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 subgroups using a new affluence index that combines wealth and income reported on the Statistics Denmark national register database. This is shown to provide consistent subgroup rankings based on age specific death rates across all ages and over all years. Its use also improves significantly on previous studies that have focused on the impact of either education of income on life expectancy or on age-standardised mortality rates. 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 captures the essential character of the raw data. The model generates smoothed death rates that allow us to work with a larger number of smaller subgroups than might be considered feasible when working with raw data.

The model produces plausible projections of age-specific death rates that preserve the subgroup 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.

JEL codes: J11, C53, G22

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 (MKL) and CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. AJGC acknowledges financial support from the Actuarial Research Centre of the Institute and Faculty of Actuaries, and from Netspar under project LMVP

[a]Joint corresponding author: Actuarial Research Centre, Institute and Faculty of Actuaries; Maxwell Institute for Mathematical Sciences, Edinburgh; and Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, EH14 4AS, UK. 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

2012.03. Finally, the authors wish to thank the two referees of the paper for their helpful and thought provoking comments.

1 Introduction

We are interested in explaining and projecting mortality in closely related populations. To do this, 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 lead to implausible projections especially in the medium and long term. For example, it is possible for projected 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 issues 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., 2016, Christiansen et al., 2015); males versus females; smokers vs non-smokers (e.g., Kleinow and Cairns, 2013); groups of annuitants vs those who hold life insurance policies (e.g., Yang et al., 2014); a specific pension plan's own mortality vs that of the national population (e.g., Cairns et al., 2011, Haberman et al., 2014, Hunt and Blake, 2017); and different socio-economic subgroups (e.g., Li et al., 2015).

In some of these cases, especially those that involve subpopulations 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 projections of the population of interest. In other settings, we might seek to manage actively the risks that we identify. This point is most apparent in the life insurance and pensions worlds where 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, Michaelson and Mulholland, 2015, and Cairns and El Boukfaoui, 2017).

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 allows us:

- To construct a new affluence index for subdividing the Danish male population into ten socio-economic subgroups of equal size.
- To consider the potential advantages of *lockdown*. Individuals potentially transfer between subgroups each year prior to the official state pension age

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

of 67 depending on their affluence, with lockdown thereafter: that is, each individual remains in the same subgroup to which they were allocated in their 67th year.

- To use the proposed index in combination with an appropriate stochastic mortality model to subdivide the male population into a larger number (10) of smaller subgroups than has previously been attempted for Danish males – using income, one component of our affluence index – while still preserving statistical significance.

- To demonstrate that the affluence index plus lockdown in their 67th year provides a much improved separation of the 10 deciles *at all ages* from 55 to 94. 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 subgroups and in all years.

- To demonstrate that the use of the affluence index reveals a greater degree of mortality inequality within the Danish males' population than is suggested by education: a commonly used alternative covariate.

- To develop a new stochastic multi-population gravity model for making mortality projections. The model fits the historical data well and produces, for the 10 subgroups, coherent projections of future death rates (in the sense of Hyndman et al., 2013) that are both biologically reasonable[2] and socio-economically reasonable over all ages and a range of time horizons.

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

- To consider a range of useful applications. For example, 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 estimates 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 outline of the paper is as follows. Section 2 introduces and explains the Danish national mortality dataset. It also discusses the approach used to process the data using the 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 confirms that our approach to modelling fits the data well, smoothing out the effects of sampling variation (also

[2]As introduced in Cairns et al. (2006).

known as Poisson risk) while still preserving the essential characteristics of the crude subgroup mortality data. Section 5 analyses properties of projected mortality: central trends, uncertainty and correlations between subgroups. Section 6 offers some suggestions for further research, and Section 7 summarises the results.

We are aware that historical trends in Danish mortality have their own idiosyncratic features (as identified, for example, by Juel, 2008, Glei et al., 2010, and Christiansen et al., 2010, in Crimmins et al., 2010). These features are evident in the current data (e.g., the trend before and after 1995 in Figure 1 below). In anticipation of the analysis, we note that the stochastic model considered in Section 3 captures these features well across all ages, years and affluence groups. The stochastic model's application to equivalent datasets for other countries would, obviously, require careful model revalidation.

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 (CPR) number either at birth or when given residence permission in the country, we are able to uniquely identify each individual across all components 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, wealth, and, ultimately, their date 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 within households. On an annual basis, we observe their marital status based on the following four categories: unmarried, married, divorced, and widow/widower. 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.

We choose at the outset to focus on measures of the financial status of each individual, since income and wealth are well known predictors of mortality. However, for comparison, we also discuss in Section 2.3 an individual's attained level of education as an alternative covariate.

The database allows us to identify three financial indicators for each individual and couple, all deflated to 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 CPR. 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, payments deducted 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.[3] For retired individuals, we observe a break in the gross annual income variable from 1994 onwards. There are several reasons for this break. The level of old age pension benefits was increased in 1994 to compensate for the government removal of a special tax rebate previously given to retirees. Moreover, individuals living in retirement homes prior to 1994 were only given a monthly allowance, but no amount to cover rent, heating, electricity, etc. as these expenses were paid for by the government. 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).

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 no longer being reported in the wealth measure.

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 subgroups, requires data for the previous calendar year, we calculate exposures and deaths data for each subgroup for the 1985-2012 period only.

One of the key objectives is to identify covariates (predictive variables) derived from information available on the SD database that have a strong predictive power in explaining the mortality of different individuals in the population. More specifically, we aim to subdivide the population at each age and in each year into a number, N, of approximately equal-sized subgroups, with a clear and unchanging ordering between subgroups in terms of mortality rates *at all ages*: that is, group 1 should

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

have the highest mortality at all ages and group N the lowest. From the outset, we aimed for a larger number of subgroups than has typically been the case in other studies,[4] settling on $N = 10$ deciles, giving us the flexibility to aggregate subgroups should this be desirable or necessary in the modelling work.

For a subsample of the US population, Waldron (2013) is able to achieve statistically meaningful results using deciles based on lifetime earnings covering a more limited range of ages (63-71) than in this paper (55-94). For Denmark, we show that separation into deciles also gives meaningful results, but requires more work to filter out or smooth the effects of sampling variation. We also found that the use of deciles revealed weaknesses in some previously considered covariates that would not be so readily apparent if we were to use, say, $N = 3$ or 4 subgroups. Extensive experimentation helped not only to identify these weaknesses (see, e.g., Figure 3 below), but also led us to identify a new and, in our view, superior covariate, namely affluence with lockdown at a certain age (67 in the case of Danish data). Lastly, greater levels of inequality are revealed using deciles.

In this particular study, we aim to identify a *single* financial covariate that is a strong predictor of mortality. Our reasons for choosing to focus on financial variables from the outset are twofold. First, income or wealth are well known as strong 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, Brønnum-Hansen and Baadsgaard, 2012, and Demakakos et al., 2015). Second, we sought to develop a modelling framework that could be applied to a variety of other countries and subpopulations; and some measure of income and/or wealth is often available, even when other types of data that are held by Statistics Denmark (e.g., education) are not.

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 these individuals' true financial positions. Ultimately, all 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 separately to be strong predictors of mortality: higher wealth and income resulted in lower 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 and unchanging mortality rankings across age for the higher (more wealthy) subgroups, while lower subgroups 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,

[4]For example, Brønnum-Hansen and Baadsgaard (2012) subdivide the Danish population into $N = 4$ income-based quartiles.

but (b) low income or low wealth 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 is, in many cases, negative for working age males; this finding reflects the fact that while liabilities include loans and mortgages, not all of the matching assets are included (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 subgroups

Following a great deal of experimentation, we settled on the following algorithm for allocating older Danish males to one of the 10 subgroups, based on a single composite financial indicator variable that we label as the *affluence index*. This index is obtained specifically from the income and wealth data available for each individual for each year. Allocation to subgroups 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:

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

where $W(i,t-1,x-1)$ is the (physical and financial) net 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 subgroups. We have chosen to use $K = 15$, but the derived subgroup death rates outlined below are found to be robust relative to the value of K.[5] Representative values for the affluence index are given in Table 1 for ages 55 and 65 in 1990 and 2010. The values given are for the boundaries between adjacent groups. Lockdown from age 67 means that the boundaries become blurred as some individuals' affluence levels drift up and some down without them being reallocated to another subgroup. However, from age 67, the relationship between *mean* levels of affluence between the groups remains relatively stable.

The second step is then to allocate all individuals to specific subgroups. Based on the affluence index, $A(i,t,x)$, all individuals, i, in year t at age x are ranked

[5]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.

