A Bayesian approach to modelling and projecting cohort effects

Andrew Hunt*

Pacific Life Re, London†

Corresponding author: andrew.hunt.1@cass.city.ac.uk

David Blake

Pensions Institute, Cass Business School, City University London

June 2018

Abstract

One of the key motivations in the construction of ever more sophisticated mortality models was the realisation of the importance of "cohort effects" in the historical data. However, these are often difficult to estimate robustly, due to the identifiability issues present in age/period/cohort mortality models, and exhibit spurious features for the most recent years of birth, for which we have little data. These can cause problems when we project the model into the future. In this study, we show how to ensure that projected mortality rates from the model are independent of the arbitrary identifiability constraints

*Material in this paper was presented under the title "Projecting mortality: Identifiability with trend changes and cohort effects" at the 17^{th} International Congress on Insurance: Mathematics and Economics in July 2013 in Copenhagen, Denmark. We are grateful to participants at this conference, and to Matthias Börger, Frank van Berkum, Andrew Cairns, Pietro Millossovich and Andrés Villegas for useful discussions regarding this work.

†This study was performed when Dr Hunt was a PhD student at Cass Business School, City University London, and therefore the views expressed within it are held in a personal capacity and do not represent the opinions of Pacific Life Re and should not be read to that effect.

needed to identify the cohort parameters. We then go on to develop a Bayesian approach for projecting the cohort parameters, which allows fully for uncertainty in the recent parameters due to the lack of information for these years of birth, which leads to more reasonable projections of mortality rates in future.
JEL Classification: C11, C15, C32
Keywords: Mortality modelling, age/period/cohort models, identification issues, projection, consistency, cohort parameters

1 Introduction

One of the key motivations in the construction of ever more sophisticated mortality models was the realisation of the importance of "cohort effects" in the historical data, as described in Willets (1999, 2004). These are lifelong mortality effects, which systematically raise or lower the observed mortality rates for individuals born in the same year. Often, these cohort effects can be linked to the specific life histories of the individuals in question and can relate to events such as epidemics, changes in lifestyles such as the rise and fall of smoking rates, increases in obesity or changes in the provision of medical care. While the specific attribution of cohort effects to these events is still controversial in some quarters (for instance, see Murphy (2009, 2010)), there is clear evidence to show that mortality models which include parameters to capture the effect of year of birth give closer fits to the historical data than those lacking such parameters (for instance, see Cairns et al. (2009) and Haberman and Renshaw (2011)).

However, the inclusion of cohort parameters in age/period/cohort mortality models brings with it significant problems. First, the collinearity of the dimensions of age, period and cohort (i.e., the fact that $period = yearofbirth + age$) generates complicated identifiability issues in sophisticated mortality models, which require both additional identifiability constraints in order to fit the model to data and extra care to be taken to ensure that the choice of these constraints does not affect the projection of future mortality rates. These are discussed in a general context in Hunt and Blake (2015b).

Second, we encounter the problem that cohort parameters for the most recent years of birth are estimated on the basis of relatively little data. This means that many of the features we see for these years of birth may be spurious and caused by a combination of the limited information and difficulty in accurately specifying an appropriate age/period structure for the model at younger ages. The second of these factors is resolved by using the "general procedure" for constructing a mortality model, proposed in Hunt and Blake (2014) as a method for selecting the appropriate age/period terms in a mortality model. However, this procedure alone will not deal with the limited information we have regarding the most recent cohorts. The spurious features observed for recent years of birth can lead to unreasonable projections of future mortality rates as these cohort reach older ages, as shown in Cairns et al. (2011). Furthermore, classical approaches for projecting the cohort parameters also understate the uncertainty in these recent cohort parameters, and assume that the cohort parameters estimated on the basis of historical data are known (subject to parameter uncertainty) rather than an initial estimate of an ongoing process.

In this study, we develop a new Bayesian approach for modelling and projecting the cohort parameters from the model constructed in Hunt and Blake (2014).[1] This approach gives projections of mortality rates which make suitable allowance for the uncertainty in the estimated cohort parameters and ensures this uncertainty blends smoothly into our projections of parameters for future years of birth. This approach must be performed in conjunction with a full analysis of the identifiability issues present in the cohort parameters at the estimation stage, which also guarantees that the projections do not depend on the arbitrary identifiability constraints we use when fitting the model. We aim to present a range of techniques for projecting mortality rates in future which are consistent with the features observed in the historical data and which make full allowance for the uncertainty in future projections.

We start in Section 2 by reviewing the model constructed in Hunt and Blake (2014) for men in the UK and, in particular, the features of the cohort parameters from it. In Section 3, we describe the identifiability issues present in the model with respect to the cohort parameters and the impact these have

[1] see **?** for an earlier study examining the Bayesian approach in the same context.

on the time series we use to project future cohort parameters. We need to allow fully for the uncertainty in the fitted cohort parameters and in Section 4, we use Bayesian techniques to combine an assumed dynamic process for generating the cohort parameters with the observations on each cohort to date. Finally, Section 5 concludes.

