluence index, in combination with sub-group lockdown at age 67, is shown to provide consistent sub-group rankings based on crude death rates across all ages and over all years in a way that improves significantly on previous studies that have focused on life expectancy. The gap between the most and least affluent is confirmed to be widest at younger ages and has widened over time.

We introduce a new multi-population mortality model that fits the historical mortality data very well and generates smoothed death rates that can be used to model a larger number of smaller sub-groups than has been previously possible without losing the essential character of the raw data.

The model produces bio-demographically reasonable forecasts of mortality rates that preserve the sub-group rankings at all ages. It also satisfies reasonableness criteria related to the term structure of correlations across ages and over time through consideration of future death and survival rates.

Keywords: Danish mortality data; affluence; CBD-X model; gravity model; multi-population mortality modelling.

[a]Joint corresponding author: Maxwell Institute for Mathematical Sciences, Edinburgh, and Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, EH14 4AS, United Kingdom. E-mail: A.J.G.Cairns@hw.ac.uk

[b]Joint corresponding author: Center for Research in Econometric Analysis of Time Series (CREATES), Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. E-mail: mkallestrup@econ.au.dk

[c]Pensions Institute, Cass Business School, City University of London

[d]Durham University Business School

1 Introduction

When trying to explain and project mortality in closely related populations, it is essential to use an appropriate modelling framework. Single-population mortality models, such as Lee-Carter and Cairns-Blake-Dowd (CBD) and their extensions,[1] are not appropriate, since they can, for example, lead to inconsistent forecasts especially in the medium and long term. It is possible for forecast death rates in an historically lower-mortality population (e.g., females) to cross over those for a higher-mortality population (e.g., males).

To avoid problems of this kind, increasing use has been made in recent years of multi-population models to explain the mortality dynamics of related populations, such as: neighbouring countries (e.g., Li and Lee, 2005, Enchev et al., 2015, Christiansen et al., 2015); males and females; smokers and non-smokers (e.g., Kleinow and Cairns, 2013); groups of annuitants and those who hold life insurance policies (e.g., Yang et al., 2014); a specific pension plan's own mortality and that of the national population (e.g., Cairns et al., 2011, Haberman et al., 2014); and different socio-economic sub-groups (e.g., Li et al., 2015).

In some of these cases, especially those that involve sub-populations of the national population, we might have limited or even no data with which to model. However, we can use the experience of the larger national population to help inform and improve forecasts of the population that we are really interested in. In other settings, we might seek to manage actively the risks that we identify. This is most apparent in the life insurance and pensions worlds where, for example, some risk management approaches require joint modelling of the mortality experiences of both an annuity provider and the national population (see, e.g., Coughlan et al., 2011, Li and Hardy, 2011, Haberman et al., 2014, and Michaelson and Mulholland, 2015).

We have been given access to a unique and extremely comprehensive database from Statistics Denmark that enables us to model the socio-economic mortality of older Danish males at a finer level of granularity than has hitherto been possible. This is valuable for a number of reasons: it allows the authorities to generate more accurate estimates of the cost and projected increases in the cost of state pension benefits in different population segments; it helps corporate pension plans improve estimate of their liabilities, given the socio-economic mix of their plan members; and it allows annuity providers to price annuities on a socio-economic basis.

The key contributions of this paper based on an analysis of older Danish male mortality data are as follows:

- We propose a new, affluence index for subdividing the Danish male population into ten socio-economic sub-groups of equal size. We allow individuals to

[1] Lee and Carter, 1992, Cairns et al., 2006b, and the extensions summarised in Mavros et al., 2014, and Hunt and Blake, 2014

transfer between sub-groups prior to the official state pension age of 67, but thereafter to remain in the same sub-group to which they were allocated at this age.

- The proposed index, in combination with the stochastic mortality model remarked on below, allows us to subdivide the male population into a larger number (10) of smaller sub-groups, while still preserving statistical significance.

- We demonstrate that the affluence index plus lockdown at age 67 provides a much improved separation of the 10 deciles *at all ages*. Specifically, whereas previous work has focused on *life expectancy* from a limited range of ages, the new method achieves a clear ranking of *death rates* at all ages from 55 to 94 across all 10 sub-groups.

- We develop a new stochastic multi-population mortality model for forecasting. The model fits the historical data well and produces coherent and biodemographically reasonable forecasts of future death rates in the 10 sub-groups across all ages and a range of time horizons.

- We analyse how future uncertain death rates and survival probabilities in the 10 sub-groups are related to each other, through a detailed analysis and discussion of the term structure of correlations.

- We discuss what lessons can be learned about the dependencies between different socio-economic sub-groups in other countries.

- We propose, but do not investigate, a method for modelling mortality at very high ages (i.e., above age 95).

The outline of the paper is as follows. Section 2 introduces and explains the Danish national mortality dataset. It also explains the approach used to process the data using the new affluence index. Section 3 sets out the proposed multi-population gravity model that we use to model mortality by socio-economic grouping. Section 4 presents the model fit for a range of years and ages and confirm that our approach to modelling fits the data well, smoothing out the effects of sampling variation while still preserving the essential characteristics of the crude sub-group mortality data. Section 5 analyses properties of projected mortality: central trends, uncertainty and correlations between sub-groups. Section 6 offers some suggestions for further research, and Section 7 concludes.

2 Danish males data

The analysis in this paper makes use of a dataset from Statistics Denmark (SD), based on administrative records. Since every individual in Denmark is given a central personal register number (CPRN) either at birth or when given residence permission in the country, we are able to uniquely identify each individual in all areas of the public register system which includes the Population Register, the Integrated Database for Labor Market Research, the Income and Tax Register, and the Cause of Death Register. Thus, for each individual, we have information on their date of birth, education, labour market status, income, tax, and, ultimately, their date of death and cause of death. We can also identify the same information for an individual's spouse or partner, thereby enabling us to allocate income and wealth between couples and within households. On an annual basis, we observe their marital status, based on the following four categories: unmarried, married, divorced, and widow/widower. We also know the precise date of marriage. Separately, cohabitation status, and with whom, is also recorded. Significantly, the information contains no survey element. In general, we have access to a very high quality dataset.

The dataset allows us to identify three financial indicators for each individual and couple, all deflated to the 2000 real values: gross individual annual income, total net household income, and household net wealth. All financial measures are based on calculations from the tax authorities which are linked to the CPRN. Gross annual income includes all taxable income, such as wage income, self-employment income, unemployment insurance benefits, social assistance (from 1994), honoraria, and all types of pension-related income. However, all payments withdrawn from labour income into non-taxable pension schemes, such as labour market pension schemes, ATP (the supplementary income-related pension scheme), private capital pension schemes, as well as annuity pension schemes, are not included in the gross annual income measure.[2] For retired individuals, we observe a break in the gross annual income variable from 1994 onwards. There are several reasons for this. The level of old age pension benefits was increased in 1994 as the government removed a special tax rebate previously given to retirees. Moreover, individuals living in retirement homes prior to 1994 were only given a monthly allowance, but had no expenses to cover rent, heating, electricity, etc. After 1994, all individuals, irrespective of their living situation, were given the full amount of old age pension benefits or disability pension benefits.

Total net household income is defined as total gross annual income for all members of the household, net of alimony, tax and interest payments. One drawback with this measure is that there is no standardized household size. Such a measure could be obtained by using standardized household income = total household income/(# of adults)$^{0.7}$, as suggested by Citro and Michael (1995).

[2]This is due to the taxation being postponed until the payout stage of these pension schemes.

Household net wealth is calculated as total assets minus total liabilities at year end. Total assets include real estate, bank deposits, bonds, stocks, mortgage deeds, shares in firms, the value of cars, boats and mobile homes, as well as cash. Liabilities include credit secured by a mortgage on real estate, debt to financial institutions, credit card debt, and all other types of debt to private companies and the government (e.g., unpaid tax). Prior to 1983, all assets were assigned to the male spouse. Thereafter, 50% is allocated to each spouse, unless otherwise reported to the tax authorities. From 1987 to 1996, net wealth also included the taxable equity value of self-employed businesses (which can be negative). In 1997, a wealth tax was abolished, which resulted in the equity value of self-employed companies, the value of some mortgage deeds (not related to real estate), shares in firms, and the value of cars, boats and mobile homes are no longer being reported in the wealth measure.[3]

Data are available for calendar years 1980 to 2012. However, as the quality of both the income and wealth data (especially for married couples) is regarded as poor at the beginning of the period, we disregard the first four years of the available data points. In addition, since our methodology for allocating individuals into sub-groups, requires data for the previous calendar year, we calculate exposures and deaths data for each sub-group for the 1985-2012 period only.

One of the key objectives of this study is to identify covariates in the SD database that have a strong predictive power in explaining the mortality of different individuals in the population. More specifically, our aim is to subdivide the population at each age and in each year into 10 approximately equal-sized sub-groups, with a clear ordering between sub-groups in terms of mortality *at all ages*: that is, group 1 should have the highest mortality at all ages and group 10 the lowest. This number of sub-groups is rather larger than is typically the case in related studies of small populations (e.g., Brønnum-Hansen and Baadsgaard, 2012, who subdivide into quartiles), and gives us the subsequent flexibility to aggregate sub-groups should this be desirable or necessary in the modelling work. For large populations, such as the US, the use of deciles produces meaningful results without smoothing (e.g., Waldron, 2013). For Denmark, we show that separation into deciles also gives meaningful results, but requires more work to filter out or smooth objectively the effects of sampling variation.

In this particular study, we aim to identify a *single* covariate that is a strong predictor of mortality, subject to the condition that the covariate represents a quantity that is readily available in other datasets, thereby enabling our modelling framework to be applicable to a wide range of countries, including those that are less comprehensive than the SD database. We therefore decided to focus on income and wealth, since these are well-known predictors of mortality (e.g., Rogot et al. 1992, Wolfson et al., 1993, Chapman and Hariharan, 1996, McDonough et al., 1997, Sabel et al., 2007,

[3]This change might have resulted in a few more transitions between groups than normal. But because the rankings are relative, we do not believe that there has been a material impact, nor is there any statistical evidence.

and Demakakos et al., 2015).

The specific candidate covariates considered were income, individual wealth, household wealth and linear combinations of these: in particular, wealth plus $K\times$income where K is some constant to be determined (e.g., $K = 5, 10, 15$ or 20). We also considered the exclusion of individuals with negative income and/or wealth and self-employed individuals, for whom there was a suspicion that the reported income/wealth values did not accurately reflect their true financial position. Ultimately, all of these individuals were included, giving us almost complete coverage of the population, with the exception of nationals living abroad.