	Year: 1990		Year: 2010	
Group Boundary	Age 55	Age 65	Age 55	Age 65
$1 \to 2$	2324	1803	2521	2026
$2 \to 3$	3141	2184	3492	2451
$3 \to 4$	3588	2518	4056	2898
$4 \to 5$	3970	2842	4542	3354
$5 \to 6$	4361	3218	5038	3866
$6 \to 7$	4836	3655	5568	4474
$7 \to 8$	5428	4219	6245	5224
$8 \to 9$	6335	5094	7261	6317
$9 \to 10$	8139	6855	9421	8356

Table 1: Boundaries (affluence, in '000's of Danish krone) between the ten affluence groups for ages 55 and 65 and in years 1990 and 2010; e.g. individuals aged 65 on 1 January 2010 will be assigned to affluence group 3 if their affluence for 2009 lies between 2,451,000 DKK and 2,898,000 DKK. Amounts are rebased to year 2000 values. 1000DKK≈150USD (18/9/2016).

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% of the total population are assigned to each subgroup.

The ranking and allocation is repeated every year and age until an individual reaches their 67th year, the main state pension age for most of the period.[6] From age 67 onwards, each individual is assumed to remain in the same subgroup that they were allocated to at age 66 (lockdown). Thus, individuals are assumed to be able to migrate between affluence subgroups before age 66, but not after. From these population subgroups, it is then possible to calculate the demographic measures exposed to risk, death counts, and death rates.

There are two key contributions of this algorithm compared to earlier ones. First, we found that the use of the affluence index is a more effective covariate *across all ages (from 55 to 94)* and in all years than income or wealth on their own. Second, the lockdown at age 67 produces much better results (including improved separation between the 10 subgroups) than not locking down.[7]

[6] 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.

[7] An additional advantage of the lockdown is that it restricts the potential for Group 1 to fill up after age 67 with individuals in severely declining health who have used up most of their personal savings on long-term care. Allowing for such mobility would artificially inflate Group 1 mortality at high ages and depress it in the more affluent subgroups.

2.2 Deaths, exposures and age-specific 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 at the last birthday at the time of death) to obtain a greater degree of precision. Here our age variable, x, refers to the attained (integer) age at the start of each calendar year. Thus

- $D(i,t,x)$ = the number of individuals allocated to subgroup i at the start of the year, born in calendar year $t-1-x$,[8] who die during calendar year t.
- $E_I(i,t,x)$ = the number of individuals allocated to subgroup i, born in calendar year $t-1-x$, and counted on the first of January in year t.[9]
- $E(i,t,x)$ is the corresponding central exposed to risk, and measures the average size of the population during year t in subgroup i born in year $t-1-x$. Migration of people into and out of Denmark prevents us from calculating $E(i,t,x)$ exactly, so we have chosen to use the common approximation $E(i,t,x) = E_I(i,t,x) - D(i,t,x)/2$.[10]

The age-specific death rate is then

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

in subgroup i, calendar year t and age x at the 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 age-specific death rates where sampling variation will be quite significant.

By way of example, age-specific death rates for the 10 subgroups in 2012 are plotted in Figure 1 (left). 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 the highest 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 socio-economically 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

[8] 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.

[9] $E_I(i,t,x)$ is referred to by actuaries as the *initial exposed to risk*.

[10] 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.

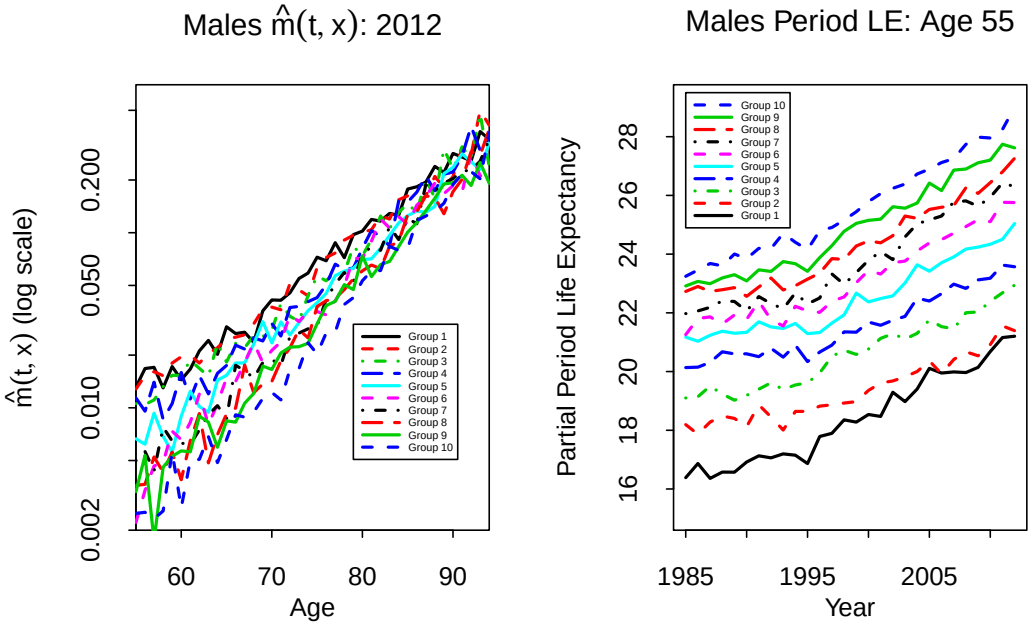

Figure 1: Left: Age-specific death rates, $\hat{m}(i,t,x) = D(i,t,x)/E(i,t,x)$, for Danish males Groups 1 to 10 by age in 2012. Right: Partial period life expectancies (LE) for Danish males Groups 1 to 10 from age 55, capped at age 95.

in the left-hand plot in Figure 1, especially at high ages, but the broad patterns are still clear. We can also observe that relative differences in death rates between subgroups 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 postulates that mortality rates 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, 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 males in England and Wales.

It is also of interest to look at the variation in life expectancies between groups. Here we choose to calculate *partial* period life expectancies using a maximum age of 95 to avoid the need to extrapolate beyond the maximum age in our dataset[11] and it is defined as the expected number of years survived from age x to age $x_u = 95$

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

conditional on having survived to age x, and assuming that mortality stays at the same levels as in year t. This measure can be approximated as follows:

$$LE(t,x) \approx \frac{1}{2} + \sum_{y=x+1}^{x_u-1} S_p(t,x,y) + \frac{1}{2} S_p(t,x,x_u),$$

where $S_p(t,x,y) = \exp\left[-\sum_{s=x}^{y-1} \hat{m}(t,s)\right]$ defines the observed period survival probabilities.[12] We can note the following. At age 55 (and the same holds for other ages), there is a clear separation between observed partial period life expectancies (LEs) in all years across all 10 subgroups (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 age-specific death rates observed in the left-hand plot in Figure 1. The gap between observed 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). Nevertheless, Group 1 improved at a somewhat faster rate than those in Group 2 (and, to a lesser extent, 3 and 4).

Similar observations have been found previously with Danish data by Brønnum-Hansen and Baadsgaard (2012) using *disposable income* as a covariate. However, these authors 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 *disposable 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 subgroups 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 subpopulation death rates with a view to assessing how subgroups 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 life expectancies in England and Wales 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.

[12] Partial life expectancy up to age 95 is only very slightly less than complete (untruncated) life expectancy unless the current age, x, is quite high.

2.3 Education

We also considered an individual's level of education as an alternative covariate (see, for example, Kitagawa and Hauser, 1968, Duleep, 1989, and, more recently, Olshansky et al., 2012, and Sasson, 2016). For consistency with Mackenbach et al. (2003) and partially with Brønnum-Hansen and Baadsgaard (2012), we subdivided the population into three education levels as shown in Table 2.

Low education	Primary and lower secondary education (ISCED levels 1, 2)
Medium education	Upper secondary education (ISCED 3A, 3C, 4A) (gymnasium or technical/vocational)
High education	Tertiary education (ISCED 5, 6)

Table 2: Education levels.

It is important to note that for the Danish population, most individuals born before 1922 have no recorded level of educational attainment (that is, we have no reliable education data for the first 31 cohorts in our ages 55-94, years 1985-2012 dataset).

We calculated two quantities: Age Standardised Mortality Rates (ASMR) for the age range 30-59 using the 1976 European Standard Population for comparison with Mackenbach et al. (2003); and partial period life expectancies (LE) from age 55 up to age 90.[13] Results are presented in Tables 3 and 4. We can see that, at both lower ages (ASMR, Table 3) and higher ages (LE, Table 4), affluence is a much stronger discriminator than education: that is, it produces a wider separation of the ASMR's and LE's. This observation supports our choice of affluence as the first choice for an analysis using a single covariate.

	Age Standardised Mortality Rate		
Group	1985	1995	2005
Low education	6.2	6.4	5.9
High education	4.0	2.8	1.9
Affluence group 1	11.2	10.2	7.4
Affluence group 10	3.1	2.5	1.5

Table 3: Age Standardised Mortality Rates (per 1000 person years) for ages 30-59 by subgroup and calendar year.