2 The fitted cohort parameters

We first use the general procedure (GP) to construct a suitable mortality model for data from the Human Mortality Database (2014) for men aged 0 to 100 in the UK over the period 1950 to 2009. The GP constructs a bespoke mortality model in the class of age/period/cohort (APC) models, discussed in Hunt and Blake (2015d), of the form

$$\ln(\mu_{x,t}) = \alpha_x + \sum_{i=1}^{7} f^{(i)}(x;\theta^{(i)})\kappa_t^{(i)} + \gamma_{t-x} \tag{1}$$

where

- age, x, is in the range $[0, 100]$, period, t, is in the range $[1950, 2009]$ and therefore that year of birth, y, is in the range $[1850, 2009]$;
- $\mu_{x,t}$ is the force of mortality at age x and for period t;
- α_x is a static function of age;
- $\kappa_t^{(i)}$ are period functions governing the evolution of mortality with time;
- $f^{(i)}(x;\theta^{(i)})$ are parametric age functions (in the sense of having a specific functional form selected a priori) modulating the impact of the period function dynamics over the age range, potentially with free parameters $\theta^{(i)}$;[2] and
- γ_y is a cohort function describing mortality effects which depend upon a cohort's year of birth and follow that cohort through life as it ages.

[2]For simplicity, the dependence of the age functions on $\theta^{(i)}$ is supressed in the notation used in the remainder of this paper, but not in the model itself.

A summary of the terms in the models and their demographic significance[3] is given in Table 1 of Hunt and Blake (2014).

In this paper, we focus on the cohort parameters fitted by the model, shown in Figure 1. These represent lifelong mortality effects specific to distinct years of birth which we interpret in terms of the life histories of the relevant cohorts in Hunt and Blake (2014). Note that we do not estimate cohort parameters for the first and last ten years of birth in the data, due to the limited number of observations of these cohorts.

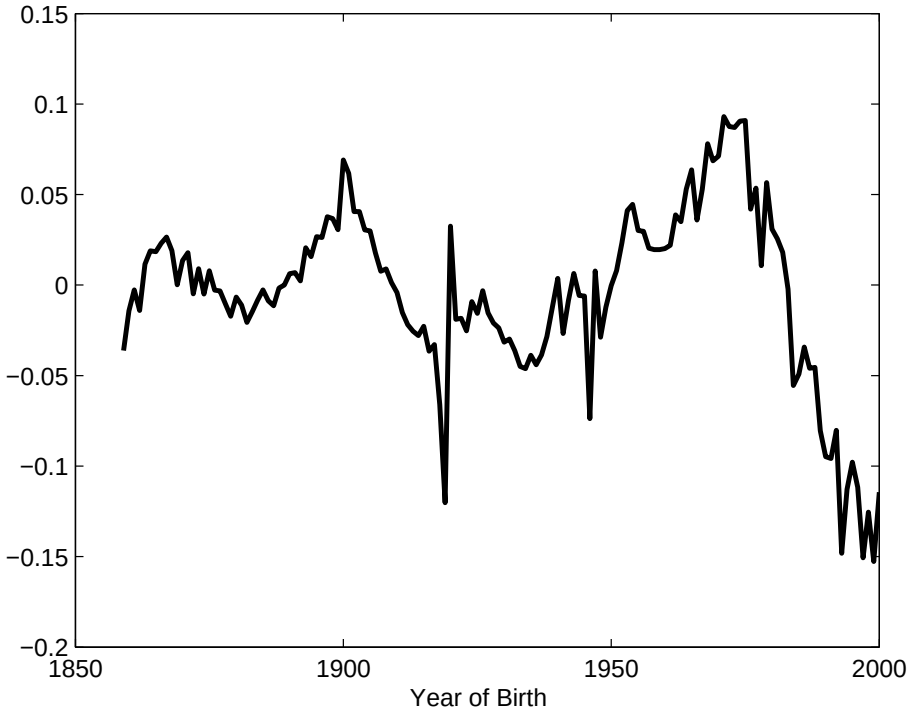

Figure 1: Cohort parameters

Given our desire for the cohort parameters to have demographic signif-

[3]Demographic significance is defined as the interpretation of the components of a mortality model in terms of the underlying biological, medical or socio-economic causes of changes in mortality rates which generate them.

icance, we would like our projections of the cohort parameters to have the following properties:

- The cohort parameters should represent genuine lifelong mortality effects, rather than being mis-classified age/period effects resulting from an incorrect specification of the model. This is an especially large problem for the most recent years of birth, since cohort parameters for these years are only estimated on the basis of data at younger ages, where it is more difficult to properly specify the age/period terms in a model. We deal with this by using the general procedure to sequentially select age/period terms which capture all the significant age/period structure in the data, before adding a set of cohort parameters to the model.