In general, wealth and income were found individually to be strong predictors of mortality. However, for most of the variants considered above, results were not satisfactory across all ages and/or in all years. Typically, we would find consistent mortality rankings across age for the higher (more wealthy) sub-groups, while lower sub-groups might be correctly ranked at some ages but not at other ages. In most of our experiments, we found that (a) high income or high wealth is a strong predictor of low mortality, but (b) low income or low wealth, taken separately, is not a strong predictor of high mortality. For example, a 70-year-old might have little capital, but still have a good pension that allows him to live a healthy lifestyle. Another 70-year-old might have no pension, but have substantial personal savings that he can draw on to provide a good income in retirement – income that would not be recorded as such by SD. We also found that the registered level of wealth was, in many cases, negative for working age males: this seems to reflect the fact that registered wealth takes account of loans and mortgages, but not all of the corresponding assets at market value (e.g., the value of a self-employed person's business set up with the help of a bank loan).

2.1 Allocation of individuals to sub-groups

Following a great deal of experimentation, we settled on the following algorithm for allocating older Danish males to one of the 10 sub-groups, based on a financial indicator variable that we label as the *affluence index*. This index is obtained specifically from the income and wealth available for each individual for each year. Allocation to sub-groups 1 to 10 is based on data that are available at or before the start of the year, meaning that we use relevant data for the previous year.

The first step in the allocation algorithm is to determine the value of the index. The affluence index, $A(i, t, x)$, for individual i, in year t at age x (at the start of the year) is defined as the individual's wealth plus K times their income in the preceding year, that is:

$$A(i, t, x) = W(i, t-1, x-1) + K \times Y(i, t-1, x-1) \qquad (1)$$

where $W(i, t-1, x-1)$ is the (physical and financial) wealth for individual i, at age $x-1$ in year $t-1$ and $Y(i, t-1, x-1)$ is the income of individual i, at age $x-1$ in year

$t-1$. K is a constant across the whole population. Income and wealth are highly correlated, but the variables combine to create a more robust index, particularly for higher-mortality sub-groups. We have chosen to use $K = 15$, but the derived sub-group death rates outlined below are found to be robust relative to the value of K.[4]

The second step is then to allocate all individuals to specific sub-groups. Based on the affluence index, $A(i,t,x)$, all individuals, i, in year t at age x are ranked from $1, 2, ..., n$, where n is the number of individuals at age x in year t. The rank of the individuals is given by $R(i,t,x)$. This rank is normalised by $U(i,t,x) = R(i,t,x)/(n+1)$, so it is evenly spread between 0 and 1. Then, if $U(i,t,x)$ lies between 0 and 1/10, the individual is allocated to Group 1, and so on. This procedure implies that 10% are assigned to each sub-group.

The ranking and allocation is repeated every year until an individual reaches the main state pension age of 67.[5] After age 67, each individual remains in the same sub-group that they were allocated to at age 67. Thus, individuals are assumed to be able to migrate between affluence sub-groups before age 67, but not after. From these population sub-groups, it is then possible to calculate the demographic measures exposed to risk, death counts, and death rates.

The key contributions of this algorithm compared to earlier ones are twofold. First, we found that the use of affluence is a more effective discriminator *across all ages* than income or wealth on their own. Second, the lockdown at age 67 produces much better results (including improved separation between the 10 sub-groups) than not locking down.[6]

2.2 Deaths, exposures and crude death rates

We exploit the detailed nature of the database to calculate deaths and exposures in a different way from standard sources (e.g., the Human Mortality Database count deaths according to the age last birthday at the time of death) to obtain a greater degree of precision. Here our age variable refers to the age at the start of each calendar year. Thus

- $D(i,t,x)$ = the number of individuals allocated to sub-group i at the start of

[4] The original choice of $K = 15$ reflects the idea that, around the age of retirement, 15×income is a very approximate estimate of the present value of an individual's future retirement income; in other words, K can be interpreted as a capitalisation factor.

[5] The state pension age was reduced from 67 in 2004 to 65, although it will increase to 67 between 2024 and 2027, with further increases after 2027 that are linked to increases in life expectancy.

[6] An additional advantage of the lockdown is that it restricts the potential for Group 1 to fill up with individuals in severely declining health who have used up most of their personal savings on long-term care. This would artificially inflate Group 1 mortality at high ages and depress it in the more affluent sub-groups.

the year, born in calendar year $t-1-x$,[7] who die during calendar year t.

- $E_I(i,t,x)$ = the number of individuals allocated to sub-group i, born in calendar year $t-1-x$, and counted on the first of January in year t.[8]

- $E(i,t,x)$ is the corresponding central exposed to risk, and measures the average size of the population during year t in sub-group i born in year $t-1-x$. Migration prevents us from calculating $E(i,t,x)$ exactly, so we have chosen to use the common approximation $E(i,t,x) = E_I(i,t,x) - D(i,t,x)/2$.[9]

The crude death rate is then

$$\hat{m}(i,t,x) = D(i,t,x)/E(i,t,x)$$

in sub-group i, calendar year t and age x last birthday *at the beginning of the year*.

Exposures, $E(i,t,x)$ range from about 4250 (at age 55) down to 13 (at age 94). Deaths, $D(i,t,x)$, ranged from 151 (peak mortality ages) down to 4 (age 94). It is evident, therefore, that subdividing an already small national population produces crude death rates where sampling variation (also known as Poisson risk) will be quite significant.

By way of example, crude death rates for the 10 sub-groups in 2012 are plotted in Figure 1 (left).[10] This plot is typical of what we see for all years 1985 to 2012. Despite the significant levels of sampling variation along each curve, we can still see a clear ranking in the plot: Group 1 has consistently high death rates across all ages; Group 2 the next highest; and so on down to Group 10. In particular, the rankings are what could be described as bio-demographically reasonable: the more affluent someone is, the less likely they are to die in the next year compared with a less affluent person of the same age. Sampling variation introduces some crossovers in the left-hand plot in Figure 1, especially at high ages, but the broad patterns are clear. We can also observe that relative differences in death rates between sub-groups are biggest at younger ages and smallest at very high ages.

We experimented with numerous ways of segmenting the population, but however we did it, we always ended up with a similar (or even narrower) spread of curves at higher ages. This finding is in line with the *compensation law of mortality* (Gavrilov and Gavrilova, 1979, 1991; see, also, Avraam et al., 2014) which observes that the mortality for different populations tend to converge with age. This 'law' implies that wealth and lifestyle-related factors have a lesser impact as we age, whilst other,

[7] Equivalently individuals: (a) who are age x at the start of calendar year t or (b) who attain their $x+1^{th}$ birthday during year t.

[8] $E_I(i,t,x)$ is the initial exposed to risk.

[9] Alternative approximations for $E(i,t,x)$ were considered, but these resulted in only very small differences compared with the formula used in this study.

[10] Knowledge of death rates up to age 94 allows us to calculate survivorship up to exact age 95.

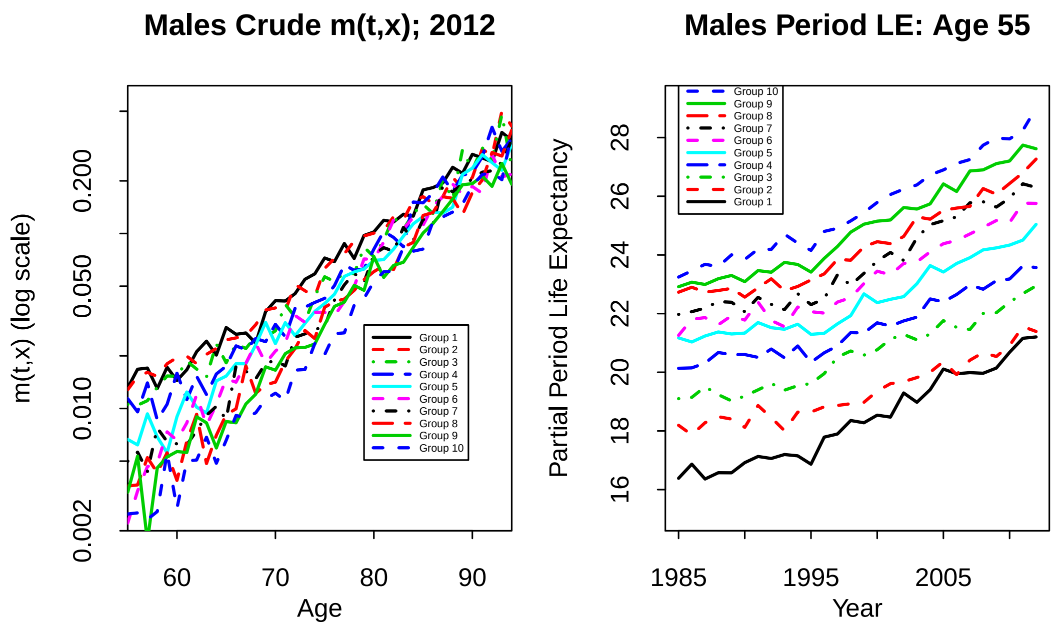

Figure 1: Left: Crude death rates, $m(i,t,x) = D(i,t,x)/E(i,t,x)$, for Danish males Groups 1 to 10 by age in 2012. Right: Crude partial period life expectancies for Danish males Groups 1 to 10 from age 55, capped at 95.

mainly genetic, factors become more significant. The narrowing gap with age is also consistent with the findings of Waldron (2007), Cristia (2009) and Bosworth and Burke (2014) for US males, and Office for National Statistics (2013) and Villegas and Haberman (2014) for England and Wales males.

Partial period life expectancies, $LE(t,x)$, for age x =55 are plotted in Figure 1 (right).[11] We can note the following points. At age 55 (and the same holds for other ages), there is a clear separation between partial period life expectancies (LEs) in all years across all 10 sub-groups (Group 1 has the lowest LE; Group 10 has the highest), even though we are using unsmoothed data. The gap between Groups 1 and 10 narrows as age increases, reflecting the narrowing gap between death rates observed in the left-hand plot in Figure 1. The gap between crude LEs for Groups 1 and 10 widened between 1985 and 2012 at both ages 55 (6.9 years widening to 7.8 years) and 67 (4.0 widening to 5.3).

[11] We define partial period life expectancy, $LE(t,x) \approx \frac{1}{2} + \sum_{y=x+1}^{x_u-1} S_p(t,x,y) + \frac{1}{2} S_p(t,x,x_u)$, to be the expected number of years survived from age x to age $x_u = 95$, assuming that mortality stays at the same crude levels as in year t. The crude period survival probabilities are defined as $S_p(t,x,y) = \exp\left[-\sum_{s=x}^{y-1} \hat{m}(t,s)\right]$. Partial life expectancy up to age 95 is only very slightly less than complete (untruncated) life expectancy, but the latter requires extrapolation of death rates beyond our upper age.