Within each birth cohort, levels of educational attainment will be stable after around age 30 allowing straightforward cohort by cohort analysis of the impact of education on mortality. But it must also be recognised that levels of educational attainment

[13]For high ages, LEs give a more meaningful view of the impact of education or affluence. We use a maximum age of 90 (lower than the upper age of 95 in Subsection 2.2, and e.g. Figure 1) because we do not have complete education data up to age 90, even in 2011.

Group	LE 2011
Low education	22.9
High Education	26.6
Affluence group 1	20.9
Affluence group 10	27.4

Table 4: Partial period life expectancies (LE) from age 55 to 90.

have been rising steadily over time (see, e.g., Brønnum-Hansen and Baadsgaard, 2012, and Sasson, 2016), 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. For example, the lowest educated group might contain an increasing percentage of the most deprived in society, a group with higher mortality (Bound et al., 2016).[14] This shift would lessen the impact of underlying mortality improvements, and separation of the two effects within the low education group might be difficult.

3 Modelling subpopulation mortality

Multi-population mortality modelling concerns the development over time of the death rates in several populations. Let $m(i,t,x)$ be the underlying age-specific 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*, and further discussion and specific models can be found, in Li and Lee (2005), Cairns et al. (2011), and Hyndman et al. (2013), whose models achieve coherence through a careful choice of time series models for period effects. Related approaches to multi-population modelling include those of Dowd et al. (2011; gravity model) Jarner and Kryger (2011; SAINT model) and Kleinow (2015; Common Age Effect model).

We propose the following gravity model of the CBD-X type for underlying age-specific death rates in the 10 subgroups:

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

where i is the subgroup, 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

[14] In other words, we need to recognise that the low education group is not homogeneous, and that its internal mix changes over time.

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, while the $\kappa_2^i(t)$ terms pick up 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 subgroups over the full range of ages from 55 to 94. Without it, we found that the mortality curves in individual years for different subgroups would cross over at high ages in a way that was not consistent with the crude age-specific death rates. As we discuss later, this model was found to fit the 10 subpopulations 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 projecting, we need to specify a stochastic model for the period effects. We propose the following:

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

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

where

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

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

As outlined below, equations (3) and (4) correspond to two underlying random walk processes, $\bar{\kappa}_1(t)$ and $\bar{\kappa}_2(t)$ that govern mortality improvements at the national level, with additional subgroup deviations from $\bar{\kappa}_1(t)$ and $\bar{\kappa}_2(t)$ that are mean reverting, to prevent subgroups from diverging from each other. 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 subgroups that prevent individual subgroup 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 properties hold for the $\Delta_{2i}(t) = \kappa_2^{(i)}(t) - \bar{\kappa}_2(t)$.

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 parameterisation implies that the correlations in log death rates between all pairs of subgroups are all the same.

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

We will move on now to discuss how parameters in the model are estimated.

Both latent state variables and process parameters were estimated using Bayesian methods.[15] These methods provide a rigorous and coherent framework for finding point estimates and also for assessing parameter uncertainty. Additionally, they allow us to mitigate problems that arise as a result of small population sampling recently highlighted by Haberman et al. (2014) and developed further by Chen et al. (2015).[16]

We seek to estimate the Bayesian 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 subgroup, year and age. The posterior distribution is proportional to

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

[15] Estimation was carried out using the statistics package R using the authors' own code.

[16] Specifically, they find that small populations introduce a significant upwards bias in the volatility of the period effects in a stochastic mortality model estimated using standard maximum likelihood (ML). In this study, the affluence based subgroups are about 1% of the size of the England & Wales (EW) population. Chen et al. (2015) compare populations that are 100% and 1% of the size of EW, and find there is significant bias in ML-based projections for the latter.

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 $\Delta(1)$.[17]

In practice, f_1 depends only on D, β and κ, and not 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.[18]

[17] 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.

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

The MCMC algorithm simulates a Markov chain, $\theta(t)$, where θ is the vector of latent state variables (the age effects and historical period effects) and the process parameters. After the 'burn-in' period, the empirical distribution of the observed (simulated) $\theta(t)$ series converges to the posterior distribution for θ. In this study, 50,000 iterations after the burn-in period were considered to be sufficient for this convergence. Following standard practice (Gilks et al., 1996)

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, *a priori*, that the random walk processes for each of the 10 subpopulations are, individually, neither more nor less volatile than the national population.

Uniform prior distributions are assumed throughout apart from:

- We selected a Beta prior for ρ to ensure it remains in the range $(0, 1)$. We tried both Beta$(2, 2)$ (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.

- We also selected a Beta prior for ψ 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.

- An inverse gamma prior was chosen for ν, centred around 1.[19] 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.

, we then record one out of every 50 of these to reduce the correlation between the consecutive recorded observations to low levels. We denote this recorded sequence by $\theta_1, \ldots, \theta_{1000}$.

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

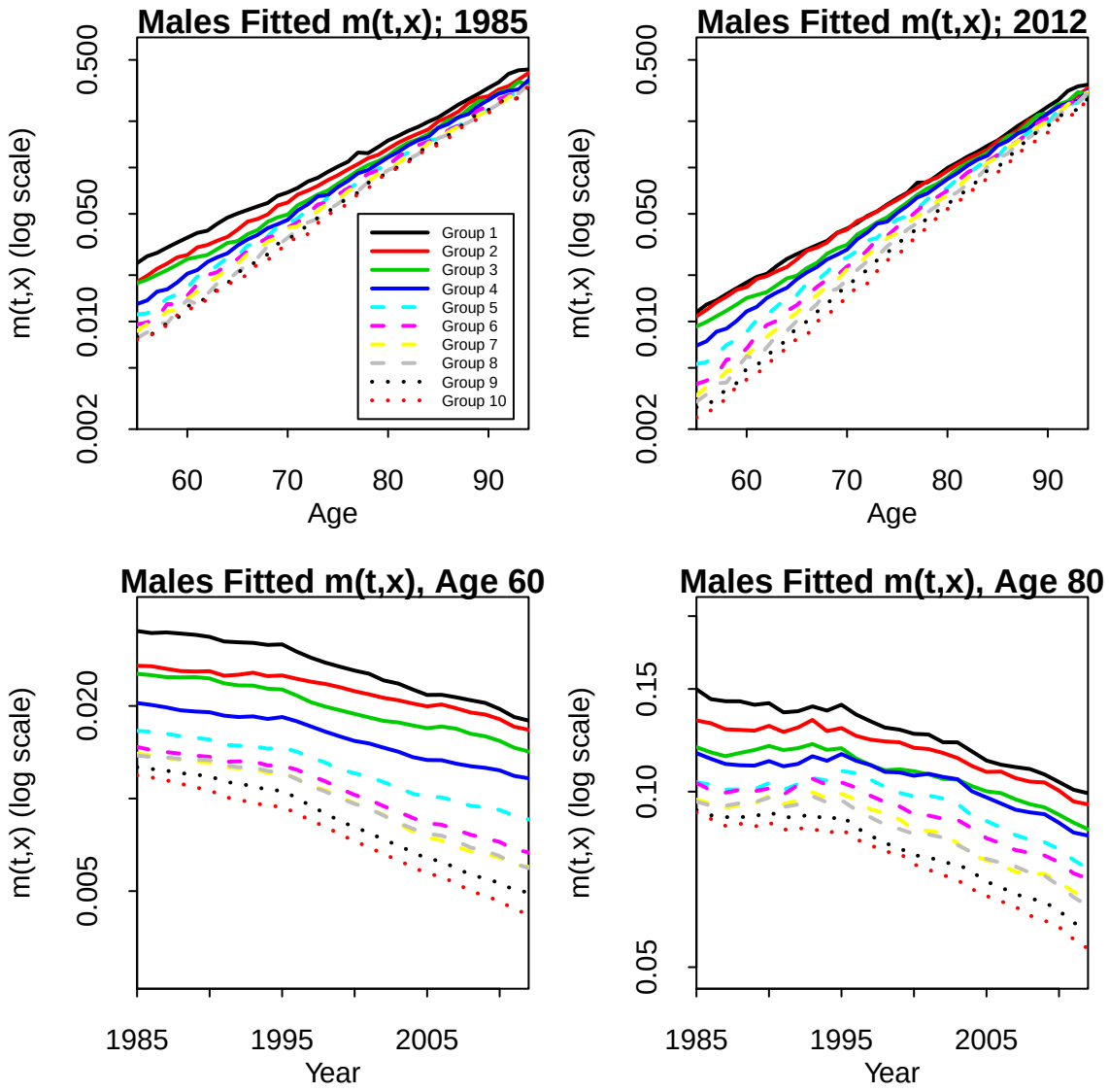

Figure 2: Fitted age-specific death rates, $m(i,t,x)$, for Danish males Groups 1 to 10 using the CBD-X model, with subgroupings based on the affluence index (wealth+15×income) and lockdown at age 67. Top row: by age in years 1985 and 2012. Bottom row: by year for ages 60 and 80.

