Phantoms Never Die: Living with Unreliable Population Data

Andrew J.G. Cairns[a]

Heriot-Watt University, Edinburgh, UK

David Blake

Pensions Institute, Cass Business School, London, UK

Kevin Dowd

Durham University Business School, UK

and Amy R. Kessler

Prudential Retirement, Newark, New Jersey, USA

Summary. The analysis of national mortality trends is critically dependent on the quality of the population, exposures and deaths data that underpin death rates. We develop a framework that allows us to assess data reliability and identify anomalies, illustrated, by way of example, using England & Wales (EW) population data. First, we propose a set of graphical diagnostics that help to pinpoint anomalies. Second, we develop a simple Bayesian model that allows us to quantify objectively the size of any anomalies. Two-dimensional graphical diagnostics and modelling techniques are shown to improve significantly our ability to identify and quantify anomalies.

An important conclusion is that significant anomalies in population data can often be linked to uneven patterns of births in cohorts born in the distant past. In the case of EW, errors of more than 9% in the estimated size of some birth cohorts can be attributed to an uneven pattern of births. We propose methods that can use births data to improve estimates of the underlying population exposures.

Finally, we consider the impact of anomalies on mortality forecasts and annuity values, and find significant impacts for some cohorts.

Our methodology has general applicability to other population data sources, such as the Human Mortality Database.

Keywords: Baby boom, cohort-births-deaths exposures methodology, convexity adjustment ratio, deaths, graphical diagnostics, population data.

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

1 Introduction

The field of stochastic mortality modelling has seen rapid growth in recent years, building on the early work of Lee and Carter (1992). Different strands of research have emerged in demography (for example, Hyndman and Ullah, 2007), in statistical methodology (for example, Brouhns et al., 2002, Czado et al., 2005, Girosi and King, 2008, Li et al., 2009), and in the development of new models. The latter, in particular, has seen rapid progress since the early 2000's, driven by emerging risk management applications in the pensions and life insurance industries (for example, Blake et al., 2006, Li and Hardy, 2011, Blake et al., 2013, Cairns, 2013, Biagini et al., 2013). Model development has been driven by the need to have a more accurate description of the underlying pattern of mortality improvements. Examples include: multifactor models (for example, Cairns et al., 2006, 2009, Hyndman and Ullah, 2007), allowance for cohort (year-of-birth) effects (for example, Willets, 2004, Cairns et al. 2009), and multi-population models (for example, Li and Lee, 2005, Cairns et al., 2011b, and Börger et al., 2014).

The great majority of studies that are concerned with modelling build on the assumption that the underlying population data (typically deaths and exposures) are accurate. Some authors have highlighted unusual patterns of mortality amongst some other birth cohorts, but they focused on finding explanations for these effects, rather than question the accuracy of the underlying data. For example, the 1918, 1919 and 1920 cohorts in England & Wales (EW) males exhibit unusual characteristics (e.g. Cairns et al., 2009) which have been attributed to a combination of lifelong *frailty* effects linked to the end of the First World War and the devastating Spanish Flu Epidemic. Further discussion of these effects can be found in Richards (2008), who speculates, additionally, that the real cause of the 1919 cohort effects is the mis-estimation of the exposure to risk due to an uneven pattern of births during the year. This line of reasoning is developed further in Section 2.

Two types of data are needed by mortality modellers. First, death counts, $D(t,x)$, by calendar year t and age x (typically age x last birthday). In some countries deaths are further subdivided by year of birth. Second, exposures (a measure of the underlying population), $E(t,x)$, again frequently available by calendar year and individual year of age. We define $P(s,x)$ as the number of persons aged x last birthday at exact time s (so *year* t runs from *exact time* t to $t+1$). Typically this quantity is only calculated once per year rather than continuously. In England & Wales, for example, the Office for National Statistics (ONS) estimates the *mid-year* population, $P(t+\frac{1}{2},x)$, for integer t. The ONS then calculates crude death rates by dividing death counts by mid-year population estimates: that is, $m(t,x) = D(t,x)/P(t+\frac{1}{2},x)$. More

commonly, the crude death rate is defined as

$$m(t,x) = D(t,x)/E(t,x) \quad (1)$$

$$\text{where } E(t,x) = \int_0^1 P(t+s,x)ds \quad (2)$$

is the *central exposed to risk*: the mean of the population aged x last birthday during year t (see, for example, Wilmoth et al., 2007).

EW is typical of a number of countries in which population figures are estimates based on data from a variety of sources, primarily decennial censuses. Censuses are affected by uncertainty in the extent of under-enumeration. In addition, between censuses, the estimation of net migration at individual ages is difficult (net migration being defined as the number of immigrants minus the number of emigrants).

Following the 2011 census, the ONS published revisions to the 2001-2011 intercensal population estimates (ONS, 2012a,b). Between ages 40 and 85 these revisions were relatively modest (no more than about 3%). However, for males aged above 90 revisions were much more significant with downwards revisions to population estimates of about 15% with matching upwards revisions to published death rates. For younger adult males, there were also significant revisions which were, most likely, the result of adjustments for migration. At the highest ages, migration is very modest and so is unlikely to be the main reason for the change. It is more likely that the population at ages 80+ was slightly overestimated at the 2001 census which led to bigger overestimates by 2010 (the last year that was based on the 2001 post-censal estimates).

These revisions have an impact on our predictions of future rates of mortality at particular ages. Such predictions, necessarily, rely on a projection model. But, in general, revisions to the historical data will have three types of impact on mortality forecasts: the base table (with obvious changes at high ages); the central trend or improvement rate (resulting in modest changes); uncertainty around the central trend (negligible changes).

Importantly, the 2011 population revisions were not a one off. Revisions of a similar magnitude after the 2001 census can be inferred from a comparison of current ONS historical population tables with the original population data for 2000 published by the ONS (ONS 2002a).

It was these censal revisions to the population data that led to the wider study of data errors that we conduct in this paper. Some of these errors might be labelled as bias terms where elements of the underlying methodology systematically induce positive or negative errors as discussed later in the paper. However, in most cases, there is no systematic bias, either positive or negative, so we will use the all-encompassing term *error* in the rest of the paper. It is, of course, inevitable that there will be errors in both population and (to a lesser extent) deaths data, and that the magnitude of these will vary from country to country. What we seek to

do in this study is analyse population and deaths data in as objective a manner as possible (a) to identify potential anomalies or errors in the data, (b) to provide tools for the early identification of emerging anomalies, and (c) to provide a methodology for 'cleaning' the data prior to embarking on a mortality modelling exercise. Understanding and quantifying errors is important for a number of stakeholders with interests in: population mortality forecasts; forecasts of sub-population mortality; calibration of multi-population mortality models; assessment of levels of uncertainty in mortality forecasts; the calculation of life insurer liabilities and economic capital; annuity pricing; pension plan buyout pricing; the assessment of basis risk in longevity hedges; and the effectiveness of hedges and hedging instruments.

In Section 2, we discuss in detail how errors might arise in deaths, population and exposures data, and propose a new methodology, which we call the Cohort-Births-Deaths (CBD) Exposures Methodology for calculating exposures data using births data at the monthly or quarterly frequency. The CBD methodology is also used to improve the calculation of mid-year population estimates based on census data. In Section 3, we develop some model-free graphical diagnostics that allow us to identify anomalies in population data. We find that the more significant anomalies can usually be linked to unusual patterns of births. We propose a Bayesian, model-based approach in Section 4 for quantifying errors in exposures data using uninformative priors. Section 5 looks at the impact of these error adjustments on mortality forecasts and annuity valuation. We provide further discussion and proposals for future work in Section 6.

2 Data Issues: Deaths, Population, Exposures

In this section, we review the reliability of the building blocks of the death rate $m(t,x) = D(t,x)/E(t,x)$. We will start with a brief discussion of death counts, before moving on to population estimates and their relationship with exposures. We will discuss how a detailed knowledge of birth patterns can help explain some of the errors that might arise in practice.

2.1 Deaths

Published death counts in EW, and in many other countries, $D(t,x)$, measure the number of deaths *registered* in calendar year t of people aged x last birthday *at the date of death*. In general, these statistics are generally accepted as accurate, but there are some issues that need to be noted.

In any given year, the total number of deaths, $\sum_x D(t,x)$, can also be regarded as accurate: it is extremely difficult for any death to go unreported in EW. But, it is possible for there to be an error in the reported age at death (e.g., the informant

does not know for sure the age of the deceased). To combat this, death registrars might carry out *some* cross checking at the point of registration where this is possible and afterwards on government databases (ONS, 2013). Cross-checking is not always possible (e.g., for immigrants) leaving open the possibility for errors to creep in and, where checks are not carried out, there might be biases towards certain ages being recorded. For example, if the deceased was thought to be 'about' age 80, then the age at death is reported as age 80. This is known as *age heaping* and can occur in both deaths and census counts (see, for example, Vaupel et al., 1998). We have assumed in the analyses that follow that these potential errors are not material in the EW data. However, we do revisit this assumption towards the end of the paper.

Deaths are typically registered very quickly, with 95% of registrations in 2013, for example, corresponding to occurrences in 2013. The remaining 5% relate to deaths at the very end of 2012 plus those delayed by the requirement for an inquest (ONS, 2014b).

2.2 Population Estimates, Exposures, Death Rates

A potentially greater source of estimation error lies in the central exposed to risk, $E(t,x) = \int_0^1 P(t+s,x)ds$. Since the population is not monitored or estimated on a continuous basis, it is not possible to calculate this expression exactly. For example, in EW, the ONS only publishes mid-year population estimates and then uses the approximation $E(t,x) \approx P(t+\frac{1}{2},x)$.[1] In contrast to the ONS, the Human Mortality Database (HMD) adopts a more sophisticated approach to exposures, as documented in Appendix E (Equation E10) of Wilmoth et al. (2007). However, the arguments applied here to identify the shortcomings of the simpler mid-year approximation can be adapted to the HMD approach.

Errors in $E(t,x)$ can occur in the following ways:

- Type 1: mid-year population estimates can be inaccurate. Typical reasons for this in EW would be: no central and universal population register; censuses are 10 years apart and suffer from under-enumeration problems; ages might be misreported at census dates; and ages or dates of birth might be transcribed or scanned incorrectly.

- Type 2: the estimation of migration flows from year to year between censuses is difficult. For many countries, the reason for this is that there is no systematic counting, at all points of entry, of persons entering and/or leaving the country especially by age. In some cases, relevant borders might not be policed, including the border between EW and other parts of the UK.

[1] Other countries publish estimates of population on 1 January each year, and employ the equally simple approximation $E(t,x) \approx \frac{1}{2}(P(t,x) + P(t+1,x))$.

- Type 3: deaths from year to year might be miscounted. For example, a death might be attributed to the wrong cohort.
- Type 4: the assumption that $E(t,x) = P(t+\frac{1}{2},x)$ might be a poor approximation at some ages.
- Type 5: the method used to derive mid-year population estimates from census data might be inaccurate at some ages.

Errors of type 1, 2 and 3 are well known, are difficult to reduce except at considerable cost, and lead to periodic revisions to population data, especially after decennial censuses. Errors of types 4 and 5 are less widely known and we will argue in this section that the current methodology for estimating exposures can be improved upon if we have access to monthly or quarterly births data.

Census errors and migration errors can be either positive or negative and, *ex ante*, in the absence of other information, we take the view that one is as likely as the other.

2.2.1 Propagation of General Errors Through Time

It is clear that once errors in population counts arise, they will persist through time with the associated birth cohort, i.e., the errors follow the cohort as it ages.

First, assume that mid-year population counts, deaths and migration are accurate. Then,

$$P(t+\frac{3}{2},x+1) = P(t+\frac{1}{2},x) - d(t+\frac{1}{2},x) + M(t+\frac{1}{2},x)$$

where $d(t+\frac{1}{2},x)$ is the number of individuals in the cohort who die between $t+\frac{1}{2}$ and $t+\frac{3}{2}$, and $M(t+\frac{1}{2},x)$ is the corresponding net migration *into* the cohort.

Second, suppose there is an error in the initial population count, but accurate deaths and net migration figures in the following year. Then we have the identity

$$\hat{P}(t+\frac{3}{2},x+1) = \hat{P}(t+\frac{1}{2},x) - d(t+\frac{1}{2},x) + M(t+\frac{1}{2},x)$$

where $\hat{P}(s+\frac{1}{2},y) = P(s+\frac{1}{2},y) + \varepsilon_P(s+\frac{1}{2},y)$ is the initial, erroneous population count and $\varepsilon_P(s+\frac{1}{2},y)$ is the error in the population count. If deaths and net migration figures are accurate, then it is straightforward to see that $\varepsilon_P(t+\frac{1}{2},x) = \varepsilon_P(t+\frac{3}{2},x+1)$: that is, the error at time $t+\frac{1}{2}$ follows the cohort.