Similar observations have been found previously with Danish data by Brønnum-Hansen and Baadsgaard (2012) using *disposable income* as a covariate. However, Brønnum-Hansen and Baadsgaard only report period life expectancy at a single age 0 and separate the population into quartiles. By contrast, we have carried out a much more detailed analysis using deciles rather than quartiles, partial period life expectancy over a range of ages from 55 and up, and death rates at individual ages from 55 to 94. Our conclusion is that the use of income as the key metric could, as discussed above, be improved upon significantly. Specifically, income, on its own, does not achieve a satisfactory separation of the 10 sub-groups across all ages and all years: a problem that has now been resolved through the use of the $A = W + 15Y$ affluence index with lockdown at age 67. Our analysis of mortality across a wide age range is important in the context of the intended applications. In particular, we seek to model sub-population death rates with a view to assessing how sub-groups will evolve over time relative to each other at all ages, and to assess the impact financially on the providers of pensions and annuities.

The widening over time of the gap in life expectancy is not unique to Denmark and can be observed elsewhere. The UK Office for National Statistics (2011) consider England and Wales life expectancies by occupation group and find a widening gap between professional/managerial and unskilled manual workers; Cristia (2009) considers US males and females subdivided by lifetime earnings and also finds a widening gap over time; Tarkiainen et al. (2012) find an increasing gap in Finland by income; and Mackenbach et al. (2003) find increasing gaps in six European countries using other socio-economic measures.

3 Modelling sub-population mortality

Multi-population mortality modelling concerns the development over time of the death rates in several populations. Let $m(i, t, x)$ be the death rate for population i in year t for individuals aged x last birthday on the first of January in the year of death. A standard hypothesis in multi-population modelling is that the relationship between the death rates at given ages in two related populations should not diverge over time: that is, the ratio $m(i, t, x)/m(j, t, x)$, for $i \neq j$, should remain stable over time. This stability condition is often referred to as *coherence*. Coherence can be achieved through the use of certain models, such as the *gravity model* which achieves stability in an intuitive and plausible way (see, for example, Li and Lee, 2005, Cairns et al., 2011, and Dowd et al., 2011). Other approaches to multi-population modelling include those of Jarner and Kryger (2011) and Kleinow (2015).

We propose the following gravity model of the CBD-X type for underlying death rates in the 10 sub-groups:

$$\log m(i, t, x) = \beta_0^{(i)}(x) + \kappa_1^{(i)}(t) + \kappa_2^{(i)}(t)(x - \bar{x}) \tag{2}$$

where i is the sub-group, t is the year and x is the age last birthday *at the start of the year*. This is a variant in the style of Plat (2009) of the CBD model (Cairns et al., 2006) that adds a non-parametric age effect, $\beta_0^{(i)}(x)$, to the basic CBD model, and models the log death rate rather than the logit of the mortality rate; see, also, Hunt and Blake (2014). The $\kappa_1^i(t)$ terms capture changes in the level of mortality and the $\kappa_2^i(t)$ terms changes in the slope of the log-mortality curve relative to the baseline $\beta_0^{(i)}(x)$. The non-parametric age effect was found to be necessary to preserve the mortality rankings between sub-groups over the full range of ages from 55 to 94. Without it, we found that the mortality curves in individual years for different sub-groups would cross over at high ages in a way that was not consistent with the crude death rates. As we discuss later, this model was found to fit the 10 sub-populations of the Danish males aged 55 to 94 well without the need for a cohort effect.

For the purpose of both fitting the model to historical data and for forecasting, we need to specify a stochastic model for the period effects. We propose the following:

$$\kappa_1^{(i)}(t) = \kappa_1^{(i)}(t-1) - \psi\left(\kappa_1^{(i)}(t-1) - \bar{\kappa}_1(t-1)\right) + \mu_1 + Z_{1i}(t) \qquad (3)$$

$$\kappa_2^{(i)}(t) = \kappa_2^{(i)}(t-1) - \psi\left(\kappa_2^{(i)}(t-1) - \bar{\kappa}_2(t-1)\right) + \mu_2 + Z_{2i}(t) \qquad (4)$$

where

$$\bar{\kappa}_1(t) = \frac{1}{n}\sum_{i=1}^{n}\kappa_1^{(i)}(t)$$

$$\bar{\kappa}_2(t) = \frac{1}{n}\sum_{i=1}^{n}\kappa_2^{(i)}(t)$$

and the random innovation terms $Z_{ki}(t)$ are multivariate normal with mean 0 and covariances

$$Cov\left(Z_{1i}(t), Z_{1j}(t)\right) = \begin{cases} v_{11} & \text{for } i = j \\ \rho v_{11} & \text{for } i \neq j \end{cases}$$

$$Cov\left(Z_{2i}(t), Z_{2j}(t)\right) = \begin{cases} v_{22} & \text{for } i = j \\ \rho v_{22} & \text{for } i \neq j \end{cases}$$

$$Cov\left(Z_{1i}(t), Z_{2j}(t)\right) = \begin{cases} v_{12} & \text{for } i = j \\ \rho v_{12} & \text{for } i \neq j \end{cases}$$

with $-1 < \rho < 1$. Additionally, the $Z_{ki}(t)$ are independent from one year to the next.

The $\kappa_1^{(i)}(t)$ share a common drift, μ_1, and the $\kappa_2^{(i)}(t)$ share a common drift, μ_2, where μ_1 and μ_2 need to be estimated. The components of equations (3) and (4)

$$-\psi\left(\kappa_1^{(i)}(t-1) - \bar{\kappa}_1(t-1)\right) \quad \text{and} \quad -\psi\left(\kappa_2^{(i)}(t-1) - \bar{\kappa}_2(t-1)\right)$$

represent gravity effects (similar to Cairns et al., 2011, and Dowd et al., 2011a) between sub-groups that prevent individual sub-group death rates from drifting away from the overall trend, with $0 < \psi < 2$ to ensure stationarity.

From equations (3) and (4), it is straightforward to show that $(\bar{\kappa}_1(t), \bar{\kappa}_2(t))$ is a bivariate random walk with drift $(\mu_1, \mu_2)'$ and one-step-ahead covariance matrix

$$\frac{1 + (n-1)\rho}{n} \begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix}.$$

Next, define

$$\Delta_{1i}(t) = \kappa_1^{(i)}(t) - \bar{\kappa}_1(t). \tag{5}$$

Then, the $\Delta_{1i}(t)$ are correlated AR(1) processes with AR(1) parameter $1 - \psi$, that revert to 0, add up to 0, and which are independent of $\bar{\kappa}_1(t)$. Similar remarks apply to $\Delta_{2i}(t) = \kappa_2^{(i)}(t) - \bar{\kappa}_2(t)$.

We have deliberately chosen to have a single short-term, contemporaneous correlation parameter, ρ, and a single gravity parameter, ψ, to keep the model simple and robust, as well as to benefit from computational efficiencies. This means that the correlations in log death rates between all pairs of sub-groups are all the same.

For forecasting, we are interested in assessing correlations between the sub-groups. We can remark that the T-year-ahead correlations between sub-groups will depend on ρ and ψ as well as the v_{ij}.

3.1 Estimation

The $\beta_0^{(i)}(x)$, $\kappa_1^{(i)}(t)$ and $\kappa_2^{(i)}(t)$ were estimated using Bayesian methods with the following elements. We seek to estimate the posterior distribution for β (representing all of the $\beta^{(i)}(x)$), κ (representing all of the $\kappa_j^{(i)}(t)$) and ϕ (representing all of the process parameters governing the dynamics of the $\kappa_j^{(i)}(t)$), given the detailed information, D, about deaths by sub-group, year and age. The posterior distribution is proportional to

$$f_1(D|\beta, \kappa, \phi) f_2(\beta, \kappa|\phi) f_3(\phi)$$

where f_1 is the probability of observing $D(i, t, x)$ deaths given β, κ, ϕ, and the exposures, f_2 is the density function for β and κ given ϕ, and f_3 is the prior density for the process parameters, ϕ.

$f_2(\beta, \kappa|\phi)$ is based on the multivariate time series structure for the $\kappa_j^{(i)}(t)$. In our formulation, β plays no role in f_2: that is, our prior assumption is that the $\beta_0^{(i)}(x)$ are independent of each other and have an improper uniform prior distribution; and we let the deaths data drive estimation of the $\beta_0^{(i)}(x)$. For the given combination of equations (2), (3) and (4), we only need two identifiability constraints to uniquely identify the posterior density. We choose to fix $\bar{\kappa}_1(1) = 0$ and $\bar{\kappa}_2(1) = 0$. Beyond these constraints, we also need a distributional assumption for $\Delta(t) = \{\Delta_{ji}(t) : j = 1, \ldots, 10; i = 1, 2\}$ (equation 5) at time $t = 1$. We assume the stationary distribution for $\Delta(1)$.[12]

In practice, f_1 depends only on D, β and κ, and not, additionally, on ϕ. For modelling the conditional distribution of the deaths given κ, we use conditionally independent normal distributions for the $\log D(i,t,x)$ (conditional on κ) with mean $\log \hat{m}(i,t,x)E(i,t,x)$ and variance $1/D_{obs}(i,t,x)$, where $D_{obs}(i,t,x)$ is the observed number of deaths. The conditional log-normal is used as an approximation to the usual conditional Poisson distribution for the $D(i,t,x)$. The use of $1/D_{obs}(i,t,x)$ is an approximation to the variance of the log of a Poisson random variable with mean $\hat{m}(i,t,x)E(i,t,x)$. It is, of course, self-referential, but it works well in practice and can be considered as an application of Empirical Bayes. For further discussion, see Cairns et al. (2016).

The use of the log-normal for deaths in combination with pre-specified variances plus the given time series model for κ results in a log-likelihood function that is quadratic in the latent state variables, $\beta_0^{(i)}(x)$, $\kappa_1^{(i)}(t)$ and $\kappa_2^{(i)}(t)$. An advantage of this specification is that when we use Markov chain Monte Carlo (MCMC) to sample from the posterior distribution for the model parameters, we can use computationally efficient Gibbs sampling from the conditional posterior distributions (i.e., the multivariate normal) to update the $\beta_0^{(i)}(x)$, $\kappa_1^{(i)}(t)$ and $\kappa_2^{(i)}(t)$.

The log-likelihood for ρ does not lead (in combination with any sensible choice of prior) to a simple conditional posterior distribution for ρ, however. To remedy this shortcoming, we use the Metropolis-Hastings (MH) algorithm for updating ρ instead of the Gibbs sampler. Estimation of the posterior distribution for ψ also uses the MH algorithm, for similar reasons.[13]

The parameters v_{11}, v_{22} and v_{12} are also estimated, but in a constrained way. Specifically, $v_{ij} = \nu \hat{v}_{ij}$, where the $\hat{v}_{ij}$ are specified constants and the scalar parameter, $\nu > 0$, has to be estimated. The prior point estimates of $\hat{v}_{11}$, $\hat{v}_{22}$ and $\hat{v}_{12}$ exploit Empirical Bayes and are based on estimated values for the total Danish population. We assume that the random walk processes for each of the 10 sub-populations are, individually, no more or less volatile than the national population. The parameter ν has an inverse gamma conditional posterior distribution, which we exploit to allow us to use the Gibbs sampler for updating ν.