- The cohort parameters should lack trends, i.e., have $\mathbb{E}\gamma_y = 0$ unconditionally for all y for both past and future years of birth. This is consistent with the notion that the cohort effects represent a deviation from the level of mortality for a "typical" cohort. We achieve this through careful choice of our identifiability constraints, as discussed in Section 3.

- The projected cohort parameters should be stationary, in the sense that the variability of the cohort parameters around the central trend should not change with time. We do not believe there is any compelling reason to suppose that the variability in the lifelong mortality factors should be any greater for future cohorts than for those observed to date. This is also consistent with the belief that cohort effects may persist for several years or decades, but should not result in permanent changes in the level of mortality, otherwise they should be re-classified as period effects.

- The projected cohort parameters should be independent of the period effects. For a full discussion of this issue, see Hunt and Blake (2015b). In addition, we believe that cohort effects have very different demographic significance from the period effects and so an assumption of independence is both practical and parsimonious. Given this independence between the cohort and period effects, we can disregard the latter and concentrate only on the former in the rest of the paper, recognising that the purpose of the General Procedure used to derive Equation 1 is to identify all period effects prior to identifying the orthogonal cohort effect.

- The projection method used for the cohort parameters should take account of "unusual" birth cohorts, such as those in 1919/1920 and 1946/1947. Based on the analysis of Richards (2008) and Cairns et al. (2015), we believe that the unusual mortality rates associated with individuals born in these years are not due to genuine cohort effects, but are artefacts of the data. These are caused by the atypical and uneven pattern of births occurring in these years as a result of the demobilisations of soldiers after the First and Second World Wars, respectively, which, in turn, led to a mis-estimation of the size of the exposed population for those years of birth.

There is currently no well-established method for projecting the cohort parameters. A number of techniques are discussed in Cairns et al. (2011). Many of these fit time series from the ARIMA family in order to make projections. The classical approach to projecting the cohort function is to use Box-Jenkins methods to fit a preferred time series process to the historical cohort parameters and then to use this process to project them into the future. This approach generates projected parameters which lack consistency between the past and future. To overcome this, we discuss how consistent projections of the cohort parameters can be obtained using a Bayesian approach which allows adequately for the uncertainty in the parameters in Section 4.

3 Identifiability issues when projecting cohort parameters

Many mortality models are not fully identified. This means that we can find transformations of the parameters in the model which leave the fitted mortality rates unchanged.[4] To uniquely specify the parameters, we impose identifiability constraints. These constraints are arbitrary, in the sense that they do not affect the fit to data, but they do allow us to impose our desired demographic significance on the terms in the model. These issues are discussed in detail in Hunt and Blake (2015a) and Hunt and Blake (2015b).

[4]These are called "invariant transformations" in Hunt and Blake (2015a,b) for this reason.

Using the results of Hunt and Blake (2015b), we observe that the following transformations involving the cohort parameters leave the fitted mortality results unchanged[5]

$$\{\hat{\alpha}_x, \hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\kappa}_t^{(3)}, \hat{\gamma}_y\} = \{\alpha_x - a_0, \kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)}, \gamma_y + a_0\} \tag{2}$$

$$\{\hat{\alpha}_x, \hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\kappa}_t^{(3)}, \hat{\gamma}_y\} = \{\alpha_x + a_1(x - \bar{x}), \kappa_t^{(1)} - a_1(t - \bar{t}), \kappa_t^{(2)}, \kappa_t^{(3)}, \gamma_y + a_1(y - \bar{y})\} \tag{3}$$

$$\{\hat{\alpha}_x, \hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\kappa}_t^{(3)}, \hat{\gamma}_y\} = \{\alpha_x - a_2((x - \bar{x})^2 - \sigma_y + \sigma_t), \kappa^{(1)} - a_2((t - \bar{t})^2 - \sigma_t), \kappa_t^{(2)} + 2a_2(t - \bar{t}), \kappa_t^{(3)}, \gamma_y + a_2((y - \bar{y})^2 - \sigma_y)\} \tag{4}$$

The degrees of freedom represented by the free parameters a_0, a_1 and a_2 in these transformations need to be used to impose three identifiability constraints on the cohort parameters when fitting the model. We choose these to be

$$\sum_y n_y \gamma_y = 0 \tag{5}$$

$$\sum_y n_y \gamma_y (y - \bar{y}) = 0 \tag{6}$$

$$\sum_y n_y \gamma_y ((y - \bar{y})^2 - \sigma_y) = 0 \tag{7}$$

where n_y is the number of observations of each cohort in the data. The justification for these constraints is that they appear to remove polynomial trends up to quadratic order in the cohort parameters at the fitting stage, so that they conform better with the demographic significance described in Hunt and Blake (2015d) and in Section 2, i.e., that the cohort parameters should be centred around zero and not have any long-term trends.