Third, the *reported* death counts, $\hat{d}(t+\frac{1}{2},x)$, and net migration, $\hat{M}(t+\frac{1}{2},x)$, are themselves subject to errors and so

$$\varepsilon_P(t+\frac{3}{2},x+1) = \varepsilon_P(t+\frac{1}{2},x) + d(t+\frac{1}{2},x) - \hat{d}(t+\frac{1}{2},x) - M(t+\frac{1}{2},x) + \hat{M}(t+\frac{1}{2},x).$$

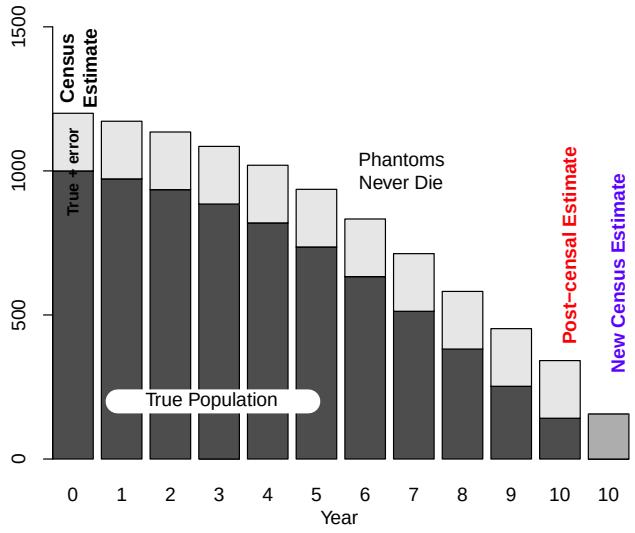

Figure 1: Stylised representation of the changing cohort size between censuses at times 0 and 10. Black bars: true cohort size. Light grey bars: additional phantoms. Dark grey bar (time 10): revised cohort size at the time 10 census.

At younger ages, errors in net migration figures from year to year will tend to dominate, but, at higher ages, the error in the population count at time $t + \frac{1}{2}$ will dominate until such time as the population counts are subject to revision (e.g., at the next census). We call the overestimated population "phantoms". This is because phantoms never die (at least between periodic revisions). That is, if we overestimate the population, then the $\varepsilon_P(t+\frac{1}{2}, x)$ phantoms persist through time, while members of the true population die off.

Figure 1 illustrates this in a stylised way. The cohort has been overestimated by 200 at the time of the census at time 0. The extra 200 form a growing percentage of the total over the next 10 years until the time of the next census. Additionally, the reported death rate (*actual* deaths divided by *over-estimated* population) will be lower than the true death rate, and the relative error in the reported death rate will grow over the 10 years.

After 10 years (final column in Figure 1), there is a further census and it is estimated that the cohort is rather smaller (dark grey bar) than the time-10 post-censal estimate (black and light grey bars), although the revised estimate is likely still to differ from the true value. The question then arises: *how do we accommodate this change in population estimates at time 10*? If the required adjustment is A at time 10, then the ONS (ONS, 2002c, Duncan et al., 2002) adjusts the previous 9 years in a simple linear fashion by cohort. Let $P_P(s+\frac{1}{2}, x+s)$ be the post-censal estimates and $P_I(s+\frac{1}{2}, x+s)$ the inter-censal estimates with $P_P(10\frac{1}{2}, x+10) - P_I(10\frac{1}{2}, x+10) = A$. We then define $P_I(s+\frac{1}{2}, x+s) = P_P(s+\frac{1}{2}, x+s) - A \times s/10$ for $s = 1, \ldots, 10$.

Implicit within this procedure is the assumption that the census measure at time 0 was correct. But the stylised example presented in Figure 1 suggests that, in some circumstances, the ONS's method only partly fixes the problem. Furthermore, the stylised example is more likely to be representative of reality (if exaggerated) for the oldest ages than at younger ages. At younger ages, it is very likely that errors at time 10 are simply due to the difficulties in estimating net migration.

Phantoms arise in situations where the population has been overestimated. A positive number of phantoms might normally be the result of either census errors, an overestimate of net migration and, to a much lesser extent, an underestimate of deaths. As discussed in the following sections, they can also arise because of types 4 and 5 errors. A population can equally, and for the same reasons, be underestimated. This gives rise to trolls, people who exist but are not counted. Their existence is due to a different combination of types 1 to 5 errors. In later sections, we suggest how types 4 and 5 errors can be estimated, but we do not try to estimate the remaining balance between error types 1, 2 and 3.

2.2.2 Census to Mid-year Shift

The next potential error concerns the census to mid-year shift. In EW, censuses occur every 10 years on variable dates in the spring (contrasting with a fixed date of April 1 for US censuses). This issue is illustrated in Figure 2. Suppose, for example, the census occurs at time T_c and the middle of the same year is, say, $T_c + \omega_c$. At $T_c + \omega_c$, we wish to estimate the population aged x last birthday. This will consist of a mixture of people aged x and $x - 1$ last birthday at the time of the census. $P(T_c + \omega_c, x)$ will consist of those aged x at time T_c who did not have a birthday between T_c and $T_c + \omega_c$ (Figure 2, group B), plus those aged $x - 1$ at time T_c who did have a birthday between the census and the middle of the year (group C). $P(T_c + \omega_c, x)$, therefore, consists of appropriate proportions of $P(T_c, x)$ and $P(T_c, x - 1)$, adjusted for deaths and net migration between T_c and $T_c + \omega_c$.

In 2001, the ONS used the assumption that birthdays were spread evenly throughout the year at all ages (Duncan et al., 2002). On this basis, a proportion ω_c of those aged $x - 1$ at the census would reach their x^{th} birthday before 30 June 2001, and a proportion $1 - \omega_c$ aged x would remain aged x. The 2001 census was on 29 April, so, for that year, $\omega_c = 62/365$. As a consequence

$$P(T_c + \omega_c, x) = \omega_c P(T_c, x - 1) + (1 - \omega_c)P(T_c, x) - \text{deaths} + \text{net migration}.$$

The assumption of an even distribution of births is reasonable for most birth cohorts when birth rates were reasonably stable from year to year. However, for a few cohorts, this assumption results in a very poor approximation.

Our analyses in Section 3 indicate that the ONS used different assumptions in the

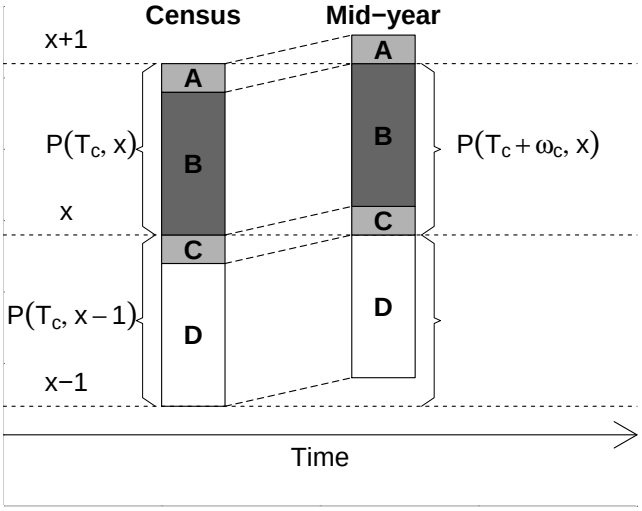

Figure 2: Population progression between a census date, T_c, and the middle of the year, $T_c + \omega_c$.

other census years including 2011.[2] In particular, in 2011, exact dates of birth were used rather than an assumed uniform distribution of birthdays.

2.2.3 The Cohort-Births-Deaths (CBD) Exposures Methodology

We now propose an alternative approach that refines the even distribution of births assumption. Our underlying hypothesis is simple: *at any point in time, t, the pattern of birthdays at age x will reflect the actual pattern of births x years earlier.* At older ages, the relative proportions across a single age needs to reflect intervening mortality. The accuracy of the assumption will depend on the proportion of the population at a given age that are counted in the EW birth statistics in the relevant year of birth, and on any differences between the average birth and patterns of immigrants and those born in EW.

In Figure 3, we plot quarterly birth registrations from 1890 to 1990. Two features are apparent: a clear seasonal pattern that persists to recent times; and significant fluctuations in birth rates over time. In particular, sudden jumps occur in birth rates after the First and Second World Wars, and we argue below that these jumps can have an impact on how mid-year population estimates should be derived from census data.

We can use the births data to estimate what proportion of those aged $x - 1$ at the time of the census will reach their next birthday before 30 June, and what proportion of those aged x at the census will still be x on 30 June. This can be used to answer

[2]This has been confirmed by the ONS (ONS seminar, 9 December 2014).

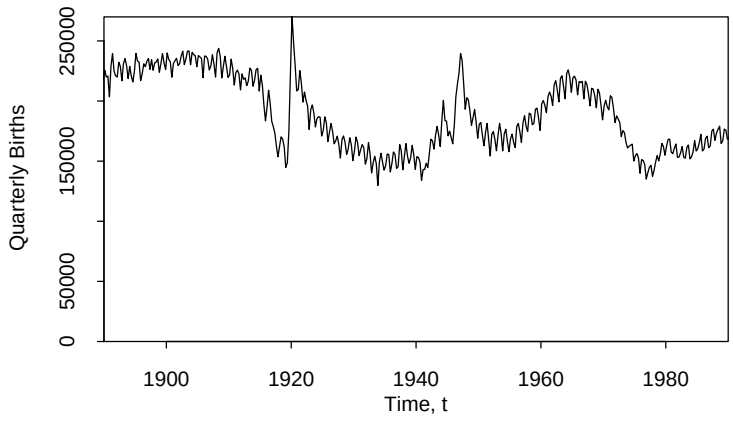

Figure 3: EW quarterly birth registrations from 1890 to 1990.

Group	A	B	C	D
Birth months	5-6/1918	7/1918-4/1919	5-6/1919	7/1919-4/1920
No. of births	113475	524566	99174	752725

Table 1: Numbers of births by birth months from May 1918 to April 1920.

the following question prospectively: *out of those aged* $x-1$ *and* x *at the time of the census, how many are due to be (i.e., ignoring deaths) aged* x *last birthday on 30 June*?[3] We then compare these CBD-based estimates for each age x with the uniform distribution of birthdays used by the ONS in 2001.

To simplify the discussion very slightly let us assume that the 2001 census took place right at the end of April, to allow us to focus on months of birth only. As an example, let us look at males born between May 1918 and April 1920 who were thus aged 82 or 81 at the time of the 2001 census. Table 1 shows the number of births over specified months in 1918 to 1920 corresponding to Groups A to D in Figure 2. At the the time of the 2001 census, 72114 were aged 82 and 115545 were aged 81 (ONS, 2002b). If we make no allowance for deaths or migration, the mid-year population estimates for age 82 under the two methodologies would be:

$$\text{ONS} \qquad \tfrac{10}{12} \times 72114 + \tfrac{2}{12} \times 115545 \quad = \quad 79352$$

$$\text{CBD} \quad \tfrac{524566}{638041} \times 72114 + \tfrac{99174}{851899} \times 115545 \quad = \quad 72741.$$

The figures 72741 (CBD) and 79352 (ONS) make the assumption that there are no deaths or net migration between the census and 30 June. The ONS published mid-

[3]We assume here that quarterly or monthly birth counts are the same as the reported registration counts. The Spanish Flu Epidemic will have caused an upsurge in death registrations which in turn might have delayed birth registrations.

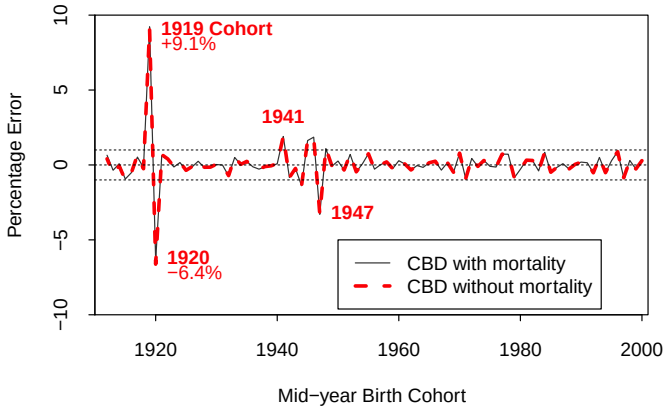

Figure 4: Estimated relative difference between mid-year population estimates by year of birth, for the ONS uniform distribution of births assumption used in 2001, and the CBD methodology. Solid grey line: ONS versus CBD with a stylised Gompertz mortality assumption and no migration. Dashed red line: ONS versus CBD with no mortality or migration. Horizontal dotted lines at -1, 0 and $+1$ are also indicated.

year population at age 82 in 2001 of 78615 includes adjustment for deaths and net migration between the census and 30 June, and so, under the CBD methodology, we would propose to multiply the ONS mid-year population estimate of 78615 by 72741/79352, giving a revised population estimate of 72065 for age 82 on 30 June 2001.