Uniform prior distributions are assumed throughout apart from:

- ρ has a Beta prior to ensure it remains in the range $(0, 1)$. We tried both

[12] Since $0 < \psi < 2$ and the $Z_{ki}(t)$ are multivariate normal, the stationary distribution of $\Delta(t)$ required for $\Delta(1)$ exists and is also multivariate normal.

[13] For an introduction to MCMC, the MH algorithm and the Gibbs sampler, see Gilks et al. (1996).

Beta$(2, 2)$ (equivalent to the Uniform distribition, and this is as weak as we could reasonably go) and Beta$(3, 3)$ priors. Both are quite uninformative and produce similar, although not identical, results for the posterior for ρ.

The posterior distribution for ρ is centred around 0.45 with a standard deviation of about 0.1, so the influence of the prior in the tails should not be critical.

In the results that follow we use the weaker, less-informative Beta$(2, 2)$ prior.

- ψ also has a Beta prior to ensure it remains in the range $(0, 1)$. We tried both Beta$(2, 2)$ and Beta$(3, 3)$ priors with similar, but not identical, results for the posterior for ψ.

  The posterior for ψ is quite skewed towards 0 and so the choice of the prior parameters can, potentially make a difference. In fact, the Beta$(2, 2)$ (which pushes ψ less strongly away from 0) produces a posterior for ψ that is a bit closer to 0 than the Beta$(3, 3)$ prior.

  In the results that follow, we use the weaker, less-informative Beta$(2, 2)$ prior.

- ν has an inverse gamma prior centred close to 1 or a bit higher.[14] Since the log-likelihood function is quadratic in relevant latent state variables, ν has an inverse gamma distribution for its conditional posterior distribution.

Sensitivity of key model outputs to the choice of priors for ρ and ψ is discussed later in Section 5.

4 Analysis of Historical Death Rates

4.1 Fitted death rates

Figure 2 shows fitted death rates for the years 1985 and 2012 across all ages and for ages 60 and 80 across all years, and can be compared to the crude death rates in Figure 1. The fitted rates smooth out the noise very considerably and achieve a crisp distinction between each of the sub-groups. At the same time, this clarity is achieved without losing any of the essential patterns and characteristics underpinning the crude rates. From these plots of smoothed fitted death rates (and others not included here), we can confirm the findings we first observed from plots of crude rates:

- Falling death rates over time at all ages.

[14] Specifically, the prior has shape parameter 11 and rate parameter 10, giving a mean of 1 and a standard deviation of 1/3.

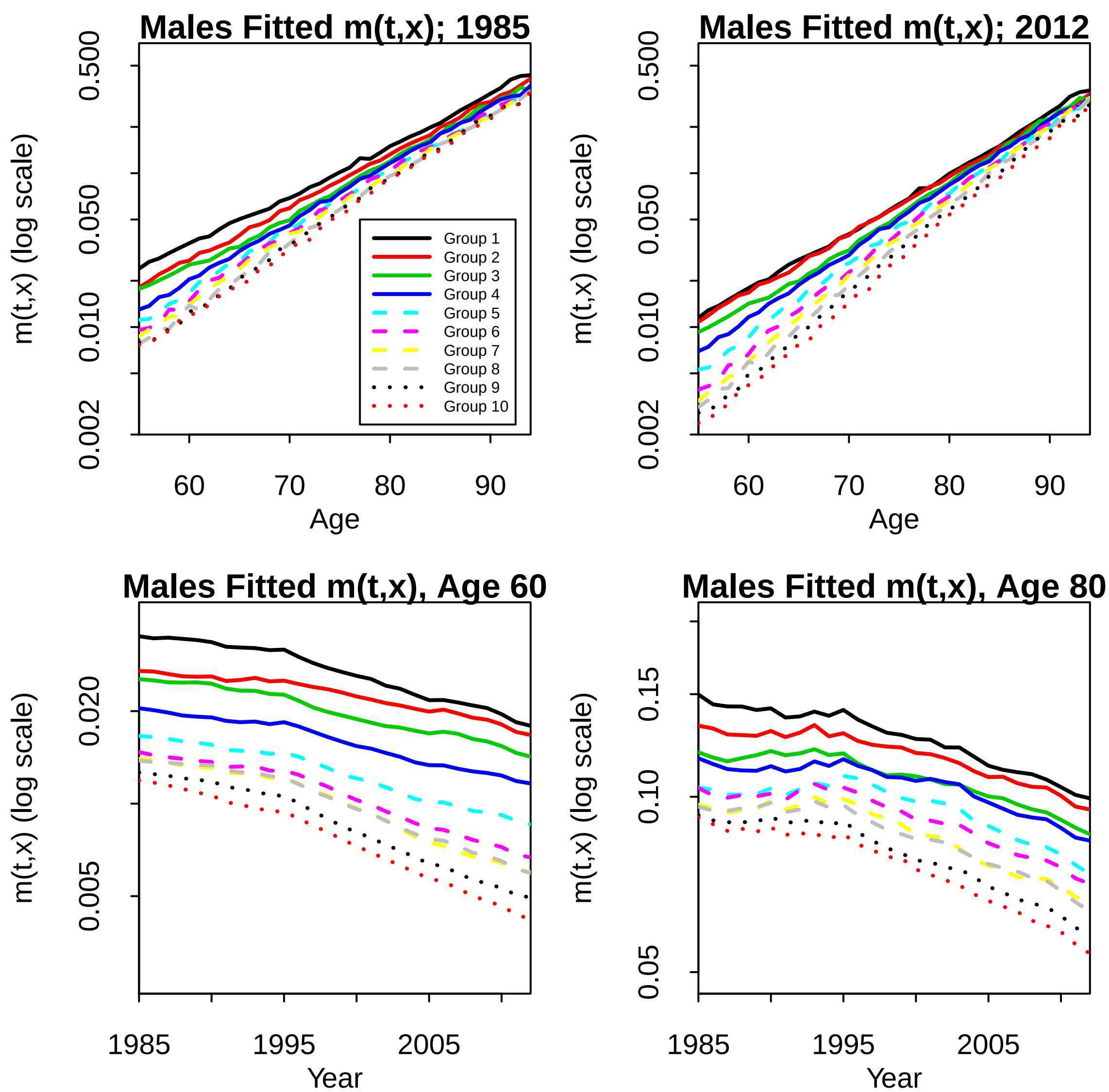

Figure 2: Fitted death rates, $m(i,t,x)$, for Danish males Groups 1 to 10 using the CBD-X model, with sub-groupings based on the affluence index (wealth+15×income) and lockdown at age 67. Top row: by age in years 1985 and 2012. Bottom row: by year for ages 60 and 80.

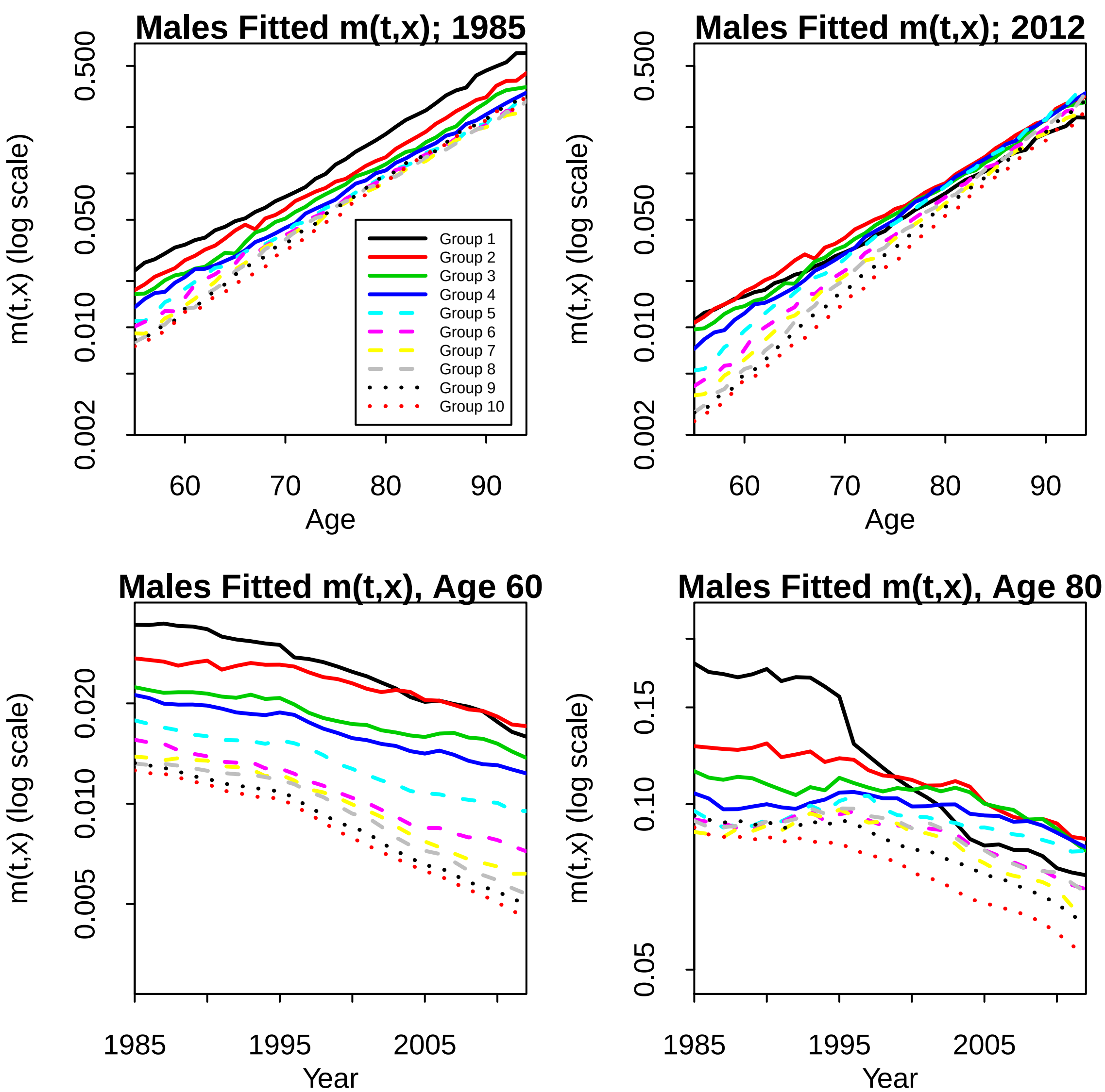

Figure 3: Fitted death rates, $m(i,t,x)$, for Danish males Groups 1 to 10 using the CBD-X model, but with sub-groupings based on income only and no lockdown at age 67.