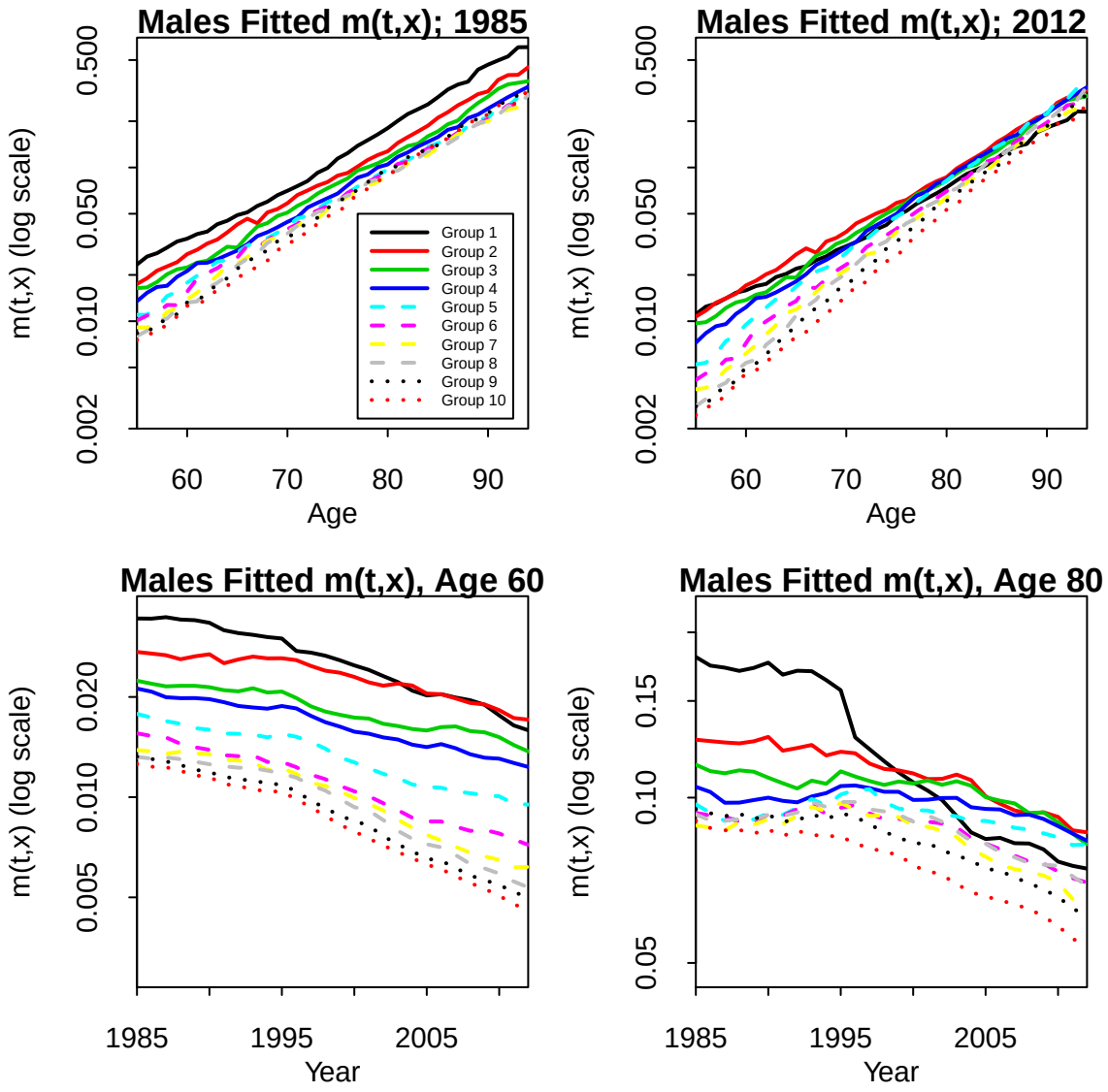

Figure 3: Fitted age-specific death rates, $m(i, t, x)$, for Danish males Groups 1 to 10 using the CBD-X model, but with subgroupings based on income only and no lockdown at age 67.

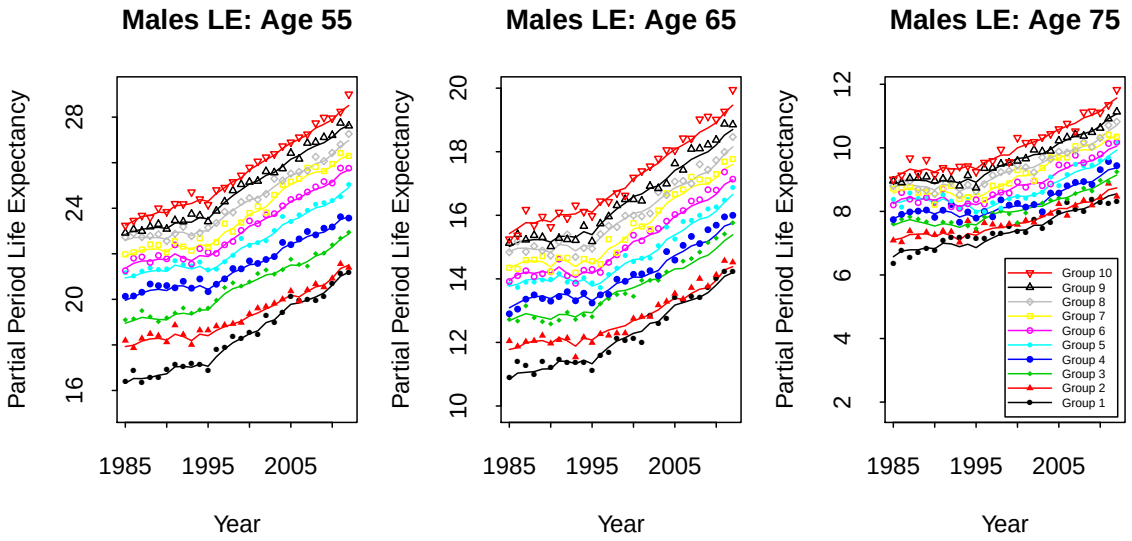

Figure 4: Partial period life expectancy (LE) for Danish males in Groups 1 to 10 for ages 55 (left) and 65 (middle) and 75 (right). Lines: LEs based on fitted age-specific death rates. Points: LEs based on observed age-specific death rates.

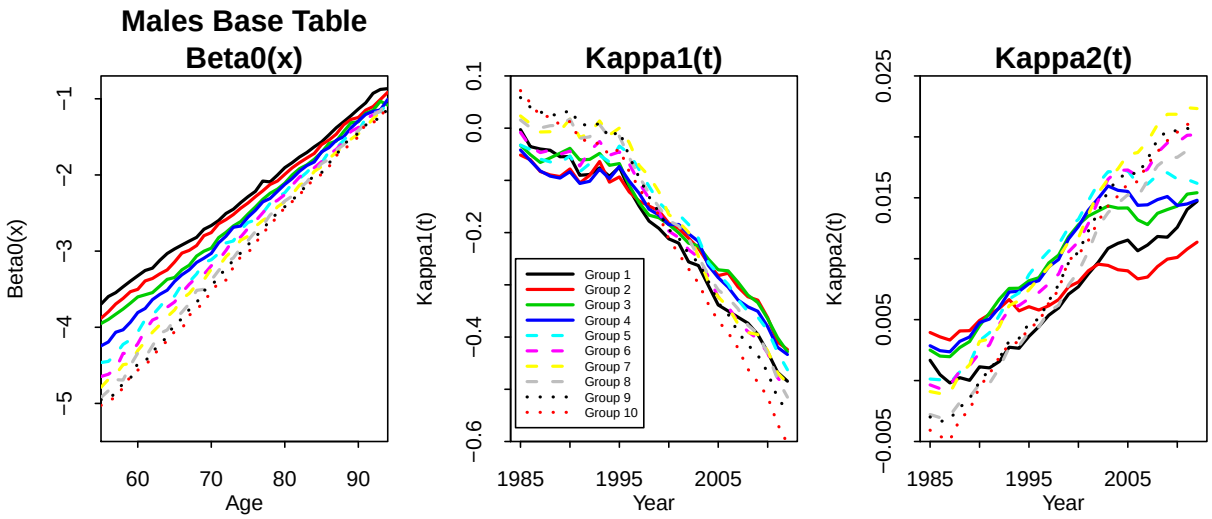

Figure 5: Estimated age and period effects for the 10 subpopulations.