The calculation for age 82 (which we denote the '1919' cohort) in mid 2001 reveals the greatest significant difference. For other cohorts, the difference between the CBD and ONS uniform distribution of deaths assumptions in the 2001 mid-year population estimates is illustrated in Figure 4 (red dashed line). While, the 1919 cohort stands out, the 1920 cohort also has a large difference. There appear to be significant differences between the methodologies used for cohorts born during and after the Second World War; there was a sharp baby boom in 1947, see Figure 3. At most other ages, the difference is at most $\pm 1\%$. However, at high ages, e.g. age 80, an error of 1% in a census year can grow to three times that over the next 10 years (see Figure 1), so even these modest errors are potentially important and deserve our attention.

The CBD methodology can be refined further by adjusting the number of births in Table 1 for mortality between the exact date of birth and the current measurement date. This would have the impact of reducing slightly the *proportion* of 82-year-olds at the census date who were born in May-June 1918. This is illustrated in Figure 4 (thin grey line) using a stylised Gompertz pattern of mortality. It can be seen that adjusting for mortality has little effect at younger ages, and only just starts to

become noticeable amongst the oldest cohorts plotted.

We conclude that the CBD methodology is robust relative to the mortality assumption. However, as already remarked, the CBD methodology makes no allowance for migration. The approach is likely to be less accurate if immigrants form a significant proportion of the population at certain ages, and if the pattern of births for these immigrants is significantly different from EW births. However, for ages 40 and above, the proportion of immigrants into England and Wales in the 2011 census ranged from 18% (ages 40-44) down to 10% (85+), suggesting that migration is unlikely to distort the CBD methodology too much[4]. As a reality check on this conjecture, the pattern of birth dates recorded in the 2011 census could be reconciled against the original birth rates.

2.2.4 $P(t+\frac{1}{2},x) \approx E(t,x)$: How Good is this Approximation?

The true exposures are $E(t,x) = \int_0^1 P(t+s,x)ds$ (equation 2), but, since the ONS publishes only mid-year population estimates, it, in common with many other agencies, makes the assumption that $E(t,x) = P(t+\frac{1}{2},x)$, leading to $m(t,x) = D(t,x)/P(t+\frac{1}{2},x)$. So how good is this approximation? The answer lies in the degree of non-linearity (and, in particular, the degree of convexity) of the underlying continuous population function, $P(t+s,x)$. The approximation will be quite good if the function $P(t+s,x)$ is reasonably linear in the interval $0 < s < 1$. However, if $P(t+s,x)$ is non-linear, then there could be a significant difference between $P(t+\frac{1}{2},x)$ and the true $E(t,x)$.

We can investigate readily how this approximation works in practice by using births data to estimate the population aged 0 at any point in time. Let $\tilde{P}(t,0)$ represent the population at time t aged 0 last birthday, assuming no deaths and net migration. Although, infant mortality was high in the early 20th century (approximately 1 in 10 died in their first year, mostly very soon after birth), the pattern of mortality changes much more slowly over time relative to the variation in birth rates that we seek to analyse here. Additionally, we only make use of age 0 exposures *relative* to the mid-year population counts. The resulting ratios will be relatively insensitive to the underlying rates of infant mortality, provided these change slowly over time.

$\tilde{P}(t,0)$ can be calculated quarterly or monthly (depending on the granularity of births data) and is, simply, the sum of the births in the preceding 12 months and is shown over the period 1910 to 1925 in Figure 5 (left-hand plot, continuous black line). The mid-year population estimates at age 0 assuming no deaths or net migration are shown as crosses. The equivalent exposure estimates are $\tilde{E}(t,0) = \int_0^1 \tilde{P}(t+s,0)ds$ and these are approximated using Simpson's rule based on the quarterly $\tilde{P}(t+s,0)$. These exposure estimates are shown in Figure 5

[4]Source: http://www.nomisweb.co.uk/census/2011 country of birth tables; accessed 11/12/2014.

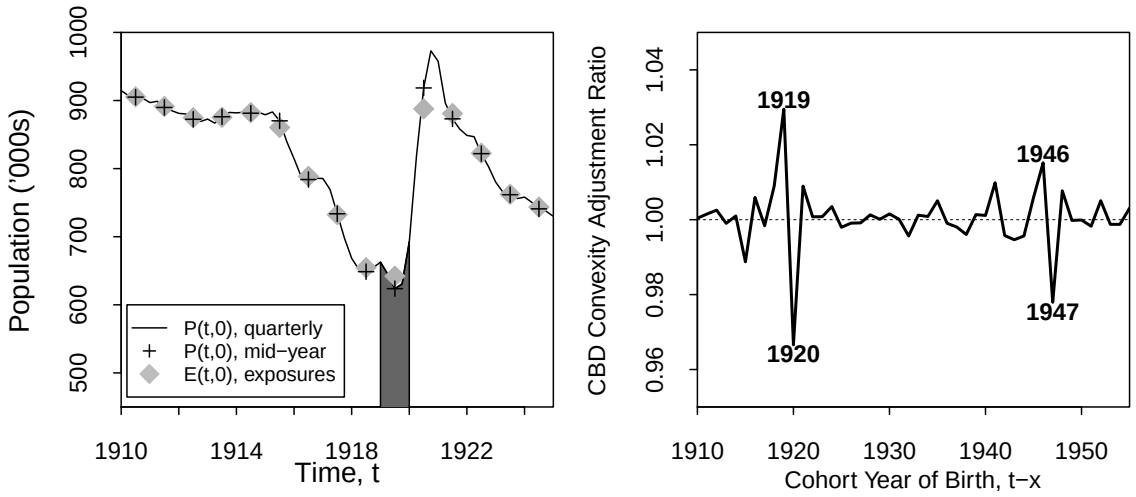

Figure 5: Left: Relationship between age 0 population and exposures estimates. Solid line: quarterly $\tilde{P}(t,0)$. Black crosses: mid-year $\tilde{P}(t+\frac{1}{2},0)$. Red diamonds: exposures $\tilde{E}(t,x)$. Right: The convexity adjustment ratio, $CAR(t-x) = \tilde{E}(t-x,0)/\tilde{P}(t+\frac{1}{2}-x,0)$, as a function of year of birth $t-x$.

(left). Mostly the $\tilde{E}(t,0)$ (grey diamonds) are very close to the mid-year $\tilde{P}(t+\frac{1}{2},0)$ (crosses). However, where there is significant non-linearity, differences can be quite striking. And the most significant non-linearities and, hence, differences occur when there is a sudden baby boom, e.g., at the end of 1919.

Irregular patterns of birth convert into similar patterns of birthdays in later years and so we argue that the relationship between the mid-year population estimate at age x and exposures should be approximately the same as it was at age 0 for the same cohort. We, therefore, propose a convexity adjustment ratio to be applied to mid-year population estimates, $P(t+\frac{1}{2},x)$, namely

$$E(t,x) = P(t+\frac{1}{2},x) \times \frac{\tilde{E}(t-x,0)}{\tilde{P}(t+\frac{1}{2}-x,0)} \tag{3}$$

where we call $CAR(t-x) = \tilde{E}(t-x,0)/\tilde{P}(t+\frac{1}{2}-x,0)$ the Convexity Adjustment Ratio (CAR) for the $t-x$ birth cohort. The CAR is plotted in Figure 5 (right). As we can see, the CAR is generally close to 1 (meaning that, the assumption that $E(t,x) = P(t+\frac{1}{2},x)$ is a reasonable approximation). But in a few years (1919, 1920, 1946, 1947) there is a significant deviation from 1. And these deviations coincide with periods when there were rapid changes in birth rates around the end of the First and Second World Wars.

2.2.5 High-age Methodology

The ONS mid-year population estimates give population counts at individual ages up to age 89 and then a single number each for males and females for age 90 and above. Death counts are available for individual ages above 90. The ONS use the Kannisto-Thatcher (KT) methodology (an adaptation of the extinct cohort methodology; see, Thatcher et al., 2002) to calculate population estimates for individual ages up to age 104 (ONS, 2014a).

The KT methodology attempts to estimate exposures at individual ages from limited underlying data and, therefore, gives rise to the possibility of errors. In particular, we will investigate in the following sections the possibility that there are discontinuity errors at the age 89-90 boundary where the high-age methodology kicks in.

3 Graphical Diagnostics and Signature Plots

In this section, we outline and illustrate a variety of graphical diagnostics that are designed to help verify the quality of the data or identify where anomalies might lie. The use of a graphical diagnostic might contain the following elements. First, a hypothesis that concerns some characteristics of the underlying data. Second, a specified way of plotting some aspect of the data. Third, if the hypothesis is true, then the plot should exhibit a recognisable structure or pattern. Fourth, if the plot does not exhibit the expected structure, then the type of deviation might provide us with pointers on how the underlying hypothesis might be varied or help identify particular types of problem with the data.

3.1 Graphical Diagnostic 1

Hypothesis 1: *Crude death rates by age for successive cohorts should look similar.*

The diagnostic that we will employ to test visually this hypothesis is to plot the curve of crude death rates for each cohort against age. Since we wish to compare cohorts (as required in the hypothesis), we compare, for each cohort, the curve of death rates with its four nearest neighbours.

Examples are presented in the top row of Figure 6 centred on the 1909, 1929 and 1919 birth cohorts. Both the top left (1909) and top centre (1929) plots exhibit the expected pattern: namely that each of the five lines in each plot all follow a consistent upwards trend that is close to linear. Each curve exhibits a certain amount of randomness as age increases. This reflects (a) random variation in the number of deaths given the true underlying death rate (known as Poisson type risk), and (b) randomness in the period effects that drive improvements in the underlying death rates. Poisson risk is likely to be most significant at high ages because of the

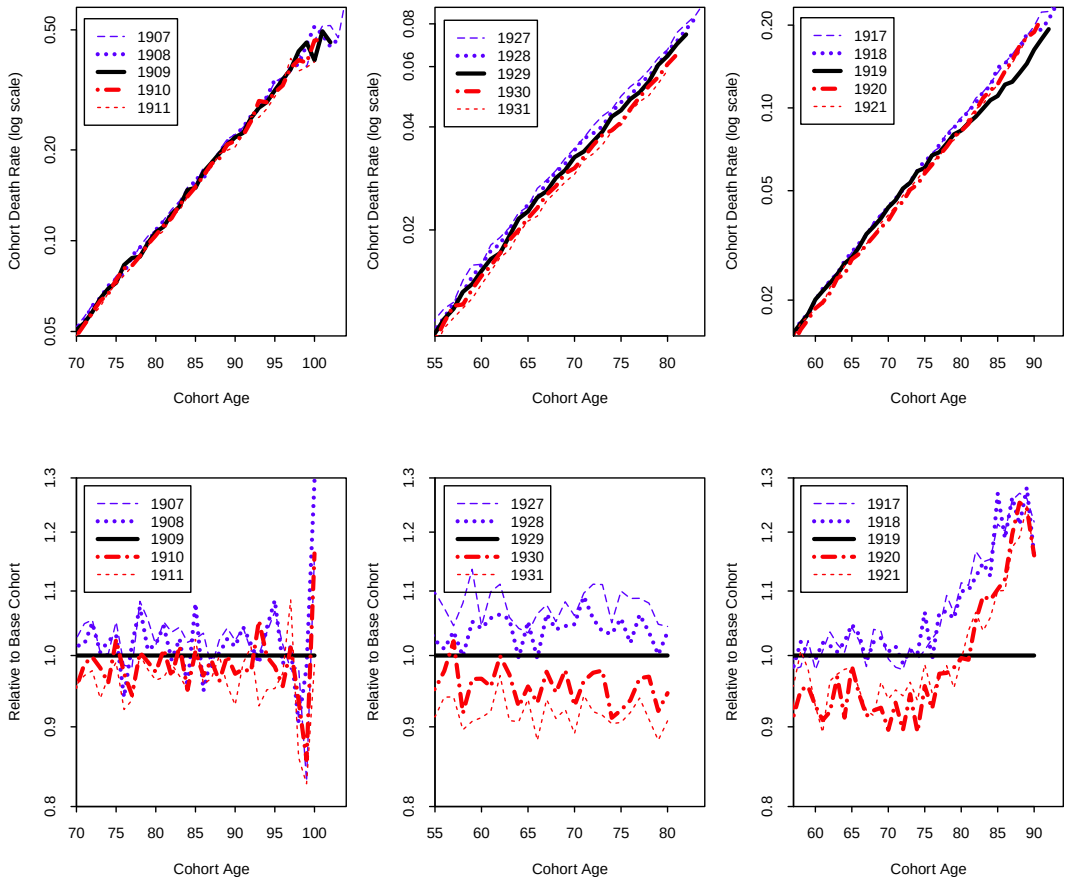

Figure 6: Top row: Cohort death rates by age for the 1907 to 1911 cohorts (left), 1927 to 2031 cohorts (centre) and 1917 to 1921 cohorts (right). The right-hand end point for each curve corresponds to 2011. ONS (revised) EW males data up to 2011. Bottom row: Cohort death rates by age for the 1907 to 1911 cohorts relative to the 1909 cohort (left), 1927 to 1931 cohorts relative to the 1929 cohort (centre) and 1917 to 1921 cohorts relative to the 1919 cohort (right).

small expected number of deaths. However, in the top row of Figure 6 this is really only a problem above age 95: below this age, the underlying shape and relative position of each curve is very clear.