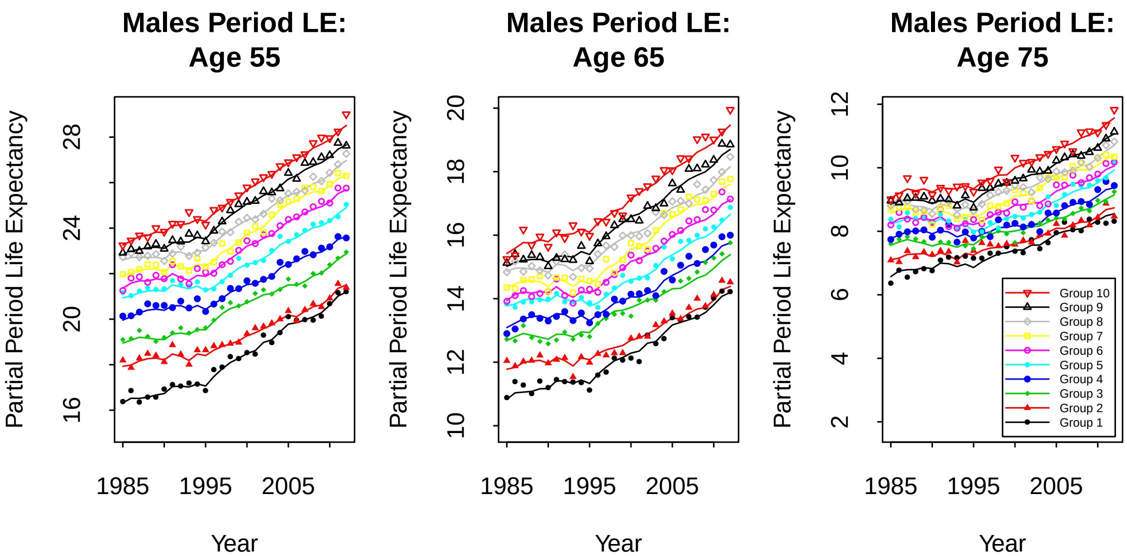

Figure 4: Partial period life expectancy (LE) for Danish males in Groups 1 to 10 for ages 55 (left) and 65 (middle) and 75 (right). Lines: LEs based on fitted death rates. Points: LEs based on crude death rates.

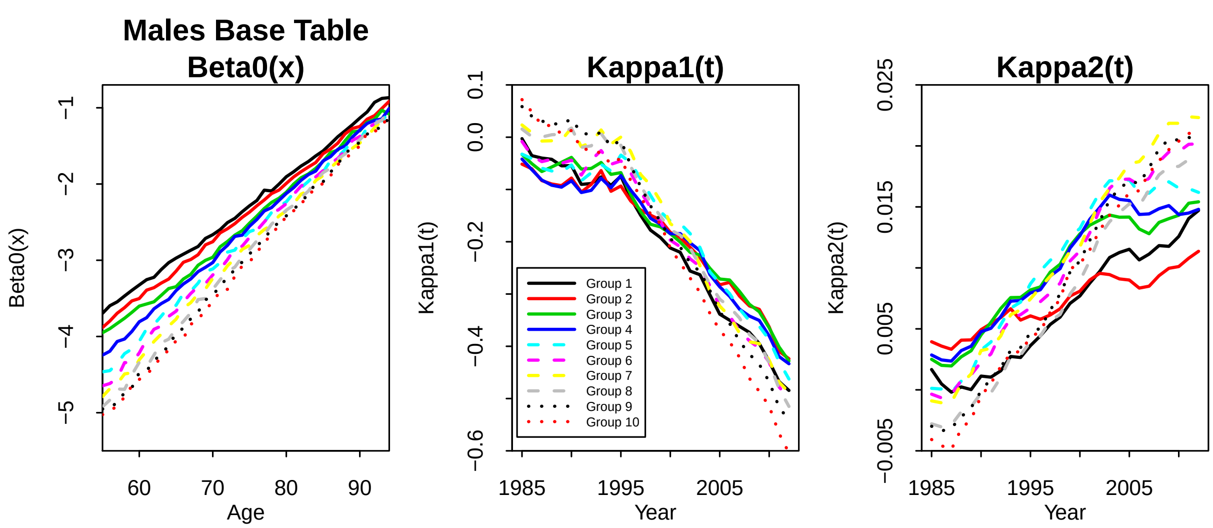

Figure 5: Estimated age and period effects for the 10 sub-populations.

- A very wide gap in death rates between the least and most affluent at younger ages, and a narrower gap at higher ages.
- Relative to 1985 death rates, improvements have been largest amongst the most affluent (that is a widening life expectancy gap) and at younger ages.

For comparison with previous grouping methods, equivalent plots are provided in Figure 3, where sub-groups are based on income only *with no* lock down at age 67. The step change in the age-80 plot in the fitted rates around 1995 for Group 1 that is evident in Figure 3 but not Figure 2 is, most-likely, the result of changes in the amounts and treatment of social assistance and old-age pensions in the previous year.[15] After 1995, the sub-group rankings are very poor especially after retirement. Similar plots have been analysed for income only with lockdown at 67, and the affluence index (wealth+15×income) with no lockdown. Out of all the numerous experiments that we conducted, only affluence *with* lockdown produces a consistent ranking across all years and all ages.

Figure 4 plots the development of partial period life expectancy (LE) over time for ages 55, 65 and 75 for each of groups 1 to 10. The dots show LEs derived from crude death rates, while the lines show LEs based on fitted death rates. We can see that the fitted LEs produce a smoother progression from year to year compared to the crude LEs, without losing the essential features of the crude LEs. More importantly, we see that the fitted process results in greater consistency from year to year between the 10 sub-groups, including improved separation. In line with earlier findings on the crude LEs, fitted LEs exhibit a wider spread at younger ages, and a slight widening of spreads between sub-groups over the period 1985 to 2012. These plots highlight, in more meaningful terms, the wide gap between the rich and poor, even in a country with a strong health care and social security system.

This can be compared with US data. Waldron (2013) considers deciles for fully-insured US males based on lifetime earnings. At age 65, Waldron finds a wider gap between Groups 1 and 10, but when one compares Groups 2 to 10, the differences between groups are quite similar to the results for Denmark: indeed the gap between Groups 2 and 10 is wider than in the US. One might tentatively conclude, therefore, that the more generous social security system in Denmark benefits primarily the least affluent 10%, in contrast to the US.

[15]The changes in the reporting of income in 1995 has resulted in high proportion of individuals in lower groups being reallocated to other groups, with group 1 losing many individuals in poor health and gaining many others (perhaps with substantial personal savings to top up their low reported income) in relatively good health.

Point estimates for fitted age and period effects – the underlying drivers of Figures 2 and 4 – are plotted in Figure 5. The base mortality tables (left-hand plot), represented by the $\beta_0^{(i)}(x)$, provide the basis for the rankings of the different sub-groups in individual years. The $\kappa_1^{(i)}(t)$ period effects (middle plot) drive changes in the overall level of mortality. The individual series all follow a consistent downwards trend (of generally improving mortality) that steepens after around about 1995.[16] [17] However, we can see that $\kappa_1^{(10)}(t)$ starts higher and falls more steeply than is the case with the other sub-groups: a feature that supports the earlier observation that the gap between the richest and poorest has been widening, potentially, at all ages. Changes in the $\kappa_2^{(i)}(t)$ (right-hand plot) indicate how the slopes of the individual mortality curves have been changing. These have tended to rise over time (by varying amounts) indicating that mortality has improved at a faster rate at lower ages than at higher ages. The biggest changes have been amongst Groups 6 to 10, with Groups 1 to 5 lagging in a variety of ways over different time periods. The interpretation of this is that the widening of the gap between the more and less affluent has been more pronounced at younger ages, as was observed in Figure 2.

4.2 Proposal for mortality modelling at very high ages

The narrowing gap in death rates up to age 94 observed in Figure 2 suggests a possible way to model *sub-group* mortality at very high ages. At the sub-group level, there is very little data to work with. To circumvent this, we propose that mortality is first modelled at very high ages at the level of the national population (although the amount of data is still small, it is not as small). Sub-group mortality can then be modelled as national mortality with modest increments or decrements at age 95, that gradually diminish with age and eventually vanish at very high ages. We leave this for future work.

5 Future Mortality and Survivorship

5.1 Projecting future death rates

We now turn to projecting future death rates. The results in this section, unless otherwise stated, include full parameter uncertainty. This means that we draw

[16]The kink around 1995 observable in Figures 2, 4 and Figure 5 that affects all groups is probably not connected to the social security and tax changes that occurred in 1994. Rather, 1995 looks like a year with randomly high mortality, making what is probably a more gradual increase in mortality improvement rates look more sudden, i.e., like a kink. We have not been able to find a good explanation for this change in improvement rates.

[17]The same kink can be observed in Figure 4 although the scale makes it slightly less obvious.

historical values for the latent state variables (the $\beta_0^{(i)}(x)$ and the $\kappa_j^{(i)}(t)$) and the process parameters (ρ, ψ etc.) at random from the MCMC output, and then use the final year's (2012) $\kappa_j^{(i)}(t)$ plus the selected process parameters to generate each stochastic projection scenario.

Fan charts for the forecast death rates for ages 65 and 75 are plotted in Figure 6. The relative positions of the fans stay fairly fixed over the period 1985 to 2012, although the fans gradually widen after 2013, reflecting the greater uncertainty in projected future rates.

The correlations between sub-group mortality will be of interest to pension plans and insurers as a key component of their overall risk assessment (see, e.g., Haberman et al., 2014). Correlations over varying time horizons are considered in Figure 7. They rise steadily the further into the future we look. Initially, the levels of the curves reflect the short-term contemporaneous correlations between the period effects. As the projection horizon lengthens, the shape reflects mean reversion towards the 'national' random walk (equations 3 and 4).

We can also see that Group 10 tends to have lower correlations than Group 5 Which, in turn, has lower correlations than Group 1. This is because Group 10 death rates are lower than Group 1 and so, in relative terms, contribute less than 10% of the risk to the national average.

The level and shape of the correlation curve depend on whether or not we include uncertainty in the underlying process parameters (see, e.g., Cairns, 2013, and Cairns et al., 2014). We investigate this in Figure 8 (left-hand plot) for Group 5 by way of illustration. This plot shows correlations under three experiments:

- Full Parameter Uncertainty (Full PU): full allowance for uncertainty in all process parameters and latent state variables in line with the posterior distribution.
- Partial PU#1: the drift parameters μ_1 and μ_2 are fixed at their posterior medians. All other elements of the posterior distribution remain random.
- Partial PU#2: the process parameters μ_1, μ_2, ρ, ψ and ν are fixed at their posterior medians.