4 Analysis of historical death rates

4.1 Fitted age-specific death rates

Figure 2 shows fitted age-specific 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 observed age-specific death rates in Figure 1. The fitted rates smooth out the noise very considerably and achieve a crisp distinction between each of the subgroups. At the same time, this clarity is achieved without losing any of the essential patterns and characteristics underpinning the observed rates. From these plots of smoothed fitted age-specific death rates (and others not included here), we can confirm the findings we first observed from plots of observed rates:

- Falling death rates over time at all ages.
- A very wide gap in age-specific death rates between the least and most affluent at younger ages, and a narrower gap at higher ages.
- Relative to 1985 age-specific death rates, improvements have been largest amongst the most affluent and at younger ages (resulting in a widening life expectancy gap).

For comparison with previous grouping methods, equivalent plots are provided in Figure 3, where subgroups are based on income only *with no* lockdown 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 to be the result of changes in the amounts and treatment of social assistance and old-age pensions in the previous year.[20] After 1995, the subgroup 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 at 67 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 observed age-specific death rates, while the lines show LEs based on fitted age-specific 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

[20] The changes in the reporting of income in 1995 have resulted in a 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.

consistency from year to year between the 10 subgroups, 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 subgroups 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 comparison with the US.

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 subgroups 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.[21] [22] However, we can see that $\kappa_1^{(10)}(t)$ starts higher and falls more steeply than is the case with the other subgroups: 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,with the interpretation 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 Model validation

The stochastic model was subjected to a number of in-sample and out-of-sample model validation tests. Standardised residuals were not found to exhibit any clus-

[21] The prominence of the kink around 1995 observable in Figures 2, 4 and Figure 5 is most likely due to 1995 being 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. A detailed analysis and explanation for the change in trend is beyond the scope of this paper. However, Christensen et al. (2010) point to changes in major lifestyle risk factors such as smoking, alcohol consumption and exercise occurring in Denmark around this time.

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

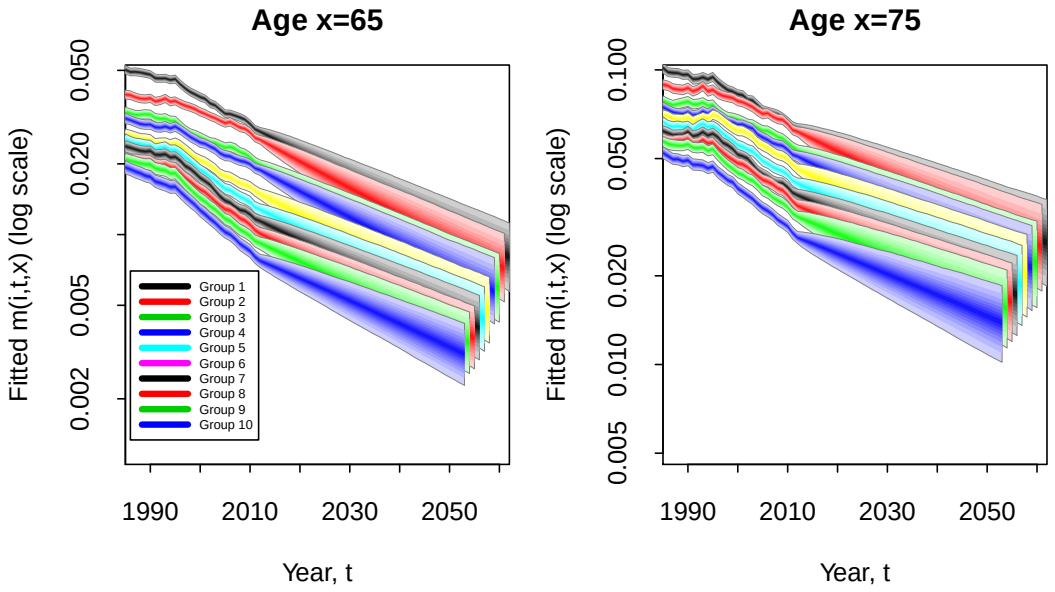

Figure 6: Fan charts for age-specific death rates at ages 65 and 75 for Danish males Groups 1 to 10. The charts show parameter uncertainty in the fitted age-specific death rates up to 2012, and combined parameter uncertainty and process risk from 2013. The lower and upper edges of the fans correspond to the 5% and 95% quantiles, respectively.

tering or obvious cohort effects and had the right level of variability to suggest a good in-sample fit, consistent with model assumptions. Estimates of age and period effects were found to be robust relative to the choice of estimation period. And out-of-sample experience was found to be consistent with projections. For further details, see Appendix A.

5 Future mortality and survivorship

In this section, we will consider simulated future mortality. Bearing in mind that different applications make use of different quantities, we will present simulation results for both age-specific death rates and cohort survival probabilities.

5.1 Projecting future age-specific death rates

We begin by projecting future age-specific death rates. The results in this section, unless otherwise stated, include full parameter uncertainty. This means that we draw 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.[23]

Fan charts for the forecast age-specific 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 subgroup 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). Forward 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 issue 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 #1 and #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

[23] Let M be the number of independent scenarios to be generated. We first generate $I_1, \ldots, I_M$ i.i.d. random variables uniformly distributed on the integers $1, \ldots, 1000$. Building on the remarks in Footnote 18, scenario k then uses the vector θ_{I_k} from which we extract the historical values of the $\beta_0^{(i)}(x)$, $\kappa_j^{(i)}(t)$, while the process parameters also extracted from θ_{I_k} provide the basis for simulating future values of the $\kappa_j^{(i)}(t)$. For further details, see Cairns et al. (2011).

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 subgroup death rates. But, since each subgroup 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 people. 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 subgroups has been pooled. 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.[24] Although this experiment is a limited one, 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 survival probabilities. A general cohort survivor index, $S(t, x)$, represents the probability that an individual aged exactly x (an integer) at time 0 (the beginning of calendar year 1) survives for t years to age $x + t$, given the knowledge of how the underlying age-specific death

[24] 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).

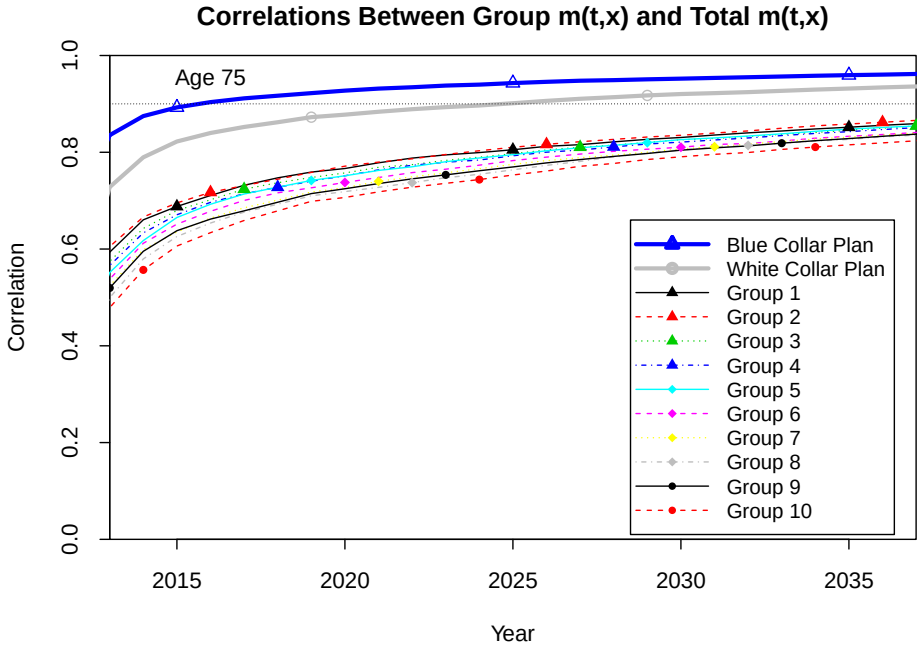

Figure 7: Empirical correlations between individual subgroups, two pension plans, and the total population, cor $(m(i,t,x), \bar{m}(t,x))$, for age 75. Simulations incorporate full parameter uncertainty. The dotted line at a correlation of 0.9 is for reference only.