In the top left-hand plot of Figure 6, the lines almost lie on top of each other across the full range of ages. In the top centre plot, the lines are more or less parallel, but the positions of the five curves reflect a greater rate of improvement in mortality rates between these cohorts (see also the lower centre plot). This greater rate of improvement between cohorts born around 1930 is well known (see, for example, Willets, 2004). These plots are typical of most birth cohorts.

Now contrast the 1909 and 1929 cohorts with the 1919 cohort (top right-hand plot, Figure 6). As a graphical diagnostic, this plot does not exhibit the characteristics that we would expect to see. Instead, we can see that the solid black curve for the 1919 cohort after about age 75 bends away from and below its neighbours pointing to anomalies with this cohort. We will discuss this in more detail later, but will just remark here that the pattern is consistent with the emergence of phantoms in the 1919 cohort exposures over the period between the censuses of 1991 (when the 1919 cohort was aged 72) and 2001: in other words, the problem with the 1919 cohort began in 1991. Thereafter, the existence of the phantoms (who never die) causes the curve to drift lower still relative to its neighbours (see Figure 4).

Figure 6 also displays death rates for adjacent cohorts relative to each other: that is, a plot of the ratio $m(t+s+k, x+s)/m(t+s, x+s)$ against age, x, where $m(t+s, x+s)$ represents the death rates over time for the central cohort in each plot. The upper row of Figure 6 is, perhaps, better when the underlying death rates are consistent with the hypothesis. Plotting the ratios in the bottom row is better where the pattern of death rates is not consistent with the hypothesis. Specifically, in the bottom right plot, centred on the 1919 cohort, we can see more easily that the trends change direction at about age 72. Additionally, the bottom-right plot suggests that death rates for the 1919 cohort might have been out by as much as 20% by the time the cohort reached their late 80s.

3.2 Graphical Diagnostics 2A and 2B

Hypothesis 2: *Underlying log death rates are approximately linear within each calendar year*

As a starting point, we calculate how far individual groups of observations (three consecutive ages within a calendar year) deviate from linearity as measured using the empirical concavity function

$$C(t,x) \;=\; \log m(t, x+t) - \frac{1}{2}\Big(\log m(t, x+t-1) + \log m(t, x+t+1)\Big)$$

We then plot this in two ways:

- 2A: plot $C(t, x)$ by cohort: that is, $(x_0 + s, C(t_0 + s, x_0 + s))$ for the $(t_0 - x_0)$ birth cohort;
- 2B: a 2-dimensional heat map of $C(t, x)$ by age and calendar year.

If the hypothesis is true, the concavity function should stay close to zero without exhibiting any systematic bias above or below zero.

Some typical results for graphical diagnostic 2A are plotted in Figure 7. The pattern exhibited in the plot for the 1924 cohort is consistent with the hypothesis: the dots are randomly above and below zero. However, the plots for the 1920, 1919, and 1947 cohorts all contain structure that is not consistent with the hypothesis. So we consider next what the irregular patterns tell us about the underlying data.

The 1919 and 1920 cohort plots each contain two specific effects. Before 1991, the plots are level but are consistently above (1919) or below (1920) the zero line. After 1992, both plots suddenly start to drift at a steady rate. The shift up or down before 1991 is consistent with our earlier discussion in Section 2.2.4, where we argued that mid-year population estimates were not always good approximations to the true exposures for some cohorts. In the 1919 and 1920 cohort plots, we have added a horizontal dashed line which makes an allowance for this effect. The adjustment means that, up to 1991, the plots of the empirical concavity function now look more reasonable. However, the positioning of the dashed line in both cases could be improved suggesting that other factors are at play beyond the proposed Convexity Adjustment Ratio. However, up to 1991, both plots are consistent with a systematic bias in the estimate of cohort exposures.

After 1992, both plots drift steadily down or up. This is consistent with the introduction of phantoms into the 1919 cohort and trolls into the 1920 cohort from 1992 onwards. We infer from this that the methodology for shifting from a census to the mid-year population estimate discussed in Section 2.2.2 was applied at the time of the 2001 census but not in previous census years. Furthermore, the apparent inconsistencies revealed in 2001 as a result of applying this methodology then led the ONS to backfill by cohort over the period 1992 to 2001 to provide a smoother progression of population estimates between the 1991 and 2001 censuses. In other words, the ONS, having introduced an error in 2001, unknowingly exacerbated the problem by spreading the error back to 1991. This will be discussed further when we consider graphical diagnostic 3 below.

The plot for the 1947 cohort is generally more dispersed. The reason for this is that the cohort is younger and, therefore, the underlying death counts are much smaller and more random. However, we can still see that the dots are biased below zero, to an extent that is consistent with a requirement to apply the Convexity Adjustment Ratio as indicated by the horizontal dashed line.

We move on to graphical diagnostic 2B and plot the empirical concavity function in two dimensions in the form of a heat map in Figure 8. Apart from some obvious

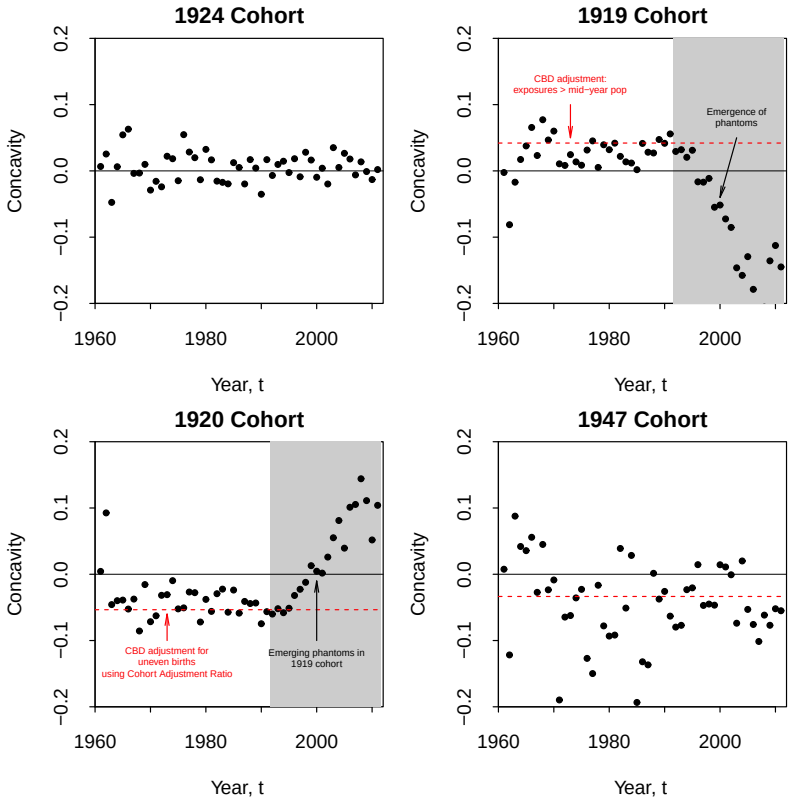

Figure 7: Empirical concavity function, $C(t+s, x+s)$, plotted against calendar year $t+s$ for birth cohorts $t-x$ =1924, 1919, 1920 and 1947. Shaded region covers the calendar years 1992 to 2011.

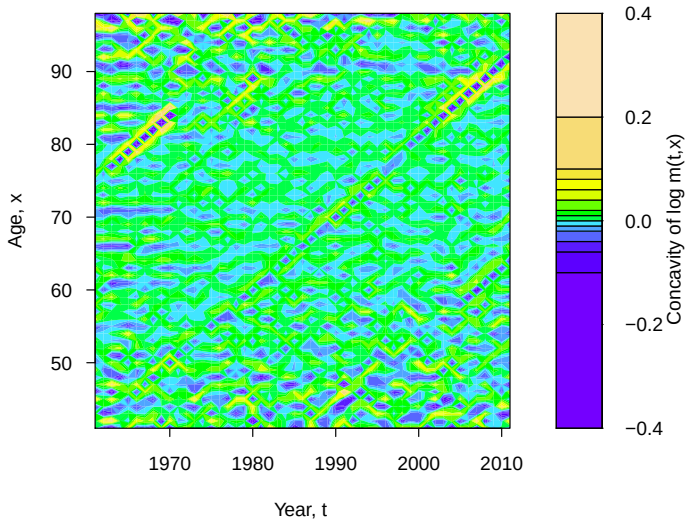

Figure 8: The empirical concavity function $C(t, x)$ for different years, t, and ages x.

structural features, we can also see greater randomness at the high ages and low ages caused by low death counts. We could normalise the empirical concavity by dividing by its standard error. We do not report the results, but the effect is to reduce the noise observable in Figure 8 at both low and high ages, while leaving the remaining structure clearly visible.

If the hypothesis about linearity in the log deaths rates were true, then the heat map should essentially be completely random apart from some autocorrelation between cells within each calendar year as a result of the common contributors to, for example, $C(t, x)$ and $C(t, x+1)$. While Figure 8 does contain plenty of randomness, it also contains significant structure. The clear diagonals reflect the cohort biases, as, for example, with the 1919, 1920 and 1947 cohorts, as well as one or two others. But there are other structural elements that are new. For example, there is evidence of clear horizontal traces. In most cases, these horizontal bands are consistent with small biases in the reporting of age at death. In general, these bands tend to be more prominent towards the left suggesting that the reporting of age at death has become steadily more accurate as time progresses.

The horizontal bands around age 90, suggest that deaths reported at age 90 and 91 are too *low* rather than too high. Furthermore, this band is consistent across all years. So we need to look for an alternative explanation to bias in the reporting of age at death. Instead, we recall the fact that exposures at age 90 and above are calculated using the KT methodology instead of the standard census and intercensal methodologies used at all ages up to 89. Our simple graphical diagnostic suggests that the current way in which the KT methodology is employed potentially introduces a small but systematic bias around age 90.

3.3 Graphical Diagnostic 3

Hypothesis 3: *Changes in cohort population sizes should match closely the pattern of reported deaths.*

Recall that we have two types of data: death counts, $D(t, x)$, and mid-year population estimates, $P(t + \frac{1}{2}, x)$. We will now define the decrement

$$\hat{d}(t + 0.5, x) = P(t + 0.5, x) - P(t + 1.5, x + 1)$$

which equals the sum of the deaths from the cohort aged x last birthday at time $t+\frac{1}{2}$, denoted by $d(t + \frac{1}{2}, x)$, less net migration, $M(t + \frac{1}{2}, x)$. At high ages, net migration will be modest compared to deaths and so $\hat{d}(t+\frac{1}{2}, x)$ will be approximately equal to $d(t + \frac{1}{2}, x)$. It is natural, therefore, to compare $\hat{d}(t + \frac{1}{2}, x)$ with the reported death counts. However, four different reported death counts contribute to the inferred decrement, $\hat{d}(t+\frac{1}{2}, x)$ (namely, $D(t, x)$, $D(t, x+1)$, $D(t+1, x)$ and $D(t+1, x+1)$).

The graphical diagnostic compares these related measures and is constructed as

follows for the $t - x$ cohort:

- Plot (solid dots) $\hat{d}(t+\frac{1}{2}, x) \times CAR(t-x)$, where $CAR(t-x) = E(t-x, 0)/P(t+\frac{1}{2} - x, 0)$ (see Section 2.2.4) to adjust for differences between the mid-year population and the exposures.
- Plot (black and blue lines respectively) $D(t, x)$ and $D(t+1, x+1)$ with no adjustment.
- Plot (green line) $D(t, x+1) \times E(t-x, 0)/E(t-x-1, 0)$, to adjust for different birth rates for different cohorts.
- Plot (red line) $D(t+1, x) \times E(t-x, 0)/E(t+1-x, 0)$ (ditto).