From Figure 8, we see that the curves for Partial PU #1and #2 are almost indistinguishable indicating that uncertainty in ρ, ψ and ν has little impact on the empirical correlations. For each parameter, uncertainty around its median can push the correlation up or down depending on whether the deviation from the median is positive or negative. By contrast, moving from Partial PU#1 to Full PU results in a big change. Uncertainty in μ_1 and μ_2 pushes up uncertainty in $\bar{\kappa}_1(t)$ and $\bar{\kappa}_2(t)$, with a corresponding impact on uncertainty in sub-group death rates. But, since

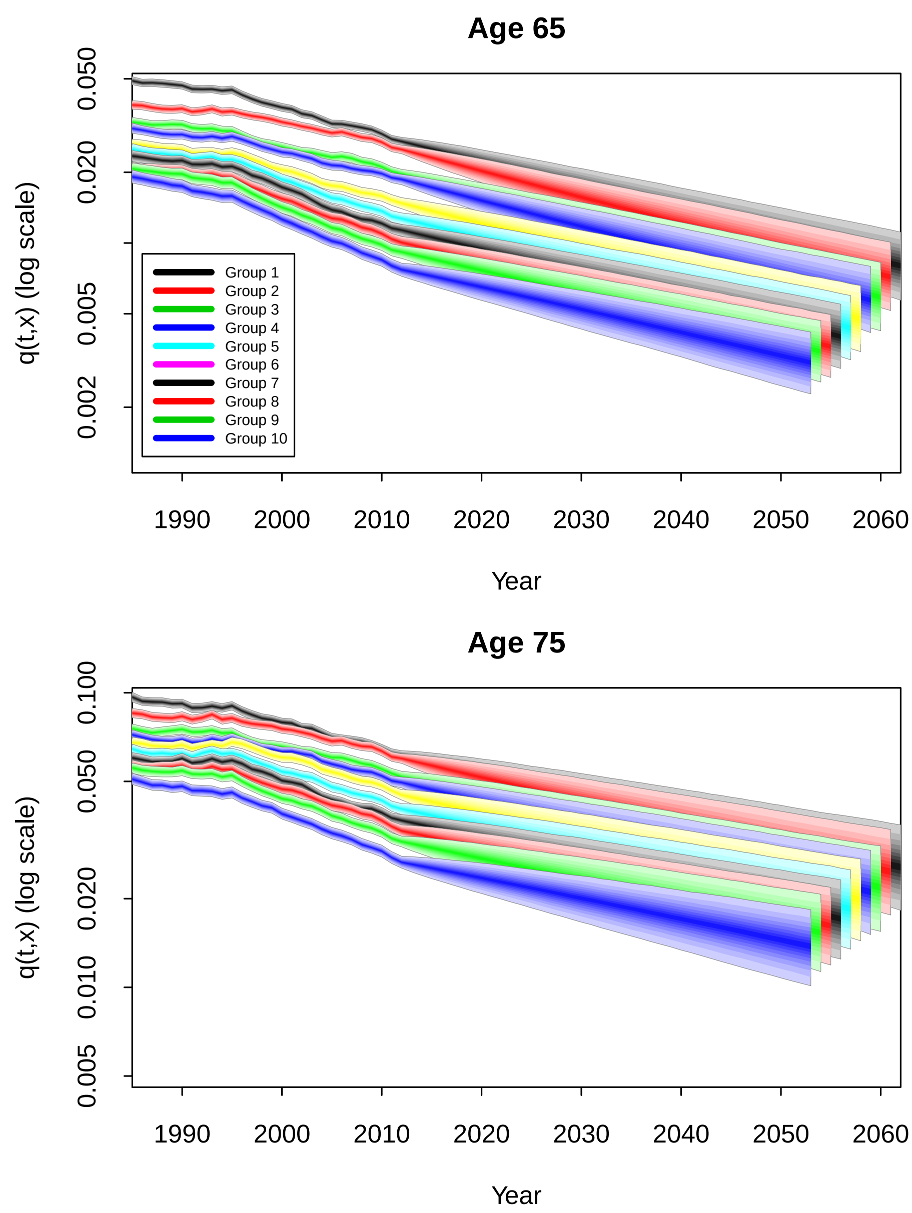

Figure 6: Fan charts for mortality at ages 65 and 75 for Danish males Groups 1 to 10. The charts show parameter uncertainty in the fitted death rates up to 2012, and combined parameter uncertainty and process risk from 2013.

each sub-group has a common dependency on $\bar{\kappa}_1(t)$ and $\bar{\kappa}_2(t)$, correlations rise, in line with the results in Cairns (2013) and Cairns et al. (2014).

Returning to Figure 7, we can also consider correlations between two stylised pension plans and the national population. The first ("white-collar") pension plan is assumed to be made up of equal numbers of Groups 8, 9 and 10: the high earners. The second ("blue-collar") pension plan is made up of equal numbers of Groups 2, 3 and 4. We exclude Group 1 from the analysis as it potentially includes unemployed or people who are in poor health (and not in employment). Both plans have much higher correlations with the national population than any of Groups 1 to 10 separately, reflecting the fact that some of the idiosyncratic risk in each of the three contributing sub-groups has been diversified. We also see that the blue-collar plan has higher correlations than the white-collar plan, for the same reason that Group 1 had higher correlations than Group 10 above.

It is noteworthy that the correlation term structure for the white-collar plan is similar to that for UK assured lives both in terms of level and shape (see the UK Continuous Mortality Investigation of Assured Lives versus England & Wales males examined in Dowd et al, 2011, Figure 13).

5.2 Sensitivity to the choice of prior distribution

In Figure 8 (right), we pick out Group 5, by way of example, and investigate how sensitive the correlation term structure is to changes in the prior distributions for ρ and ψ. Each has either a Beta(2, 2) or Beta(3, 3) prior distribution as outlined in the legend. In each case, although differences can be seen, the three sets of priors produce very similar results in each of the plots.[18] Although this is a limited experiment, it does suggest that our estimates of the correlations over a range of time horizons are robust relative to the choice of prior.

5.3 Survivor indexes

As an alternative to death rates, we can also look at cohort survivorship. A general survivor index, $S(t, x)$, represents the probability that an individual aged exactly x at time 0 (the beginning of calendar year 1) survives for t years to age $x + t$, given the knowledge of how the underlying mortality rates, $q(t, y)$, evolve from time 0 to t.[19] Thus

$$S(t, x) = \prod_{u=1}^{t} \left(1 - q(u, x + u - 1)\right).$$

[18] These differences are small in comparison with the case that allows for the inclusion of parameter uncertainty in μ_1 and μ_2 – see Figure 8 (left).

[19] We use the approximation $q(i, t, x) = 1 - \exp\left(-m(i, t, x)\right)$.

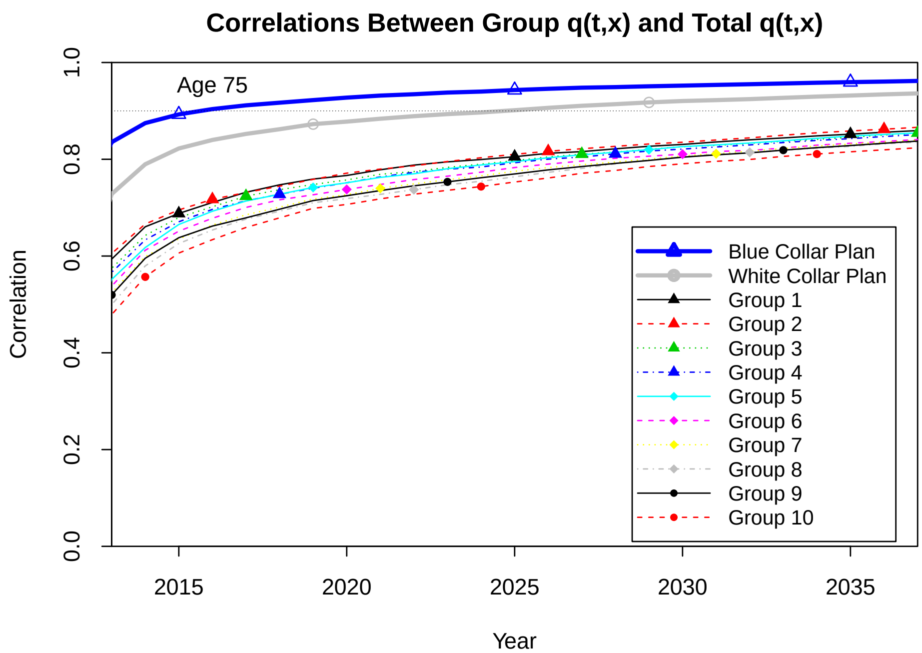

Figure 7: Empirical correlations between individual sub-groups, two pension plans and the total population, $\mathrm{cor}\left(q(i,t,x), \bar{q}(t,x)\right)$, for age 75. Simulations incorporate full parameter uncertainty. The dotted line at a correlation of 0.9 is for reference only.

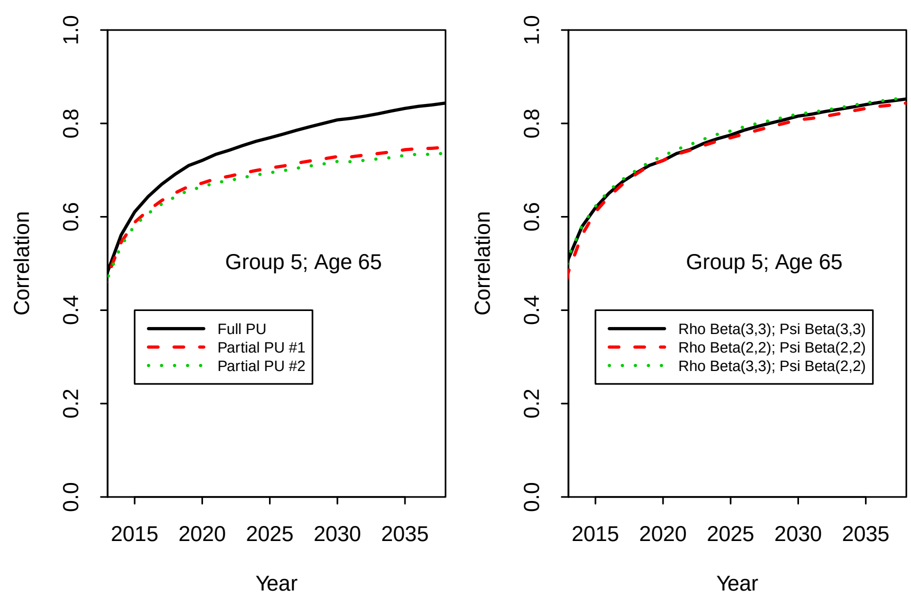

Figure 8: Empirical correlations between Group 5 and the total population, $\mathrm{cor}\left(q(i,t,x), \bar{q}(t,x)\right)$, for age 65, showing sensitivity to changes in inputs and assumptions. Left: impact of different levels of parameter uncertainty: Full parameter uncertainty (PU); PPU #1 has μ_1 and μ_2 fixed at their posterior medians; PPU #2 has μ_1, μ_2, ρ, ψ and ν fixed. Other elements of the posterior distribution remain random. Right: the three lines show sample correlations under three different combinations of prior distributions for ρ and ψ.