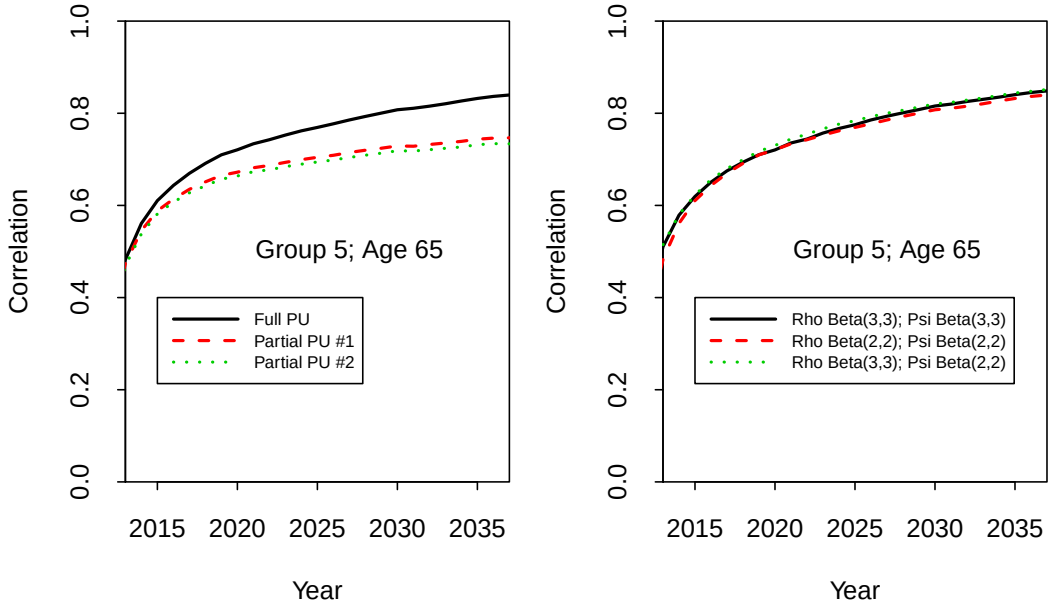

Figure 8: Empirical correlations between Group 5 and the total population, cor $(m(i,t,x), \bar{m}(t,x))$, for age 65, showing sensitivity to changes in inputs and assumptions. Left: impact of different levels of parameter uncertainty: Full parameter uncertainty (PU); PPU #1 has μ_1 and μ_2 fixed at 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.

rates, $m(t, y)$, evolve from time 0 to t. Thus[25]

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

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 discussed above. 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 subgroups, 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 subgroups, 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 subgroups contribute more to the uncertainty in $S_{TOT}(t, x)$. In later years, however, the less affluent subgroups 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 survivorship correlation,

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

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 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)

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

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.

[25] We use the approximation that, for an individual aged exactly integer x at the start of year t, the probability of death during year t is $q(i, t, x) = 1 - \exp\left(-m(i, t, x)\right)$.

These correlation plots help us to identify a number of desirable criteria from the perspective of biological and socio-economic 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 death rates or between survivor indexes for different populations should vary smoothly with the time horizon and should be increasing 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.[26]

It is difficult in general terms 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 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.[27]

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, perhaps suggesting that reported levels of income and wealth do not reflect the true level of affluence of the females in Group 1. Further work needs to be done on female mortality, 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 could look at the explanatory power of including other covariates, such as education (previously discussed), marital status, occupation and area of

[26] That is, correlations between different socio-economics groups should be less than 1 at all ages and time horizons.

[27] 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 neither plausible nor realistic when correlations are (significantly) less than 1 at other ages.

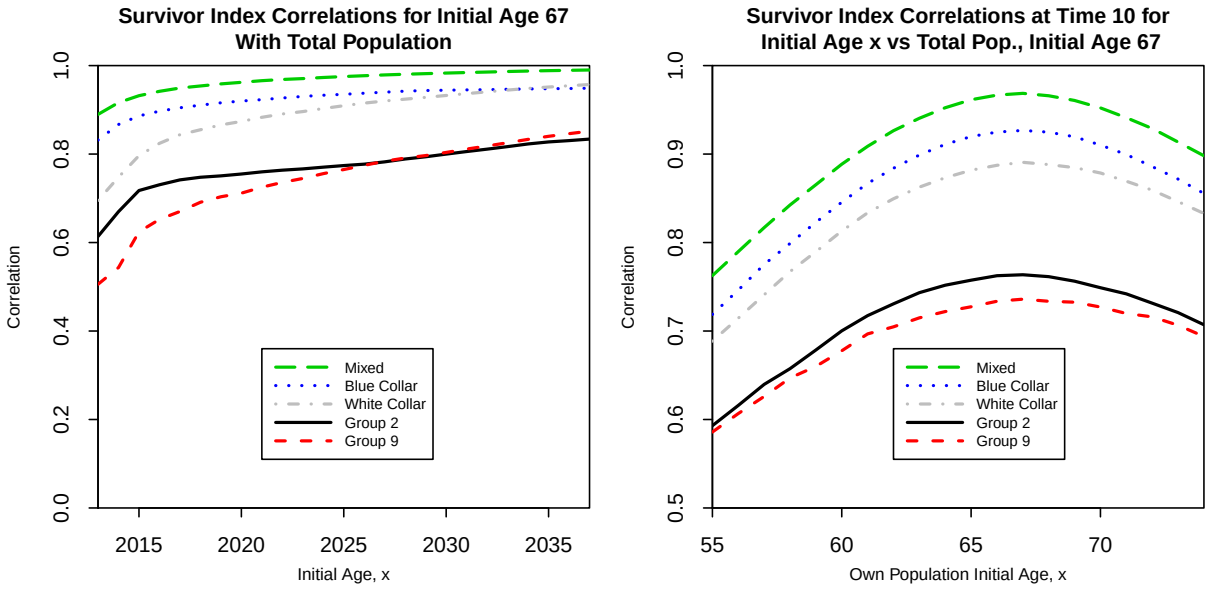

Figure 9: Left: Empirical correlations between $S_X(t, 67)$ and $S_{TOT}(t, 67)$ for $t = 2013, \ldots, 2037$. Right: Correlations between $S_X(10, x)$ and $S_{TOT}(10, 67)$ for $x = 55, \ldots, 74$. In both plots, results for Groups $X = 2$, 9, blue collar, white collar, and mixed pension plans are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

residence. In each case, we need to think carefully about sample sizes to allow us to elicit statistically significant results.

- We could look at how individuals migrate between different affluence subgroups. From an insurance and pension perspective, it is of most interest to determine the probabilities of ending up in each of Groups 1 to 10 at age 67.
- We could consider stochastic models that offer an alternative to the random walk with constant drift, e.g., we might adopt the approach of Liu and Li (2017).
- We could 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. For example, Case and Deaton (2015) attribute stagnating mortality in the midlife, white, non-Hispanic, US population primarily to increasing mortality from suicide, drug and alcohol poisoning, and chronic liver diseases and cirrhosis.
- We could consider how to model mortality above age 95, in a way that exploits the convergence in death rates at high ages observed in Figure 2.

There are potentially many other "big data" analyses that could be conducted, including cohort-based analyses where we track individuals over time and assess the

impact of various risk factors on mortality including all of those discussed above. 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 Summary

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 using income or education as a covariate, namely at the decile level, and for single ages, using just a single covariate to allocate males to a subgroup. 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 subgroups, 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. In their final year before reaching age 67, the state pension age in Denmark for some of the period under investigation, they would be allocated to a subgroup and remain in that subgroup for the remainder of their lives, a procedure we call *lockdown*. Lockdown at a particular age combined with the affluence index was found to be critical for ensuring a consistent ranking of subgroup death rates across all years and all ages. The use of affluence-based deciles was found to reveal greater levels of inequality than, say, quartiles.

We also introduced a flexible multi-population stochastic gravity-type mortality model for both fitting death rates and projecting 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 subgroups, but also preserves the subgroup rankings over different time horizons.

Model-based smoothing was employed to dampen the effect of sampling variation in the underlying crude age-specific death rates, and we used these smoothed death rates to rank the 10 subgroups at each age and in each year. This smoothing was achieved without losing any of the essential patterns and characteristics underpinning the crude age-specific death rates. Subgroup 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.

Another key element of the paper was an analysis of how correlations between subgroups and the national population change as the projection horizon lengthens. We looked at both projected 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 across a range of the 10 subgroups 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 with to generate clear and consistent rankings of death rates at all ages down to at least the decile level using just a single, well-chosen covariate to allocate individuals to a particular subgroup.
- 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 compared to looking at income alone which most previous studies have concentrated on. It is also superior to education on its own.
- 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 subgroups prior to the lockdown age without violating the subgroup rankings, but they needed to be locked in to a particular subgroup once they reached a certain age in order to preserve the subgroup rankings at higher ages.
- The age at which lockdown happens might well be related to 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 subgroup death rates by fitting a multi-population stochastic gravity-type mortality model. Fitting this model has the benefit of reducing the effect of idiosyncratic mortality risk in small populations and can then be used to make mortality projections, with the gravity effect helping preserve subgroup rankings.
- Correlations between subgroups and the national population rise with the time horizon and are especially high for subgroups that contain a diverse mixture of socio-economic groups, such as in a large pension plan.