Figure 9 shows typical examples of how the plot should look if the hypothesis is true. Specifically, the black dots and the four lines for the reported deaths should all follow each other quite closely.[5] This is, indeed, the case, indicating that the adjustment for unequal births is about right. Without this adjustment, some curves would be significantly higher or lower than the central case. The similar shape for each curve links to Hypothesis 1 underpinning Graphical Diagnostic 1. At younger ages and earlier years, the decrements $\hat{d}(t+\frac{1}{2}, x)$ are rather more variable, reflecting year to year variations in net migration.

Closer inspection reveals some unusual effects that indicate that there are issues with the population data. For example, for the 1914 and 1916 cohorts, the decrement at age 90 is significantly and systematically below where we would expect it to be. This provides support for our conclusions under Graphical Diagnostic 2B that the alternative methodology for estimating exposures at age 90 and above needs revision. For the 1914 cohort, for example, the decrement at age 90 is about 1500 below where it should be. The mid-year population at age 90 is 24,842 and so the excess of 1500 equates to 6% of the cohort size at that age. This again consistent with our conclusions after Graphical Diagnostic 2B.

The plots in Figure 10 show examples where the plot is more obviously not consistent with the hypothesis. For the 1919 and 1947 cohorts, we can see that in the years 1992 to 2001 inclusive (grey bar), the dots are significantly and systematically below (1919) or above (1947) the reported curves of deaths. The 1919 cohort and the left-hand 1947 cohort plots are based on the inter-censal population estimates as revised after the 2011 census. Now compare the left-hand 1947 plot with the right-hand 1947 plot based on the post-2001-censal population estimates. Differences only exist in the years 2002 to 2010, and we can see that the inter-censal values (middle plot) have been subjected to what appears to be a parallel shift down relative to the original post-censal values (right-hand plot). This shift down suggests that the previous

[5] It is not necessary to be able to distinguish between the four lines. Their similarity is one of the keys to why this diagnostic is effective.

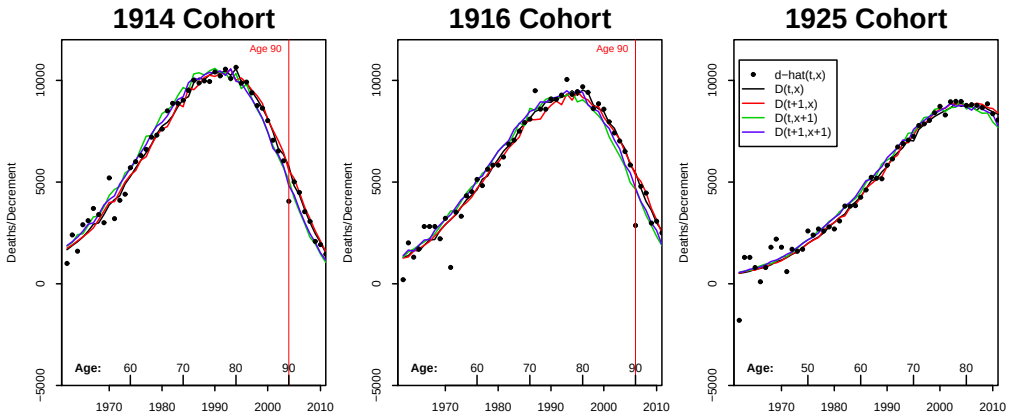

Figure 9: Death curves for the 1914 (left), 1916 (centre) and 1925 (right) cohorts by calendar year or, alternatively, by age (black and blue curves), and dots (cohort population decrements multiplied by the Convexity Adjustment Ratio). Red and green curves: death counts in adjacent cohorts adjusted for unequal cohort sizes.

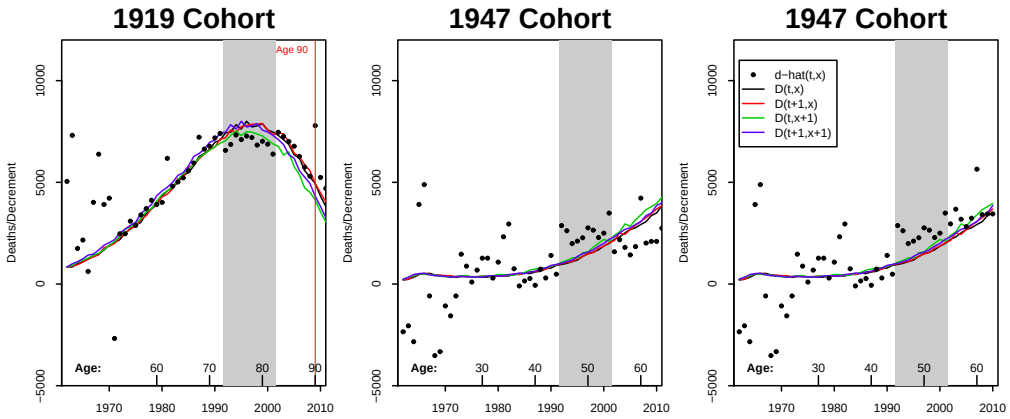

Figure 10: As Figure 9. Additionally, the left and middle plots use the revised ONS data published in 2012 for years 1961 to 2011. The right plot uses the post-2001-censal ONS data for 1961 to 2010.

error over the 1992-2001 period has been subjected to a reversal following the 2011 census, as we have previously discussed.

Similar patterns of parallel shifts exist for other cohorts including 1920, 1944, 1945 and 1946, amongst others. However, the 1992-2001 and 2002-2010 cohort shifts are not always similar to the 1947 cohort shifts.

These parallel shifts up and down are consistent with the following sequence of events:

- The census to mid-year population adjustments up to 1991 seem to have produced reasonable results. The methodology used is not reported;
- The methodology for adjusting from census to mid-year outlined in Section 2.2.2 was applied in 2001;
- Differences between the resulting mid-year population estimates and the post-1991-censal population estimates were then backfilled evenly over the years 1992 to 2001, producing the *signature* parallel shift down or up;
- The census to mid-year methodology applied in 2011 might have changed back, but it is difficult to establish from the data, as the older 1919 and 1920 cohorts are now above age 90 and subject to the 90+ methodology for estimating mid-year populations;
- Adjustments in 2011 attributable to those born in the 1940's might be due to post-censal errors in assumed net migration.

The method of backfilling employed by the ONS in 2001 (Duncan et al., 2002) is illustrated in a stylised way in Figure 11. It produces a signature plot for Graphical Diagnostic 1 for the 1919 cohort (Figure 6): the build up of phantoms over the years 1992 to 2001 causes death rates for the 1919 cohort to drift downwards relative to neighbouring cohorts.

For the 1919 cohort, the difference between the CBD census-to-mid-year methodology and the methodology employed by the ONS amounted to 9.1% (Figure 4) or a potential error of about 6600 people. If this error did not exist in 1991 and was spread evenly over the years 1992 to 2001 then the adjustment would be 660 per year: this is entirely consistent with the magnitude of the parallel shift that we observe in Figure 10 for the 1919 cohort.

3.4 Summary

The combination of these three graphical diagnostics reveal that there remain significant anomalies in the England & Wales males population data and how it is

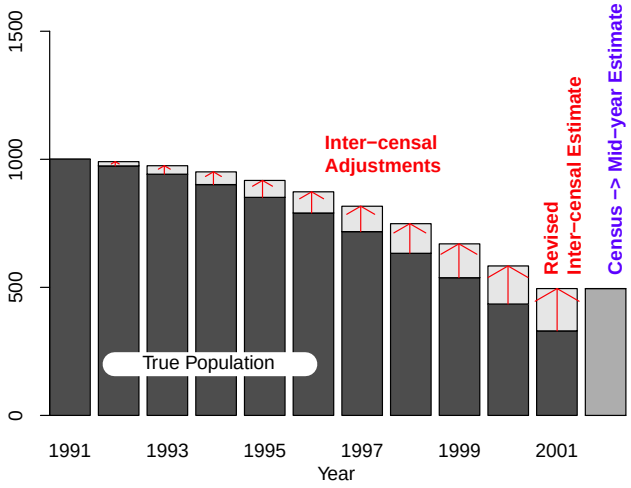

Figure 11: Stylised example of backfilling. Black bars: true population equal post-1991-censal population estimates. Dark grey bar: Mid-year population estimate following 2001 census to mid-year adjustment. Light grey bars: inter-censal adjustments build up evenly over the preceding 10 years, with zero adjustment to 1991.

used to construct crude death rates. Furthermore, although the three diagnostics are qualitatively different, they produce consistent findings.

The 1919 and 1920 cohorts have been known for some time to have unusual patterns of mortality. The graphical diagnostics indicate that the anomalies are the result of unusual patterns of births and its interaction with the methodology employed in any given census year to produce mid-year population estimates.

The three diagnostics produce a stronger set of conclusions than would be possible using only one. Sometimes an anomaly will reveal itself clearly in only one of the three diagnostics and might go unobserved in the other two.

4 Model-Based Analysis of Historical Population Data

In Section 3, our use of a variety of graphical diagnostic tools allowed us to identify a number of anomalies in the England & Wales males population data. However, these tools do not allow us to quantify how big the individual anomalies are. To overcome this, we propose a methodology that allows us to quantify the individual errors and outputs a revised set of exposures that we believe is as free from anomalies as possible.

The methodology we propose uses only given sets of *exposures* and deaths data. Here, exposures have been equated to mid-year population estimates as employed

by the ONS; alternative methodologies for calculating exposures have been proposed (for example, the Human Mortality Database; Wilmoth et al., 2007) that derive exposures from population and deaths data in a more sophisticated way. In addition, it would be possible to exploit monthly or quarterly births data as we did in our discussion of Graphical Diagnostic 2. However, we seek to develop an approach that can be applied to a wide variety of datasets, many of which will not have births data in addition to mid-year population and deaths data.

4.1 Underlying Modelling Assumptions

In seeking to model errors in exposures, we will be guided by three assumptions:

- Assumption A: death counts are accurate;
- Assumption B: exposures are subject to errors, and errors following cohorts are correlated through time;
- Assumption C: within each calendar year the curve of underlying death rates is "smooth".

Based on these assumptions, we adjust exposures to achieve a balance between B and C. In order to make these adjustments, we need to translate these assumptions into a model for the errors in the exposures.

4.2 Model for Errors in Exposures

We summarise here the main characteristics of the proposed model, with further details provided in Appendix A. We use the following definitions and notation:

$$\begin{aligned}
\hat{E}(t,x) &= \text{ONS published exposures} \\
E(t,x) &= \text{corresponding true but unobservable exposures} \\
D(t,x) &= \text{corresponding death count (assumed accurate)} \\
m(t,x) &= \text{true, but unobservable death rate} \\
\epsilon(t,x) &= \log E(t,x), \quad \text{with } \textit{a priori} \text{ mean } \hat{\epsilon}(t,x) \\
Y(t,x) &= \log m(t,x) \\
\phi(t,x) &= \epsilon(t,x) - \hat{\epsilon}(t,x).
\end{aligned}$$

The $\hat{\epsilon}(t,x)$ are chosen so that the *a priori* mean of $E(t,x)$ equals $\hat{E}(t,x)$: in other words, we assume that there are no systematic biases in the ONS estimates.

Conditional on $m(t,x)$ and $E(t,x)$, deaths, $D(t,x)$, have a Poisson distribution with mean and variance equal to $m(t,x)E(t,x)$. Assumption B led us to model

each $\phi(t+s, x+s)$, given (t, x), as an AR(1) process. Assumption C (after some experimentation) led us to model, for fixed t, the $Y(t, x)$ as an ARIMA(0,3,0) process with low-variance innovations. With a low variance, the choice of model means that quadratic curvature is acceptable in terms of "smoothness". The choice of process for $Y(t, x)$ is discussed further in Appendix A.2.

Our use of a Bayesian framework brings a number of advantages. As described Above, it provides a natural setting in which we can build in our prior beliefs about the error structure. The model plus data then allow us to estimate the true exposures, as well as provide a full posterior distribution for the true exposures, the true death rates, and the associated process parameters. This can then be used to make mortality forecasts that incorporate uncertainty in exposures in a very straightforward way.

The latent state variables and process parameters are estimated using Markov chain Monte Carlo (MCMC) and the Gibbs sampler (see Appendix A). This produces a vector-valued Markov chain, $M(i)$ under which, for each i, $M(i)$ records a single drawing of the latent state variables $\phi_i(t, x)$ and $Y_i(t, x)$ from the posterior distribution. This means that not only can we obtain a point estimate of each error in the exposures, but the Markov chain also gives us additional information about how much uncertainty there is around these point estimates of the errors.