Correspondingly, we have survivor indexes, $S_1(t,x), \ldots, S_{10}(t,x)$, for each of Groups 1 to 10. Now suppose that a pension plan ("X") consists of a mixture of Groups 1 to 10 with weights $w_1, \ldots, w_{10}$ (with $\sum_i w_i = 1$; different ages might have different weights). Special cases of X include the stylised white- and blue-collar plans. The plan X then has its own cohort survivor index $S_X(t,x) = \sum_{i=1}^{10} w_i S_i(t,x)$. Lastly, the total population from age $x = 67$ has survivor index $S_{TOT}(t,x) = \sum_{i=1}^{10} S_i(t,x)/10$.

We consider the correlation between survivor indices for individual sub-groups, pension plans and the total population from two perspectives in Figure 9. The plots include a third stylised plan ("Mixed", "M") that has weights proportional to the vector $(0,0,1,2,3,4,5,6,7,8)$, which might reflect either growing numbers of individuals in more wealthy sub-groups, or more equal numbers with the weights reflecting the growing amounts of pensions. Figure 9, left, looks at the effect of the time horizon, and we can see correlation curves that mimic the shape of those in Figure 7. Unlike the death rates, the survival index correlations depend on multiple death rates from prior years. Additionally, we see Groups 2 and 9 cross over around 2027. Initially, the less affluent sub-groups contribute more to the uncertainty in $S_{TOT}(t,x)$. In later years, however, the less affluent sub-groups will have died off much more quickly, so that they contribute less to $S_{TOT}(t,x)$, while, e.g., Group 9 contributes relatively more.

There is, however, a much more general 'term-structure' of correlation,

$$\mathrm{cor}\left(S_i(t_i,x_i), S_j(t_j,x_j)\right),$$

for any two populations i and j and potentially different time horizons and ages. In some applications, it is important that this term structure takes a reasonable or plausible form. By way of example, we take, again, $i = X = 2, 9, B, W, M$ and $j = TOT$ with $t_i = t_j = 10$. The right-hand plot in Figure 9 looks at how correlations change as we vary the ages in the two populations. Specifically, we keep the initial age (x_j) for the total population fixed at 67, and calculate the time-10 correlation with different plans, X, over a range of ages (x)

$$\mathrm{cor}\left(S_X(10,x), S_{TOT}(10,67)\right).$$

We see that choosing matching ages makes a big difference in the correlations: cohorts in the two populations that are far apart in terms of age are less strongly correlated.

These correlation plots help us identify a number of desirable criteria from the perspective of biological reasonableness (see, e.g., Cairns et al., 2009, and Haberman et al., 2014). Specifically, a multi-population mortality model should ideally satisfy the following:

- Correlations between mortality rates and survivor indexes for different populations should vary smoothly with the time horizon and should be increasing

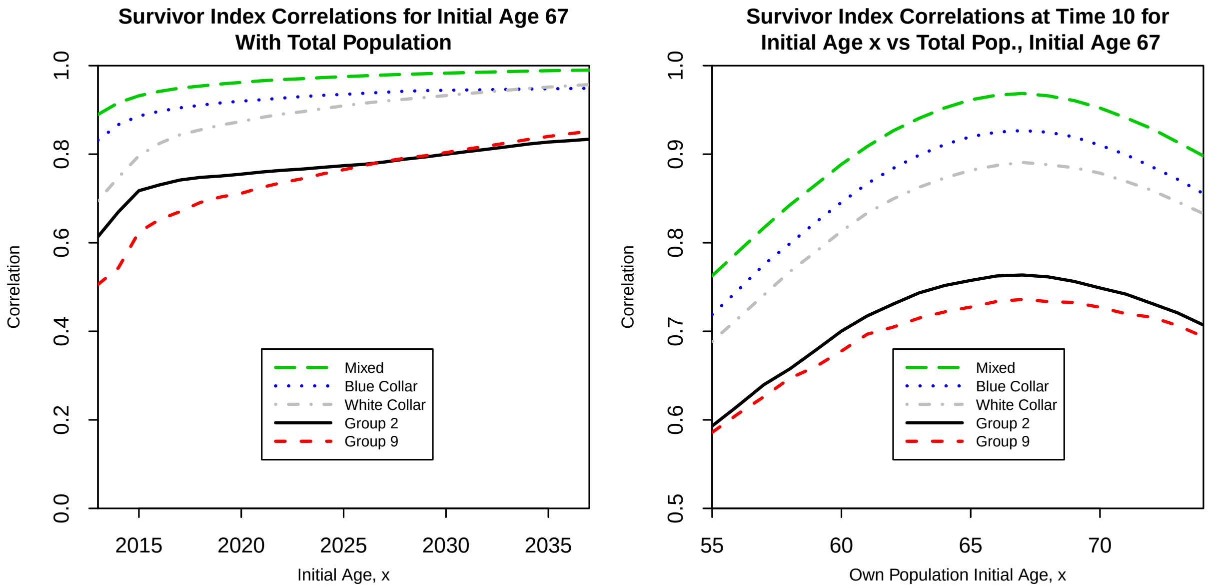

Figure 9: Left: Empirical correlations between $S_X(t, 67)$ and $S_{TOT}(t, 67)$ for $t = 2013, \ldots, 2037$. Right: Correlations between $S_X(10, x)$ and $S_{TOT}(10, 67)$ for $x = 55, \ldots, 74$. In both plots, results for Groups $X = 2$, 9, blue collar, white collar and mixed pension plans are plotted.

with the time horizon if it is felt that mean reversion between populations is itself a desirable element of the model, and correlations should not be exactly equal to 1 without good reason.

- For a fixed time horizon, correlations between populations should vary smoothly with the reference ages of both populations, and correlations should not be exactly equal to 1 without good reason.

It is difficult to define what the boundary is between what would be a reasonable and unreasonable plot, and, as in some previous cases (e.g., Cairns et al., 2009), each plot of its type for each model needs to be considered on its own merits. From time to time, a model produces a plot that is clearly unreasonable for reasons that can only be inferred from the plot itself rather than anticipated in advance.[20]

[20]For example, consider the Li and Lee (2005) model: $\log m(i, t, x) = \alpha_i(x) + B(x)K(t) + \beta_i(x)\kappa_i(t)$. For different populations i and j it is possible to have $\beta_i(x_i) = \beta_j(x_j) = 0$ for significantly different ages x_i and x_j. We then have, for all t, $\text{cor}\left(\log m(i, t, x_i), \log m(j, t, x_j)\right) = 1$. This is not plausible or realistic when correlations are (significantly) less than 1 at other ages.

6 Extensions and further work

This study has focused on older male mortality in Denmark. For females, we obtain similar results with the exception that the affluence index is less effective at producing the anticipated ranking amongst Groups 1, 2 and 3. Group 1, in particular, seems to have rather lower mortality than Groups 2 and 3, suggesting that reported levels of income and wealth do not truly reflect the affluence of the females in Group 1. Further work needs to be done, although we can report that good rankings can be observed amongst Groups 4 to 10.

We have not attempted to explore the many other covariates and pieces of information within the SD database. The work could, therefore, be extended in a number of ways:

- We can look at the explanatory power of including other covariates, such as marital status, education and area of residence: education, in particular, is known to have strong explanatory power (see, e.g., Olshansky et al., 2012, and Sasson, 2016). We need to think carefully about sample sizes to allow us to elicit statistically significant results. We also need to be mindful of the fact that levels of educational attainment have been rising steadily over time (see, e.g., Brønnum-Hansen and Baadsgaard, 2012, and Sasson, 2012), so the impact of having a tertiary education might be different in an older cohort than a younger cohort where its occurrence will be more common.
- We can look at how individuals migrate between different affluence sub-groups over different periods of time. From an insurance perspective, it is of most interest to determine the probabilities of ending up in each of Groups 1 to 10 at age 67.
- We can consider cause of death data (e.g. Arnold-Gaille and Sherris, 2016) and investigate whether affluence, as a covariate, has a greater impact on some causes of death.

There are potentially many other "big data" analyses that could be conducted. However, many of these might be very specific to the detailed nature of the SD database, making it difficult to apply the results in other contexts (e.g., using the more limited data available to an annuity provider).

7 Conclusions

Understanding socio-economic differences in mortality is important both to policy-makers planning and projecting state pension budgets and to private sector providers of mortality-linked products, such as pensions and annuities.

We have been able to explain mortality differences in older Danish males at a much finer level of granularity than hitherto attempted, namely at the decile level, using just a single index to allocate males to a decile sub-group. The index, which we call the *affluence index*, is based on income and wealth data available on the Statistics Denmark database. The male population aged 55 to 94 in years 1985 to 2012 was subdivided into 10 sub-groups, based on the relative value of their affluence index (measured as wealth+15×income). Prior to age 67, they would be allocated to a particular subgroup annually, based on the value of the index in the previous year. Once they reached 67, they would be allocated to a subgroup and remain in that sub-group for the remainder of their lives, a procedure we called *lockdown*. Lockdown at a particular age (we found that age 67, which also happened to equal the state pension age in Denmark for most of the period under investigation, worked well) combined with the affluence index was found to be critical for ensuring a consistent ranking of sub-group death rates across all years and all ages.

We also introduced a flexible multi-population stochastic gravity-type mortality model for both fitting death rates and forecasting future death rates. The structure of the model combined with the gravity effect links group-specific mortality improvements to the national trend. The model allows for some flexibility in the relationship between the 10 sub-groups, but also preserves the sub-group rankings over different time horizons.

Model-based smoothing was employed to filter out sampling variation in the underlying crude death rates, and we used these smoothed death rates to rank the 10 sub-groups at each age and in each year. We were able to do this without losing any of the essential patterns and characteristics underpinning the crude death rates and without the need for a cohort effect. Sub-group rankings were again consistent and clear across all ages and years, with a very wide gap between the most and least affluent at young ages, narrowing significantly with age, but widening slightly over the period 1985 to 2012.

We also proposed a way to model sub-group mortality at very high ages, given that there is very little data to work with at these ages. Mortality is first modelled at the national population level, with sub-group mortality then modelled as national mortality with small increments or decrements at age 95, which subsequently decline and ultimately vanish as age increases above 95.

Another key element of the paper was an analysis of how correlations between sub-groups and the national population change as the forecasting horizon lengthens. We looked at both forecast death rates and survival rates. Correlations were found to start at moderate levels 1 year ahead (in the range 0.5-0.6) and climb quickly to very high levels (over 0.8), especially for populations that comprise a mixture of individuals from several of the 10 sub-groups modelled. Amongst other things, knowledge of this term structure of correlation is important in some financial applications where risk management strategies will be more effective if correlations are higher.