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.

Booth, H., Maindonald, J., and Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies, 56*, 325-336.

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.

Bound, J., Geronimus, A.T., Rodriguez, J.M., and Waidmann, T.A. (2015). Measuring recent apparent declines in longevity: The role of increasing educational attainment. *Health Affairs, 34(12)*, 2167-2173.

Brønnum-Hansen, H., and 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. *Journal of the Royal Statistical Society, Series A, 179*, 975-1005.

Cairns, A.J.G., and El Boukfaoui, G. (2017) Basis Risk in Index Based Longevity

Hedges: A Guide For Longevity Hedgers. Working paper. Edinburgh: Heriot-Watt University.

Case, A., and Deaton, A. (2015). Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. *Proceedings of the National Academy of Sciences of the United States of America, 112*, 15078-15083.

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.

Chen, L., Cairns, A.J.G., and and Kleinow, T. (2015). *Small Population Bias and Sampling Effects in Stochastic Mortality Modelling.* Working paper. Edinburgh: Heriot-Watt University.

Christensen, K., Davidsen, M., Juel, K., Mortensen, L., Rau, R., and Vaupel, J.W. (2010). The divergent life-expectancy trends in Denmark and Sweden – and some potential explanations. In Crimmins et al. (2010, pp. 385-407).

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 Academies 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.

Crimmins, E.M., Preston, S.H., and B. Cohen, B. (eds.) (2010). *International Differences in Mortality at Older Ages: Dimensions and Sources.* Washington, DC: The National Academies Press.

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.* doi:10.1136/jech-2015-206173.

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.

Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M. (2010). Backtesting stochastic mortality models: An ex-post evaluation of multi-period-ahead density forecasts. *North American Actuarial Journal, 14*, 281-298.

Duleep, H.O. (1989). Measuring socioeconomic mortality differentials over time. *Demography, 26*, 345-351.

Enchev, V., Cairns, A.J.G., and and Kleinow, T. (2016). Multi-population mortality models: Fitting, forecasting and comparisons. To appear in *Scandinavian Actuarial Journal*.

Feldman, J.J., Makuc, D.M., Kleinman, J.C., and Cornoni-Huntley,J. (1989). National trends in education differentials in mortality. *American Journal of Epidemiology 129(5)*, 919-933.

Gavrilov, L.A., and Gavrilova, N.S. (1979). Determination of species length of life. *Doklady Akademii Nauk SSSR, 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.

Glei, D., Meslé, F., and Vallin , J. (2010). Diverging trends in life expectancy at age 50: A look at causes of death. In Crimmins et al. (2010, pp. 17-67).

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.

Hunt, A., and Blake, D. (2017) Mortality modelling for pension schemes. To appear in *ASTIN Bulletin.*

Hyndman, R., Booth, H., and Yasmeen, F. (2013). Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography, 50*, 261-283.

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

Juel, K. (2008) Middellevetid og dødelighed i Danmark sammenlignet med i Sverige. Hvad betyder rygning og alkohol? [Life expectancy and mortality in Denmark compared to Sweden. What is the effect of smoking and alcohol?] *Ugeskrift Laeger, 170(33)*, 2423-2427. (In Danish.)

Kitagawa, E.M., and Hauser, P.M. (1968). Education differentials in mortality by cause of death: United States, 1960. *Demography, 5*, 318-353.

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.

Lee, R., and Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography, 38*, 537-549.

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.

Liu, Y., and Li, J.S.-H. (2017). The locally-linear Cairns-Blake-Dowd model: A note on delta-nuga hedging of longevity risk. To appear in *ASTIN Bulletin*.

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*, 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., Boersch-Supan, A.,

Cacioppo, J.T., Carnes, B.A., Carstensen, L.L., Fried, L.P., Goldman, D.P., Jackson, J., Kohli, M., Rother, J., Zheng Y., and Rowe, J. (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.

This version: December 2, 2016

A Model validation

A.1 Standardised residuals

Standardised (Pearson) residuals based on the Poisson model for deaths are defined as

$$\epsilon(i,t,x) = \left(\hat{m}(i,t,x) - m(i,t,x)\right) / \sqrt{m(i,t,x)/E(i,t,x)}.$$

If the model fits well, then the $\epsilon(i,t,x)$ should be approximately i.i.d. standard normal. Heat plots of these (see, e.g., Cairns et al., 2009) for the 10 affluence groups are provided in Figure 10 using the posterior mean for the $m(i,t,x)$. As a graphical diagnostic, each of these plots looks almost completely random, with no systematic clustering, and hence consistent with the independence criterion. Additionally, (a) the $\epsilon(i,t,x)$ have mean -0.07 and variance 1.004, and (b) a QQ plot of the residuals against a standard normal is reasonably linear with only marginal evidence of a fat tail (kurtosis = 3.14).

Overall, these results show that, although the fit is not perfect, it is very much better than can typically be achieved for some other populations or models (see, e.g., Cairns et al., 2009, Section 6.1.2). This finding allows us to conclude that the model does provide us with a satisfactory in-sample fit over the period 1985-2012.

A.2 Robustness

We tested for robustness of parameter estimates, following Cairns et al. (2009). Parameter estimates using data from three periods: 1985-2012; 1995-2012; 1995-2004. By way of example, results for Group 6 are plotted in Figure 11. We can see that the broad shape of each of $\beta_0^{(6)}(x)$, $\kappa_1^{(6)}(t)$ and $\kappa_2^{(6)}(t)$ is largely unchanged in the years where estimates overlap. The approximate parallel shifts in $\kappa_1^{(6)}(t)$ and $\kappa_2^{(6)}(t)$ are the artificial result of the constraint that $\bar{\kappa}_1(t) = \bar{\kappa}_2(t) = 0$ in the first year of observation (1985 or 1995). The same constraints result in a small shift (from the $\bar{\kappa}_1(t)$ constraint) and a tilt (from $\bar{\kappa}_2(t)$) in the $\beta_0^{(6)}(x)$ curves. Based on this limited test, there is no evidence to suggest that the model fit is not robust.

A.3 Optimal historical calibration period

For out of sample backtesting, we need, first, to choose an historical period (not including the most recent years) to which the model is calibrated. We then simulate from the year after the end of the calibration period up to the most recent year that has been actually observed and compare these with the forecasts. To choose the best calibration period, we follow the approach of Booth et al. (2002). We focused on the posterior means of the fitted $\bar{\kappa}_1(t)$, this being the main driver of uncertainty