Figure 12 provides a heat plot of the posterior means of the errors, $\phi(t, x)$. Blues indicate that the true exposures are lower than the ONS estimates; greens and yellows indicate that exposures need to be adjusted upwards. Mostly the errors are quite small. However, we have highlighted some (but not all) regions of the plot where errors are significantly different from zero and consistently positive or negative. Most of these significant regions lie on diagonals that follow specific cohorts. The 1919 and 1886 cohorts exhibit the strongest negative errors, while the strongest positive errors are for the 1900 cohort, with others associated with individual cohorts born in the 1940's, 50's and 60's. The pattern of errors for the 1919 and 1920 cohorts are entirely consistent with the conclusions that we drew from Section 3. While the earlier cohorts are either extinct or close to being so, the younger cohorts are still large in number, so any measurement errors associated with them will be materially significant.

The Bayesian analysis was repeated using exposures that had been adjusted for the uneven pattern of births using the Convexity Adjustment Ratio (CAR), as outlined in Section 2.2.4. The result of this is shown in Figure 13. The most obvious impact is that the significant negative errors for the 1920 cohort and positive errors for the 1919 cohort up to the early 1990's have now been eliminated. However, other errors, although modified, persist. For example, for the 1919 cohort in the last 15 years, errors are still significant and, from our investigation in Section 3, these errors can be attributed to a change in census-to-mid-year methodology in 2001. For the younger cohorts, the more extreme errors observable in Figure 12 have been

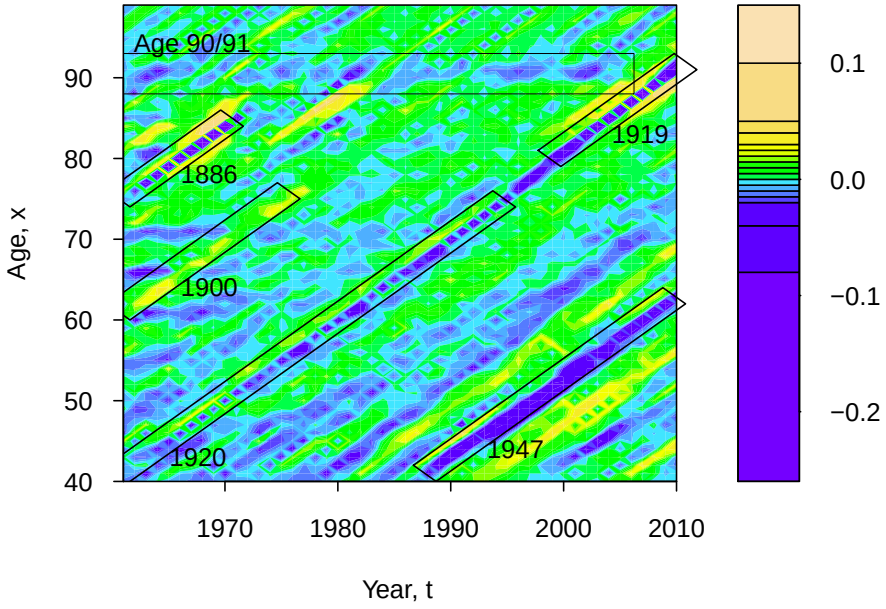

Figure 12: Heat plot of the posterior mean of the exposures error function, $\phi(t, x)$. Exposures are assumed to be equal to the mid-year population estimates.

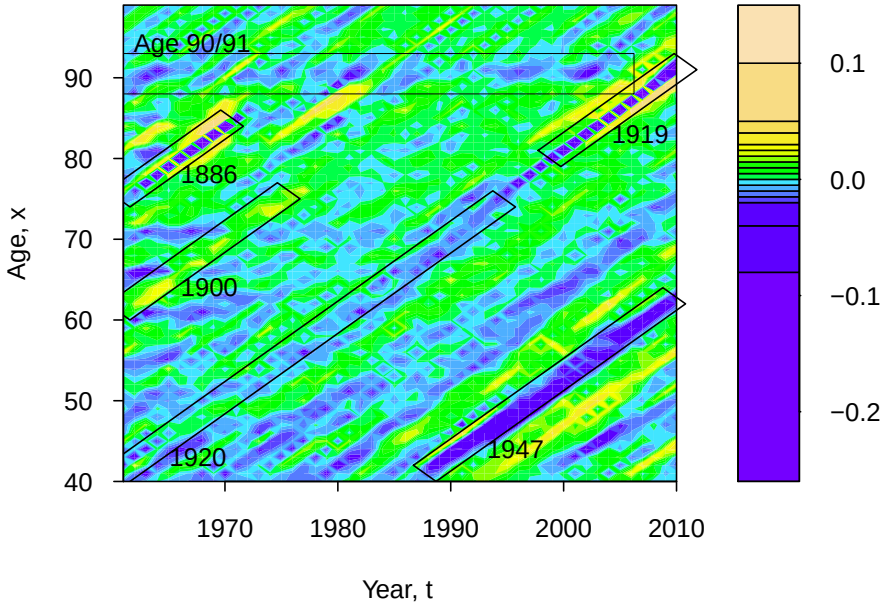

Figure 13: Heat plot of the posterior mean of the exposures error function, $\phi(t, x)$. Exposures are assumed to be equal to the mid-year population estimates multiplied by the Convexity Adjustment Ratio.

mitigated slightly, but clearly there are additional problems with the population estimates for some of these cohorts. Adjusting for these types of errors is less easy, since, at least in some cases, they reflect changes from time to time in methodology. In contrast, the Convexity Adjustment Ratio provides a straightforward, objective and constant adjustment to the exposures for each cohort. Further, a key feature of this section is that the simple model for errors and the Bayesian methodology provides us with an objective procedure for estimating the true exposures that was missing from Section 3.

Figure 12 also exhibits some persistent but small negative errors around ages 90 and 91, that are consistent with the concavity function in Figure 8. It confirms observations in Section 3 that there are some modest issues that need to be addressed in the high-age methodology for estimating population and exposures above age 90. The band around age 90 is the only one that persists across all years.

5 Impact of Exposure Errors on Future Mortality Projections and Annuity Calculations

We next consider what the impact of these errors is on future mortality-contingent problems. We will compare two cases: model-based forecasts using the unadjusted ONS exposures data (2012 revisions); and model-based forecasts using adjusted exposures data outlined in Section 4, incorporating uncertainty in the estimated exposure errors.

By way of example, we use the following model (model M7 in Cairns et al., 2009):

$$\text{logit } q(t,x) \;=\; \kappa^{(1)}(t) + \kappa^{(2)}(t)(x-\bar{x}) + \kappa^{(3)}(t)\left((x-\bar{x})^2 - \sigma_X^2\right) + \gamma^{(4)}(t-x) \quad (4)$$

where the mortality rate $q(t,x)$ is assumed to be equal to $1-\exp[-m(t,x)]$, $x = x_0,\ldots,x_1$, $\bar{x} = (x_1-x_0)/2$ is the mean age, $\sigma_X^2 = n^{-1}\sum_x (x-\bar{x})^2$, the period effects, $(\kappa^{(1)}(t),\kappa^{(2)}(t),\kappa^{(3)}(t))$, are assumed to follow a multivariate random walk, and the cohort effect, $\gamma^{(4)}(t-x)$, is assumed to be an AR(1) process independent of the period effects.

N stochastic scenarios incorporating uncertainty in $\phi(t,x)$ were generated. In Figure 14, we compare mortality fan charts for unadjusted (green) and adjusted (grey) exposures data. Broadly speaking, the fans are quite similar in terms of central trajectory and spread. The similar spread indicates that parameter uncertainty in the adjusted exposures is not that large compared to uncertainty in future period and cohort effects. However, in terms of the detail, the grey fans are quite smooth, while the green fans are relatively lumpy. This lumpiness is linked to the estimated cohort effect (see Figure 15). Specifically, the blips associated with the progress through time of the 1919 and 1947 cohorts (Figure 14) correspond to jumps in the fitted cohort effect in model M7. The cohort effect is shown in Figure 15. The red

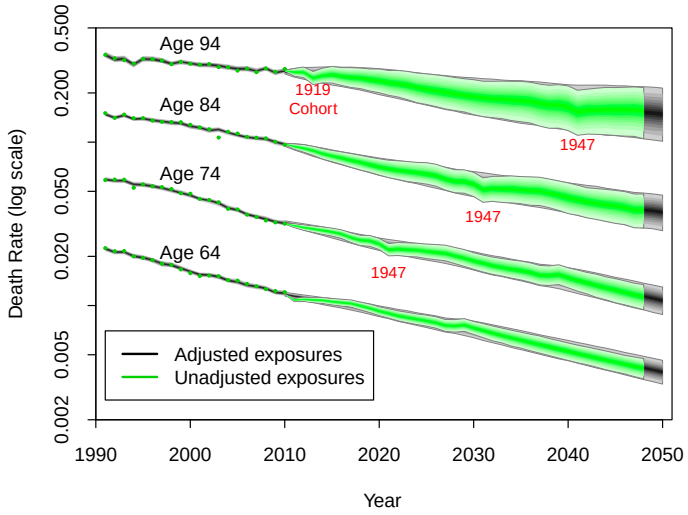

Figure 14: Mortality fan charts for ages 64, 74, 84 and 94 for the stochastic mortality model (equation 4). Green dots and fans: historical data and forecasts (90% prediction interval) using the unadjusted ONS exposures data. Grey fans up to 2010: adjusted exposures incorporating parameter uncertainty. Grey fans after 2010: mortality forecasts using adjusted exposures data with parameter uncertainty. Forecasts use data from 1991 to 2010.

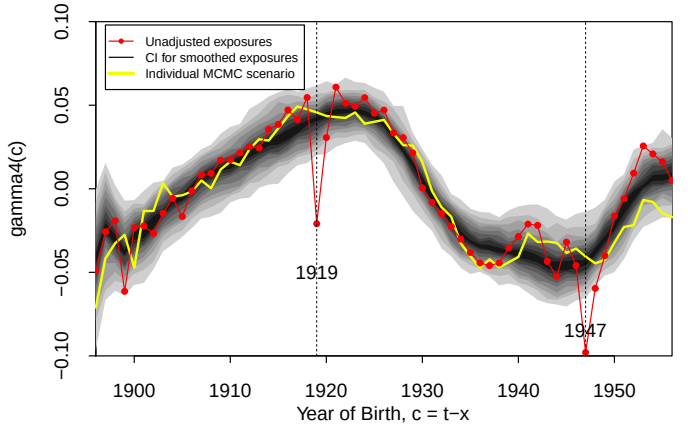

Figure 15: Cohort effect, $\gamma^{(4)}(c)$ for the stochastic mortality model (equation 4) using unadjusted (solid red line with dots) and adjusted (grey fan and solid yellow line without dots) exposures data. Grey fan: 90% credibility interval (bounded by the 5% to the 95% quantiles of the posterior distribution) for $\gamma^{(4)}(c)$. Solid line without dots: a typical realisation of $\gamma^{(4)}(c)$ from the posterior distribution.

line with dots shows the cohort effect using the unadjusted exposures data. Mostly this is fairly smooth, but there are extreme jumps associated with the 1919 and 1947 cohorts. When the model is refitted to adjusted exposures data, the fitted cohort effect ceases to have these extreme jumps (see, for example, the solid yellow line). Otherwise the fitted cohort effect has a similar level of smoothness as before.

5.1 Annuity values and cohort effects

While the modelling of mortality rates are of interest in their own right, their real significance is that they provide a fundamental building block for other mortality linked quantities, both demographic and financial. An example that combines both is an annuity of £1 per annum from the end of year $t-1$ (exact time t) payable annually in arrears to a male aged exactly x at the end of year $t-1$. The expected present value of this annuity is

$$\begin{aligned} a(t,x) &= \sum_{s=1}^{\infty}(1+r)^{-s}E[S(t,s,x)|\mathcal{M}_t] \qquad (5) \\ \text{where } S(t,s,x) &= \prod_{j=0}^{s-1}(1-q(t+j,x+j)) \end{aligned}$$

is a survivor index that represents the proportion of males aged x at exact time t who survive to exact time $t+s$, and $\mathcal{M}_t$ represents the mortality information available at time t. An interest rate of $r = 2\%$ has been assumed.

Annuity values, calculated for the end of 2010, are tabulated in Table 2 for selected cohorts. Mean values are presented using both unadjusted and adjusted exposures data, along with the standard deviation of the random present value of the annuity (per £1 of annuity for a large population). We can see that while the 1919 cohort has the largest proportional change, the 1947 cohort has the largest absolute change, reflecting the fact that this cohort will live much longer after 2010. On the other hand, the impact of uncertain survivorship for this cohort is more distant and so its effect is dampened by the discount factor.

The 1919 cohort is the most interesting visually in much of our analyses. However, for the majority of pension plans and annuity providers, cohorts born in the 1940's and 1950's for whom we also detect significant (although smaller than the 1919) anomalies should be of most concern from a financial perspective. Additionally, although price impacts are currently modest, we should nevertheless take account of them, and monitor the development of the anomalies over time.