Our paper provides important general lessons for researchers with other datasets who are interested in modelling socio-economic differences in mortality at high ages at a fine level of granularity:

- It is possible to generate clear and consistent rankings of death rates at all ages down to at least the decile level using just a single covariate to allocate individuals to a particular decile sub-group.
- That covariate is likely to be some measure of the relative affluence of the individuals in the dataset and will involve some combination of the wealth and income available to the individual – this is intuitively appealing and is obviously preferable to looking at income alone which most previous studies have concentrated on.
- It is likely that, in order to preserve the rankings across ages and over time (including future projections), there will need to be a lockdown at a certain age – in other words, we found that individuals could switch between decile sub-groups prior to the lockdown age without violating the sub-group rankings, but they needed to be locked in to a particular sub-group once they reached a certain age in order to preserve the sub-group rankings at higher ages.
- The age at which lockdown happens might well be related to the state pension age or the age at which individuals retire – this also makes intuitive sense: individuals have much more flexibility to change their labour market behaviour (and hence their relative affluence) before retirement than after.
- It is possible to smooth sub-group death rates by fitting them to a multi-population stochastic gravity-type mortality model, with the gravity parameter helping to preserve the sub-group rankings – this has the benefit of reducing the effect of idiosyncratic mortality risk in small populations.
- The multi-population mortality model can then be used to make mortality projections that preserve the subgroup rankings.
- It can also be used to model sub-group mortality at very high ages, where data are likely to be sparse, by first modelling at the national population level, with sub-group mortality then modelled as increments or decrements to the national mortality which decline and ultimately vanish with increasing age.
- Correlations between sub-groups and the national population rise with the time horizon and are especially high for sub-groups that contain a diverse mixture of socio-economic groups.

Acknowledgements:

The authors wish to acknowledge the numerous discussants at conference presentations of this work. MKL and CPTR gratefully acknowledge financial support from the Danish Council for Independent Research, Social Sciences, grants: 11-105548/FSE (M. Kallestrup-Lamb) and CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. AJGC acknowledges financial support from Netspar under project LMVP 2012.03.

References

Arnold-Gaille, S., and Sherris, M. (2016) International cause-specific mortality rates: New insights from a co-integration analysis. *ASTIN Bulletin, 46*, 9-38.

Avraam, D., Arnold, S., Jones, D., and Vasiev B. (2014). Time-evolution of age-dependent mortality patterns in mathematical model of heterogeneous human population. *Experimental Gerontology, 60*, 18-30.

Börger, M., Fleischer, D., and Kuksin, N. (2013). Modeling mortality trend under modern solvency regimes. *ASTIN Bulletin, 44*, 1-38.

Bosworth, B.P., and Burke, K. (2014). *Differential Mortality and Retirement Benefits in the Health and Retirement Study*. Washington, D.C.: Brookings Institution.

Brønnum-Hansen, H., Baadsgaard, M. (2012). Widening social inequality in life expectancy in Denmark. A register-based study on social composition and mortality trends for the Danish population. *BMC Public Health, 12*, 994.

Cairns, A.J.G. (2013). Robust hedging of longevity risk. *Journal of Risk and Insurance, 80*, 621-648.

Cairns, A.J.G., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance, 73*, 687-718.

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

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., and Khalaf-Allah, M. (2011) Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin, 41*, 29-59.

Cairns, A.J.G., Blake, D., Dowd, K., and Coughlan, G.D. (2014). Longevity hedge effectiveness: A decomposition. *Quantitative Finance, 14*, 217-235.

Cairns, A.J.G., Blake, D., Dowd, K., and Kessler, A. (2016). Phantoms never die: Living with unreliable population data. To appear in *Journal of the Royal Statistical Society, Series A*.

Chapman, K.S., and Hariharan, G. (1996). Do poor people have a stronger relationship between income and mortality than the rich? Implications of panel data for health analysis. *Journal of Risk and Uncertainty, 12*, 51-63.

Christiansen, M.C., Spodarev, E., and Unseld, V. (2015) Differences in European mortality rates: A geometric approach on the age-period plane. *ASTIN Bulletin, 45*, 477-502.

Citro, C. F., and Michael, R. T. (1995). *Measuring Poverty: A New Approach.* Washington, DC: National Academy Press.

Coughlan, G.D., Khalaf-Allah, M., Ye, Y., Kumar, S., Cairns, A.J.G., Blake, D. and Dowd, K., (2011). Longevity hedging 101: A framework for longevity basis risk analysis and hedge effectiveness. *North American Actuarial Journal, 15*, 150-176.

Cristia, J.P. (2009). Rising mortality and life expectancy differentials by lifetime earnings in the United States. *Journal of Health Economics, 28*,984-995.

Demakakos, P., Biddulph, J. P., Bobak. M., and Marmot, M. G. (2015). Wealth and mortality at older ages: A prospective cohort study. To appear in *J. Epidemiol. Community Health*.

Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., and Khalaf-Allah, M. (2011). A gravity model of mortality rates for two related populations, *North American Actuarial Journal, 15*, 334-356.

Enchev, V., Cairns, A.J.G., and and Kleinow, T. (2015). Multi-population mortality models: Fitting, forecasting and comparisons. To appear in *Scandinavian Actuarial Journal*.

Gavrilov, L.A., and Gavrilova, N.S., (1979). Determination of species length of life. *Dokl. Akad. Nauk SSSR Biol. Sci., 246*, 905-908. Gavrilov, L.A., and Gavrilova, N.S. (1991). *The Biology of Life Span: A Quantitative Approach.* New York, N.Y.: Harwood Academic Publisher.

Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Haberman, S., Kaishev, V., Millossovich, P., Villegas, A., Baxter, S., Gaches, A., Gunnlaugsson, and Sison, M. (2014). *Longevity Basis Risk: A Methodology for Assessing Basis Risk*. Sessional research paper, Institute and Faculty of Actuaries, 8 December 2014. Available online at www.actuaries.org.uk/events/pages/sessional-research-programme.

Hunt, A., and Blake, D. (2014). A general procedure for constructing mortality models. *North American Actuarial Journal, 18*, 116-138.

Jarner, S.F., and Kryger, E.M. (2011). Modelling adult mortality in small populations: The SAINT Model. *ASTIN Bulletin, 41*, 377-418.

Kleinow, T., and Cairns A.J.G. (2013). Mortality and smoking prevalence: An empirical investigation in ten developed countries. *British Actuarial Journal, 18*, 452-466.

Kleinow, T. (2015). A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics, 63*, 147-152.

Lee, R.D., and Carter, L.R. (1992). Modeling and forecasting U.S. mortality, *Journal of the American Statistical Association, 87*, 659-675.

Li, J.S.-H., and Hardy, M.R. (2011). Measuring basis risk in longevity hedges, *North American Actuarial Journal, 15*, 177-200.

Li, N., and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography, 42*, 575-594.

Li, J.S.H., Zhou, R., Hardy, M.R. (2015). A step-by-step guide to building two-population stochastic mortality models. *Insurance: Mathematics and Economics, 63*, 121-134.

Mackenbach, J.P., Bos, V., Andersen, O., Cardano, M., Costa, G., Harding, S., Reid, A., Hemström, Ö., Valkonen, T., and Kunst, A.E. (2003). Widening socioeconomic inequalities in mortality in six Western European countries. *International Journal of Epidemiology, 32*, 830-837.

McDonough, P., Duncan, G. J., Williams, D., and House, J. (1997). Income dynamics and adult mortality in the United States, 1972 through 1989. *American Journal of Public Health, 87(9)*, 1476-1483.

Mavros, G., Cairns, A.J.G., Kleinow, T., and Streftaris, G. (2014). A parsimonious approach to stochastic mortality modelling with dependent residuals. Working paper. Edinburgh: Heriot-Watt University.

Michaelson, A., and Mulholland, J. (2015). Strategy for increasing the global capacity for longevity risk transfer: Developing transactions that attract capital markets investors. *Pension and Longevity Risk Transfer for Institutional Investors*, Fall, 28-37.

Office for National Statistics (2011). Trends in life expectancy by the National Statistics Socio-economic Classification 1982-2006. *Statistical Bulletin*, 22 February 2011.

Office for National Statistics (2013). Trends in mortality by NS-SEC at older ages in England and Wales, 1982-86 to 2002-06. *Statistical Bulletin*, 22 February 2013.

Olshansky, S.J., Antonucci, T., Berkman, L., Binstock, R.H., Fried, A., Goldman, D.G., Jackson, J., Kohli, M., Rother, J., and Zheng Y. (2012). Differences in life expectancy due to race and educational differences are widening, and many may not catch up. *Health Affairs, 31(8)*, 1803-1813.

Plat, R. (2009). On stochastic mortality modelling. *Insurance: Mathematics and Economics, 45*, 393-404.

Rogot, E., Sorlie, P.D., Johnson, N.J., and Schnitt, C. (1992). *A Mortality Study of 1.3 million Persons by Demographic, Social, And Economic Factors: 1979-1985.* Bethesda, MD: NIH.

Sabel, C.E., Dorling, D., and Hiscock, R. (2007). Sources of income, wealth and the length of life: An individual level study of mortality. *Critical Public Health, 17(4)*, 293-310.

Sasson, I. (2016) Trends in life expectancy and lifespan variation by educational attainment: United States, 1990-2010. *Demography, 53,* 269-293.

Tarkianen, L., Martikainen, P., Laaksonen, M., and Valkonen, T. (2012). Trends in life expectancy by income from 1988 to 2007: Decomposition by age and cause of death. *Journal of Epidemiology and Community Health, 66*, 573-578.

Villegas, A., and Haberman, S. (2014). On the modeling and forecasting of socioeconomic mortality differentials: An application to deprivation and mortality in England. *North American Actuarial Journal, 18*, 168-193.

Waldron, H. (2007). Trends in mortality differentials and life expectancy for male social security–covered workers, by socioeconomic status. *Social Security Bulletin, 67(3)*, 1-28.

Waldron, H. (2013). Mortality differentials by lifetime earnings decile: Implications for evaluations of proposed social security law changes. *Social Security Bulletin, 73(1)*, 1-37.

Wolfson, M., Rowe, G., Gentleman, J.F., and Tomiak, M. (1993). Career earnings and death: A longitudinal analysis of older Canadian men. *Journal of Gerontology, 48*, S167-79.

Yang, S.S., Huang, H.C., and Jung, J.K. (2014). Optimal longevity hedging strategy for insurance companies considering basis risk. Presented at *The Tenth International Longevity Risk and Capital Markets Solutions Conference*, Chile, September 2014.

First version: August 2014
This version: April 18, 2016