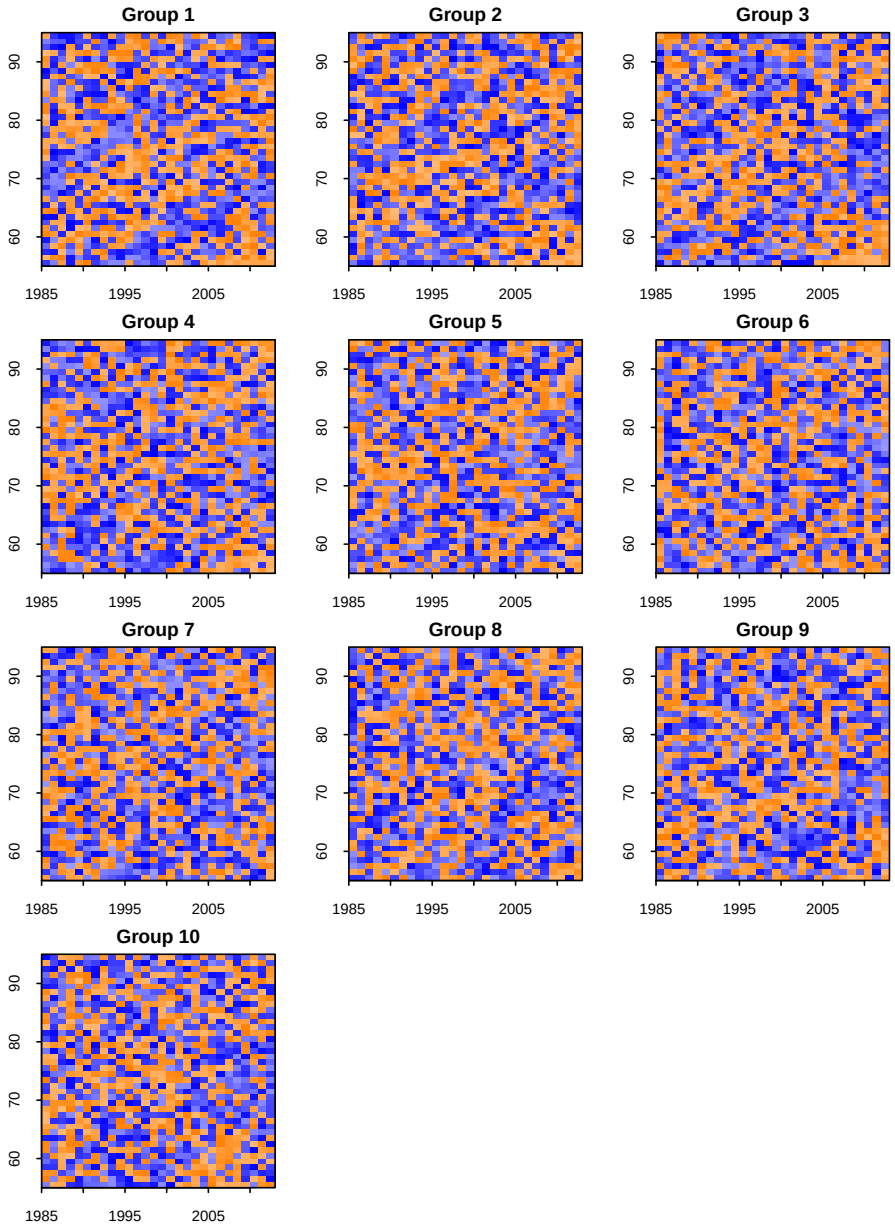

Figure 10: Plots of standardised (Pearson) residuals, $\epsilon(i,t,x)$, for affluence groups 1 to 10. Deep orange: $\epsilon > 0$, close to 0. Light orange: $\epsilon > 0$, further from 0. Deep blue: $\epsilon < 0$, close to 0. Light blue: $\epsilon < 0$, further from 0.

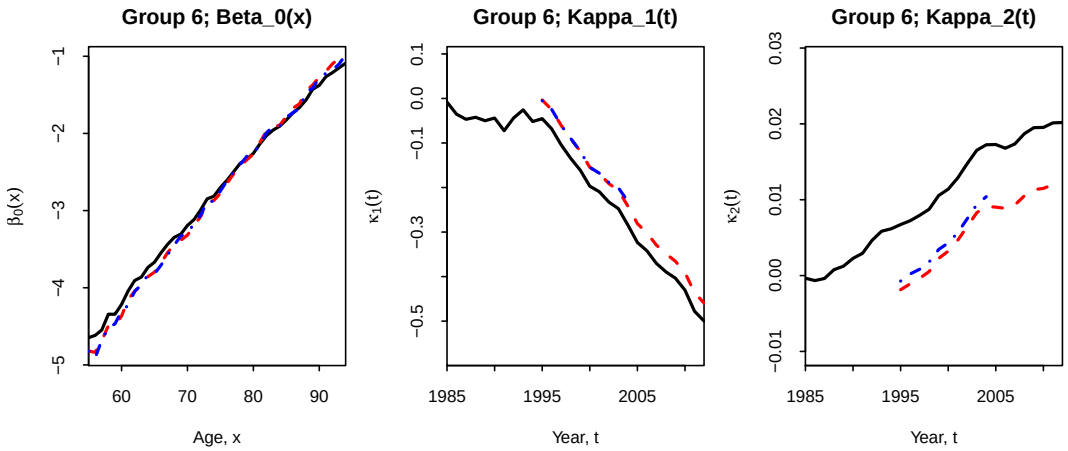

Figure 11: Point estimates of age and period effects for Group 6. Left: $\beta_0^{(6)}(x)$. Middle: $\kappa_1^{(6)}(t)$. Right: $\kappa_2^{(6)}(t)$. Solid lines: values based on 1985-2012 calibration. Dashed lines: 1995-2012. Dot-dashed lines: 1995-2004.

in the level of mortality in all groups. Over the full period 1985-2012, we observe a change in trend around 1995 (see Figure 5 for the individual $\kappa_1^{(i)}(t)$). Booth et al. (2002) remark that trend changes can, and do happen (e.g., in Australia) and they propose a method for choosing the optimal start date for calibration. By directly examining the $\bar{\kappa}_1(t)$ (rather than the individual death counts in addition), we found the optimal start date to be 1995.

A.4 Out-of-sample backtesting

Following Lee and Miller (2001) and Dowd et al. (2010), we compare out of sample projections of partial period life expectancy (LE) from age 55 based on a model calibration using data from 1995 to 2004. Results are presented in Figure 12. Grey fans show the out-of-sample projections of the LE with no allowance for sampling variation. Red fans show the outcome for LEs based on the posterior distribution for the underlying death rates. Blue dots show the outcome for LEs based on observed mortality (i.e. including sampling variation). The actual outcomes (red fans) for the 10 groups look quite reasonable, with some groups above and some below their central projection.

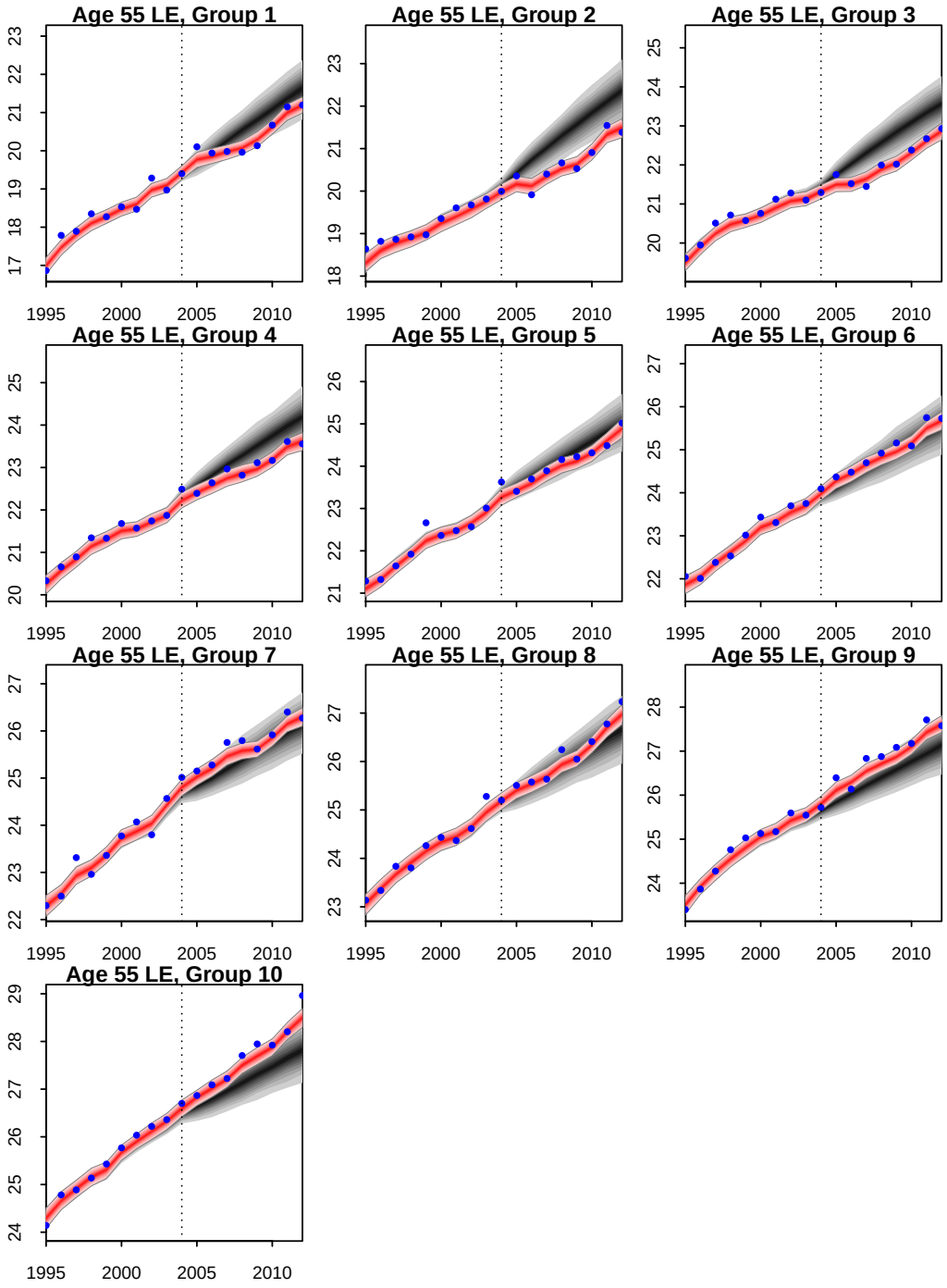

Figure 12: Fan charts of historical and projected partial period life expectancy (LE) from age 55 for each of affluence groups 1 to 10. Grey fan: credibility intervals (5% to 95% quantiles) for underlying LE (historical and projected) using data from 1995 to 2004 and simulated mortality from 2005 to 2012. Red fan: credibility intervals using data from1995 to 2012. Blue dots: empirical LEs using observed mortality.