Finally, consider the impact on annuity values and its relation to the impact on cohort effects. In Figure 16, we show, as a grey fan, the difference between the cohort effects after and before adjustment (grey fan minus the red line with dots in Figure 15), to demonstrate the impact of the adjustment on cohort effects. Generally

Age		PV: Unadjusted		PV: Adjusted	Percentage
x	Cohort	Mean	St.Dev.	Mean	Difference
63	1948	17.40	2.2%	17.32	+0.5%
64	1947	17.00	2.3%	16.74	+1.5%
65	1946	16.12	2.4%	16.14	−0.1%
92	1919	3.04	4.2%	2.91	+4.0%

Table 2: Mean of annuity present values for various cohorts from the end of 2010, assuming $r = 2\%$, using unadjusted and adjusted exposures. Age = age at the end of 2010. Standard deviations of $PV = \sum_{s=1}^{\infty}(1+r)^{-s}S(t,s,x)|\mathcal{M}_t$ for t = end 2010 using unadjusted exposures are also presented (as a percentage of the mean). Mean present values are shown before and after adjustment of the exposures.

the differences are quite smooth and centred close to zero. However, there are some significant jumps in Figure 16 associated with the 1919 and 1947 cohorts. We then compare these differences with the impact on annuity values (Figure 16, solid purple line). To illustrate this, Figure 17 shows, by reference to cohort survivorship fan charts, why the annuity price changes by so much. Over a period of only 8 years, the proportion of survivors is more than 10% lower using the adjusted exposures. From Figure 16, we can see that the impact on annuity values is strongly correlated, as we would expect, with the impact on cohort effects.

6 Conclusions

The 2011 census revisions to England & Wales population estimates by the ONS drew attention to the possibility that there are widespread errors in how population data are measured and reported in many countries. In particular, we have discovered that an uneven pattern of births within a given calendar year is a major cause of errors in population and exposures data. Different countries or agencies might derive population and exposures in different ways. But however they do this, the estimates will be subject to potentially significant errors, unless they take into account monthly or quarterly births data.

We have developed a range of methodologies to help identify specific errors in population, exposures and deaths data:

- graphical diagnostics providing a powerful model-free toolkit for identifying anomalies in the form of signature plots;
- a Bayesian framework enabling the size of these anomalies to be quantified;
- two-dimensional diagnostics enabling the detection of small systematic errors

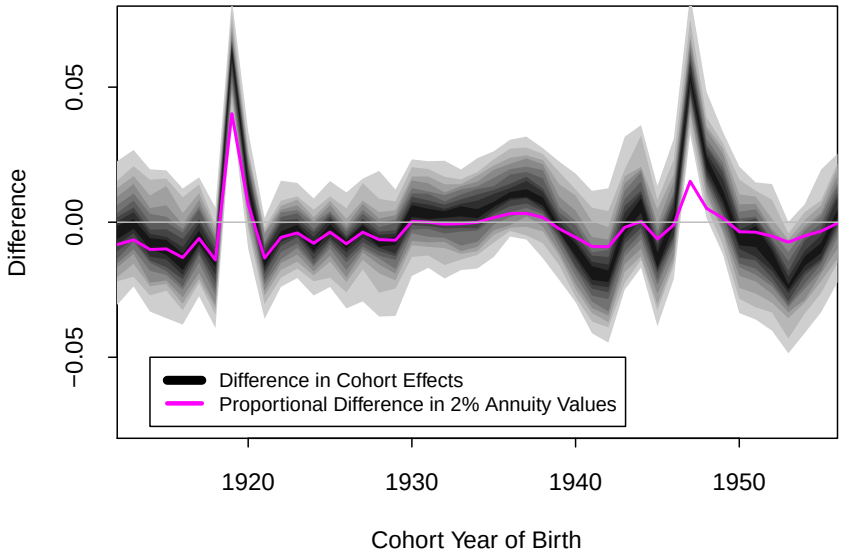

Figure 16: Comparison of changes in the estimated cohort effects versus changes in annuity prices. Grey fan: distribution of the difference between the estimated cohort effect, $\gamma^{(4)}(c)$, before and after the adjustment of the exposures. Purple line: Impact of the adjustment to exposures on annuity values for different cohorts at the end of 2010 using $r = 2\%$ as the interest rate.

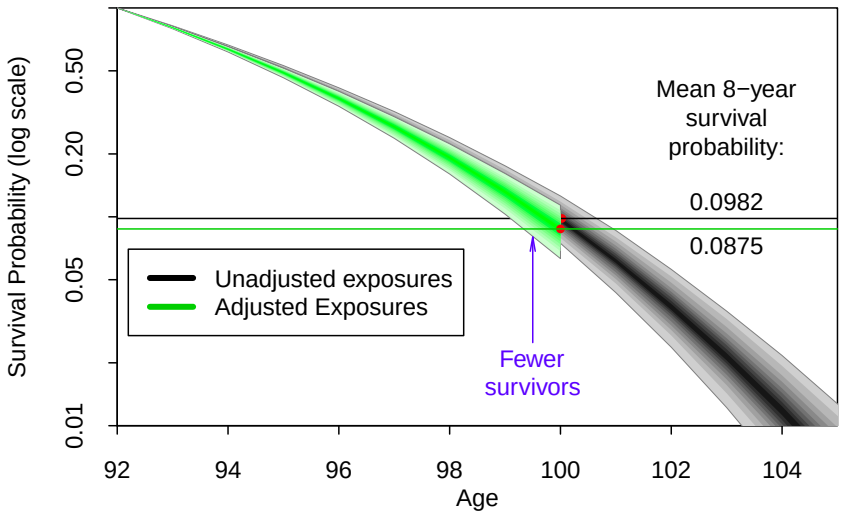

Figure 17: Survivor fan charts under the stochastic mortality model (equation 4) for the 1919 cohort from the end of 2010 before (grey fan) and after (green) adjusting the exposures.

in exposures and deaths of less than 1%.

Apart from analysis of historical data, our recommendation is that these methodologies should be used on an annual basis as an early warning system for the detection of emerging anomalies in data.

Our analysis using these methodologies has shown that errors remain in ONS population data. We have developed the Cohort-Births-Deaths (CBD) Exposures Methodology which can be used to explain many of the bigger errors. The first component was the convexity adjustment ratio which can be used to explain how persistent cohort-related errors arise when exposures are equated to mid-year population estimates. The second component was the improved methodology for deriving mid-year population estimates in census years from census data. Other errors were identified but only partially explained by the CBD Exposures Methodology. In addition, anomalies in death rates have been identified that we believe are due to: potential small biases in the reported age at death; and use of the Kannisto-Thatcher high-age methodology resulting in a discontinuity at age 90.

Collectively, these errors can make substantial differences, particularly in respect of cohorts that are still large enough to have a significant financial impact, an example being the 1947 cohort within an annuity portfolio.

The same sources of errors – with possible variants – will apply to other countries. Some countries are similar to England & Wales in that they derive their population data from periodic (typically decadal) censuses. As one example, data collection in the US shares similar characteristics and reveals similar anomalies to England & Wales. Many other countries, however, have compulsory systems of registration, which means that both deaths and population counts are more accurate, even if they are not completely error-free. Nevertheless, there is still the potential for errors to be introduced when the crude death rates are calculated. As an example, a country might publish accurate start-of-year population data. But, if the central exposed to risk involves taking an arithmetic average of the population at the beginning and end of each year (possibly with further adjustments as in the HMD methods protocol; Wilmoth et al., 2007), then errors can be introduced as a result of an uneven pattern of births. The implication of all this is that definitions of published data need to be read carefully and published numbers that are exact should be used wherever possible, while population figures that are derived from these (possibly by other agencies) should be used with caution.

The graphical diagnostics and Bayesian quantification of errors can be just as easily applied to assess the quality of data for other countries. This includes the rich and extensive database maintained by the Human Mortality Database (HMD). The HMD, as outlined earlier, processes data provided by national statistics agencies in a standardised way that does not make use of monthly or quarterly births data. Although there is no space here to report on individual analyses, we can remark

that many countries exhibit similar and some alternative patterns of anomalies to those in the England & Wales.

Acknowledgements

We would like to thank the ONS for providing the deaths and population data and for providing feedback at an extended in-house seminar (particularly Julie Jefferies, Angele Storey, Steve Smallwood, Adrian Gallop, Julie Mills and Jo Zumpe); Stephen Richards for providing the births data and for contributing the idea that the 1919 birth cohorts anomaly might be connected to births, and Andrew Hunt for comments on an earlier version of the paper. We would also like to thank seminar participants at the US Social Security Administration (particularly Steve Goss, Karen Glenn, Michael Morris and Alice Wade), the Institutional Investor Breakfast Meeting, Harvard Club of New York City (particularly Peter Nakada), the United Nations Population Directorate (particularly Kirill Andreev), and the Continuous Mortality Investigation (particularly Jon Palin). Finally, we thank the associate editor and two anonymous referees for their detailed and very helpful suggestions to improve the paper. Cairns acknowledges financial support from Netspar under project LMVP 2012.03.

References

Biagini, F., Rheinländer, T., and Widenmann, J. (2013) Hedging mortality claims with longevity bonds. *ASTIN Bulletin*, 43: 123-157.

Blake, D., Cairns, A.J.G., and Dowd, K. (2006) Living with mortality: Longevity bonds and other mortality-linked securities. *British Actuarial Journal*, 12: 153-197.

Blake, D., Cairns, A., Coughlan, G., Dowd, D., MacMinn, R. (2013) The new life market. *Journal of Risk and Insurance*, 80: 501-558.

Börger, M., Fleischer, D., and Kuksin. N. (2014) Modeling mortality trends under modern solvency regimes. *ASTIN Bulletin*, 44:1-38.

Brouhns, N., Denuit, M., and Vermunt, J.K. (2002) A Poisson log-bilinear regression approach to the construction of projected life tables. *Insurance: Mathematics and Economics*, 31:373-393.

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. (2011b) Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin*, 41: 29-59.

Czado, C., Delwarde, A., and Denuit, M. (2005) Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics*, 36: 260-284.

Duncan, C., Chappell, R., Smith, J., Clark, L., and Ambrose, F. (2002) Rebasing the annual mid-year population estimates for England and Wales. *Population Trends* 109: 9-14. Office for National Statistics.

Girosi, F., and King, G. (2008) *Demographic Forecasting.* Princeton, Princeton University Press.

Hyndman, R.J., and Ullah, M.S. (2007) Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis*, 51: 4942-4956.

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

Li, J.S.-H., Hardy, M.R., and Tan, K.S. (2009) Uncertainty in model forecasting: An extension to the classic Lee-Carter approach. *ASTIN Bulletin,* 39: 137-164.

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(3): 575-594.

ONS (2002a) *Mortality Statistics – General – Review of the Registrar General on deaths in England & Wales, 2000, Series DH1 No.33.* London, HMSO.

ONS (2002b) *Census 2001: First results on population for England and Wales.* London, TSO.

ONS (2002c) Methods used to revise the 1982-2000 annual mid-year population estimates for England and Wales. Office for National Statistics.

ONS (2012a) Explaining the Difference between the 2011 Census Estimates and the Rolled-Forward Population Estimates. Office for National Statistics, 16 July 2012.

ONS (2012b) Population Estimates for England & Wales , Mid-2002 to Mid-2010 Revised (National). Office for National Statistics, 13 December 2012.

ONS (2013) Quality and methodology information. Office for National Statistics, 10 July 2013.

ONS (2014a) Calculating population estimates of the very elderly. Office for National Statistics. Accessed 9/5/2014.

ONS (2014b) Deaths registered in England and Wales (Series DR), 2013. Office for National Statistics. 29 October 2014.

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

Thatcher, R., Kannisto, V., and Andreev, K. (2002) The survivor ratio method for estimating numbers at high ages. *Demographic Research* 6: 1-18.

Vaupel, J.W., Wang, Z., Andreev, K.F., and Yashin, A.I. (1998) *Population Data at a Glance: Shaded Contour Maps of Demographic Surfaces over Age and Time.* Odense, Odense University Press.

Willets, R.C. (2004) The cohort effect: Insights and explanations. *British Actuarial Journal*, 10: 833-877.

Wilmoth, J.R., Andreev, K., Jdanov, D., and Glei, D.A. (2007) Methods protocol for the Human Mortality Database. University of California at Berkeley and Max Planck Institute for Demographic Research. See www.mortality.org. (Accessed 16/3/2014.)

A Modelling Uncertainty in Exposures

We assume that, for each t and x, $E(t,x)/\hat{E}(t,x)$, the ratio of true to published exposures, has a log-normal distribution with mean 1: that is, there is no *a priori* reason to assume that the ONS would deliberately over or under-estimate exposures. The logarithm of the latent death rates, $Y(t,x) = \log m(t,x)$, is assumed to have an underlying smoothness across ages in a given calendar year, t.

A.1 Model for the propagation of errors in exposures

Recall $\epsilon(t,x) = \log E(t,x)$ and $\phi(t,x) = \epsilon(t,x) - \hat{\epsilon}(t,x)$. Define $\phi(t) = (\phi(t,1), \ldots, \phi(t,n_x))'$ to be the vector of errors in year t. $\phi(t)$ is assumed to follow a vector autoregressive process of order 1, $VAR(1)$. Thus, for $t = 2, \ldots, n_y$,

$$\phi(t) = A\phi(t-1) + \sigma_\phi Z_\phi(t)$$

where the $Z_\phi(t)$ are i.i.d. standard multivariate normal random vectors of dimension $n_x \times 1$, $\sigma_\phi > 0$ is a scalar, and the autoregression matrix A is $n_x \times n_x$ with elements $A_{11} = \theta$, $A_{i,i-1} = \theta$ for $i = 2, \ldots, n_x$, and $A_{ij} = 0$ otherwise. We define V_ϕ to be the covariance matrix of $\phi(t)$ given $\phi(t-1)$. It follows that V_ϕ is diagonal with the constant σ_ϕ^2 down the diagonal.

$\phi(1)$ is assumed to have the stationary distribution of $\phi(t)$. Thus, $\phi(1)$ has a multivariate normal distribution with mean vector 0 and covariance matrix $V_{0\phi}$, and is independent of the $Z_\phi(t)$ for $t > 1$. If we let v_{ij} for $i, j = 1, \ldots, n_x$ represent the elements of $V_{0\phi}$, then, for the given structures for A and V_ϕ, we have $v_{ii} = \sigma_\phi^2/(1-\theta^2)$ for all i, $v_{ij} = \sigma_\phi^2/(1-\theta^2)\theta^{2(j-1)}$ for all i and $j > i$, and $v_{ij} = v_{ji}$ for all i and $j < i$.

Samples of the $\phi(t)$ drawn from the MCMC output produce estimates of the $Z(t)$ which are consistent with the i.i.d. standard multivariate normal assumption.

A.2 Model for the underlying log death rates

Individual years are treated in isolation and we assume the following:

$$Y(t,x) \;=\; 3Y(t,x-1) - 3Y(t,x-2) + Y(t,x-3) + \sigma_Y Z_Y(t,x)$$

where the $Z_Y(t,x)$ are i.i.d. standard normal random variables. Thus, for a given t, $Y(t,x)$ is an ARIMA(0,3,0) process, giving locally quadratic predictions plus noise proportional to σ_Y.

The motivation for using an ARIMA model is to assess *in-sample* roughness in the underlying $Y(t,x)$, and it is not intended as a means of modelling $Y(t,x)$ beyond the range of ages in the dataset. The choice of an ARIMA(0,3,0) model provides a convenient way of penalising deviations from a locally quadratic shape. The log-likelihood for year t (see equation 7 in the next section) is equivalent to a roughness penalty of $\sum_{x=4}^{n_x} \left(Y(t,x) - 3Y(t,x-1) + 3Y(t,x-2) - Y(t,x-3)\right)^2 / \sigma_Y^2$, where the inverse of the ARIMA volatility σ_Y^2 takes the role of the smoothing parameter. We choose σ_Y to be small, so that, for each t, the observed process $Y(t,x)$ follows a smooth path that is locally as close as possible to a quadratic curve. The choice of an ARIMA(0,3,0) (locally quadratic smoothness) was preferred to an ARIMA(0,2,0) (locally linear smoothness) because the former resulted in a much greater degree of robustness in the exposures errors relative to the choice of minimum age.

Samples of the $Y(t,x)$ drawn from the MCMC output produce estimates of the $Z_Y(t,x)$ that are consistent with the i.i.d. standard normal assumption.

A.3 Model for deaths

For each t and x, and given $\phi(t,x)$ and $Y(t,x)$, we assume that death counts, $D(t,x)$, are conditionally independent and have a log-normal distribution: that is, $d(t,x) = \log D(t,x)$ has a normal distribution with mean $\mu_D(t,x)$ and variance $S_D(t,x)$. The log-normal is used for computational purposes as an alternative to the usual Poisson distribution for $D(t,x)$. As such we equate the mean and variance of the log-normal to the mean and variance of the matching Poisson: both $\exp[\hat{\epsilon}(t,x) + \phi(t,x) + Y(t,x)]$. Thus, $\mu_D(t,x) = \hat{\epsilon}(t,x)+\phi(t,x)+Y(t,x)-\frac{1}{2}S_D(t,x)$. Additionally, since the mean and variance are equal, we have $\exp[S_D(t,x)] - 1 = 1/\exp[\hat{\epsilon}(t,x) + \phi(t,x) + Y(t,x)]$. A first order Taylor expansion then gives us $S_D(t,x) \approx 1/\exp[\hat{\epsilon}(t,x) + \phi(t,x) + Y(t,x)]$ or $S_D(t,x) \approx 1/E[D(t,x)|\phi(t,x), Y(t,x)]$, an approximation that will be more accurate if the mean, $E[D(t,x)|\phi(t,x), Y(t,x)]$, is large. We can note that, for each (t,x), $\phi(t,x)+Y(t,x)$ is used twice in the mean and variance but, where the number of deaths is relatively large, proximity of $d(t,x)$ to its mean is the main driver of the likelihood function around $d(t,x)$, with the coupled changes in $S_D(t,x)$ having very little impact. As a consequence we choose to fix $S_D(t,x)$ rather than set it at $1/E[D(t,x)|\phi(t,x), Y(t,x)]$, an approach that facilitates the use of the Gibbs sampler outlined later in this appendix. Furthermore, we propose an empirical Bayes approach as an appropriate mechanism for fixing $S_D(t,x)$. Thus, we replace $E[D(t,x)|\phi(t,x), Y(t,x)]$ by the observed deaths, $D(t,x)$: that is, $S_D(t,x) = 1/D(t,x)$.

For larger expected numbers of deaths across all ages, this will provide a good approximation to the true posterior distribution. It is important, nevertheless, to verify that the proposed use of empirical Bayes does not distort or bias the estimation procedure to any significant extent for this particular population size. We experimented, therefore, with three variants. The first two variants used $S_D(t,x) = C/D(t,x)$ where $C = 0.5$ and $C = 2$. The third used $S_D(t,x) = 1/\hat{D}(t,x)$ where $\hat{D}(t,x)$ was equated to the posterior mean of $E(t,x)\exp(Y(t,x))$ using the MCMC output from the original run. In all three cases, the differences between the posterior distributions for the $\phi(t,x)$ were very small. However, we stress again that the approximation would work less well if the numbers of deaths are much smaller.

Apart from the time-series structures specified above, the matrices of latent state variables, Y and ϕ, are assumed to have uninformative, uniform prior distributions on $(-\infty, \infty)^{n_y \times n_x}$. Additionally, we set the process parameters exogenously as $\sigma_Y = 0.01$, $\sigma_V = 0.02$ and $\theta = 0.9$. Now $\sigma_Y = 0.01$ reflects our desire for a high degree of smoothness in $Y(t,x)$, $\sigma_V = 0.02$ reflects the revisions that we have seen in the ONS exposures data, while $\phi = 0.9$ reflects the strong persistence of errors in the exposures data. Estimated exposure errors have been tested for sensitivity to these three parameter settings and the results have been found to be robust.

The log-likelihood for our model is:

$$f(\phi, Y|d, \hat{E}) = f_1(d|Y, \phi, \hat{E}) + f_{2A}(\phi(1)) + \sum_{t=2}^{n_y} f_{2B}(\phi(t)) + f_3(Y) \quad (6)$$

where

$$\begin{aligned} f_1(d|Y, \phi, \hat{E}) &= -\frac{1}{2}\sum_{t=1}^{n_y} (d(t) - Y(t) - \hat{\epsilon}(t) - \phi(t))' S_D(t)^{-1} (d(t) - Y(t) - \hat{\epsilon}(t) - \phi(t)) + \text{const.} \\ f_{2A}(\phi(1)) &= -\frac{1}{2}\phi(1)' V_{0\phi}^{-1} \phi(1) + \text{const.} \\ f_{2B}(\phi(t)) &= -\frac{1}{2} (\phi(t) - A\phi(t-1))' V_\phi^{-1} (\phi(t) - A\phi(t-1)) + \text{const.} \end{aligned}$$

and

$$f_3(Y) = -\frac{1}{2}\sum_{t=1}^{n_y} \frac{1}{\sigma_Y^2} Y(t)' \Delta' \Delta Y(t) + \text{const.} \quad (7)$$

In the above equations,

$$\Delta = \begin{pmatrix} 1 & -3 & 3 & -1 & 0 & \cdots & \cdots \\ 0 & 1 & -3 & 3 & -1 & 0 & \cdots \\ \vdots & 0 & 1 & -3 & 3 & -1 & \cdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & 0 & 1 & -3 & 3 & -1 \end{pmatrix} \quad (\text{an } n_x - 3 \times n_x \text{ matrix});$$

d is the matrix of the logs of the deaths, $D(t, x)$; $S_D(t)$ is a diagonal matrix whose elements are $D(t, x)^{-1}$ for $x = x_1, \ldots, x_{n_x}$; and A, V_ϕ and $V_{0\phi}$ are defined in A.1.

Since Y and ϕ have uniform prior distributions, the log-likelihood above (Equation 6) is equal to the log-posterior density. The posterior distribution for the combination (Y, ϕ) can be seen to be multivariate normal. However, the dimension of the state space $(2n_x n_y)$ is very large, and so identifying the mean and covariance matrix for the posterior is computationally almost impossible. Instead, we adopt a Markov chain Monte Carlo (MCMC) approach using the Gibbs sampler as a proposal distribution. The MCMC algorithm proceeds as follows:

- Let $Y(t)^C$ represent all elements of (Y, ϕ) apart from the vector $Y(t)$, and $\phi(t)^C$ represent all elements of (Y, ϕ), apart from the vector $\phi(t)$.
- For $t = 1, \ldots, n_y$, the conditional posterior for $Y(t)$, given $Y(t)^C$, is multivariate normal, with mean $\mu_Y \equiv \mu(Y(t)^C)$ and covariance matrix $H_Y^{-1} \equiv H(Y(t)^C))^{-1}$ where

$$\begin{aligned} H_Y &= S_D(t)^{-1} + \frac{1}{\sigma_Y^2}\Delta'\Delta \\ \mu_Y &= H_Y^{-1} S_D(t)^{-1} (d(t) - \hat{\epsilon}(t) - \phi(t)) . \end{aligned}$$

- The conditional posterior distribution for $\phi(1)$, given $\phi(1)^C$, is multivariate normal with mean $\mu_\phi(1, \phi(1)^C)$ and covariance matrix $H_\phi(1, \phi(1)^C)^{-1}$, where

$$\begin{aligned} H_\phi &= S_D(1)^{-1} + A' V_\phi^{-1} A + V_{0\phi}^{-1} \\ \mu_\phi &= H_\phi^{-1} \left(S_D(1)^{-1} (d(1) - \hat{\epsilon}(1) - Y(1)) + A' V_\phi^{-1} \phi(2) \right). \end{aligned}$$

- The conditional posterior distribution for $\phi(t)$, given $\phi(t)^C$, for $t = 2, \ldots, n_y - 1$, is multivariate normal with mean $\mu_\phi(t, \phi(t)^C)$ and covariance matrix $H_\phi(t, \phi(t)^C)^{-1}$ where

$$\begin{aligned} H_\phi &= S_D(t)^{-1} + V_\phi^{-1} + A' V_\phi^{-1} A \\ \mu_\phi &= H_\phi^{-1} \left(S_D(t)^{-1} (d(t) - \hat{\epsilon}(t) - Y(t)) + V_\phi^{-1} A \phi(t-1) + A' V_\phi^{-1} \phi(t+1) \right). \end{aligned}$$

- The conditional posterior distribution for $\phi(n_y)$ given $\phi(n_y)^C$ is multivariate normal with mean $\mu_\phi(n_y, \phi(n_y)^C)$ and covariance matrix $H_\phi(n_y, \phi(n_y)^C)^{-1}$ where

$$\begin{aligned} H_\phi &= S_D(n_y)^{-1} + V_\phi^{-1} \\ \mu_\phi &= H_\phi^{-1} \left(S_D(n_y)^{-1} (d(n_y) - \hat{\epsilon}(n_y) - Y(n_y)) + V_\phi^{-1} A \phi(n_y - 1) \right). \end{aligned}$$