However, it is important to note that the choice of these constraints is arbitrary and it is important that they do not affect our projections of mortality rates. We see that Equation 2 adds a constant to γ_y, Equation 3 adds

[5]Here, X is the number of ages in the data, $\bar{x} = \frac{1}{X}\sum_x x$ and $\sigma_x = \frac{1}{X}\sum_x (x - \bar{x})^2$, and similarly for $\bar{t}$, $\bar{y}$, etc. Also note that, to aid understanding of these complex relationships, Equations 2, 3 and 4 do not incorporate the normalisation factors required on the age functions in order to ensure that $\sum_x |f^{(i)}(x)| = 1 \ \forall i$. These will need to be included before the model is fitted to data.

a term linear in year of birth to γ_y and Equation 4 adds a term quadratic in year of birth to γ_y. These can be combined and written as

$$\hat{\gamma}_y = \gamma_y + a_0 + a_1(y - \bar{y}) + a_2\left((y - \bar{y})^2 - \sigma_y\right) = \gamma_y + AX_y \tag{8}$$

where $X_y = \begin{pmatrix}1, & y, & y^2\end{pmatrix}^\top$. This transformation converts one set of fitted parameters (using one set of identifiability constraints) into an alternative set of parameters (which satisfy a different set of identifiability constraints). These two sets of parameters, γ_y and $\hat{\gamma}_y$, are equivalent: they give the same fitted mortality rates and so there is no statistical reason for preferring one over the other.

As discussed in Hunt and Blake (2015b), identifiability under this transformation means that we need to allow for linear and quadratic trends within the cohort parameters, even if they are not apparent visually. The desire for a stationary distribution around these central, deterministic trends leads us to use an ARMA time series process of the form

$$\Phi(L)(\gamma_y - \beta X_y) = \Psi(L)\epsilon_y \tag{9}$$

where β is a matrix of regression coefficients found from analysing the fitted parameters and L is the lag operator. We can see that this is well-identified by applying the transformation in Equations 8 to Equation 9 to obtain an equivalent set of parameters, which we then substitute into Equation 9 to give

$$\Phi(L)(\hat{\gamma}_y - AX_y - \beta X_y) = \Phi(L)(\hat{\gamma}_y - \hat{\beta} X_y) = \Psi(L)\epsilon_y \tag{10}$$

Doing this has changed the numerical values of the regressors in β, but nothing fundamental about the time series, such as the moving average and autoregressive terms, Φ and Ψ. Hence, if the time series process was appropriate for γ_y, it is also appropriate for $\hat{\gamma}_y$ and, therefore, appropriate for all different sets of identifiability constraints. Hence, this time series model is well-identified.

The specific nature of the time series can be set by choosing the polynomials $\Phi(L)$ and $\Psi(L)$. Classically, these are selected via a modified Box-Jenkins process, which takes care to include the βX_y term. Alternatively, we can work backwards from our desired demographic significance of the cohort

parameters to select $\Phi(L)$ and $\Psi(L)$, whilst also including the βX_y term to ensure that the process is well-identified.

For instance, an AR(1) process, with $\Phi(L) = 1-\rho L$ and $\Psi(L) = 1$, might be felt to be consistent with the desired demographic significance as it is stationary, parsimonious, but still allows for persistent cohort effects. AR(1) processes are often used for the cohort parameters in mortality models, for instance in Cairns et al. (2011). In order to make this well-identified, however, we could choose to project using an AR(1) process around a quadratic trend by including a βX_y term. This is the "AR(1) process around a quadratic drift" process discussed in Hunt and Blake (2015b) for the model of Plat (2009).

When we project using the AR(1) process around a quadratic drift, we obtain $\mathbb{E}\gamma_y = \beta X_y$ unconditionally. Consequently, it might be felt that there is a conflict between the need for the time series process to be well-identified and our desired demographic significance for the cohort parameters, namely that they lack trends. We need to allow for quadratic trends in order to give well-identified projections, but we would like these trends to be zero (i.e., we would like to have $\beta = 0$) based on our (subjective) interpretation of demographic significance. Clearly, the need to have well-identified projections which do not depend upon arbitrary identifiability constraints is more important. However, it is possible to achieve both aims simultaneously.

As shown by Equation 10, the value of β found depends upon the identifiability constraints imposed. In Hunt and Blake (2015b), we argued that the choice of identifiability constraints is arbitrary, and no one set of identifiability constraints is preferable on statistical grounds to any other. We also know that the transformation in Equation 8 allows us to change between different, equivalent sets of parameters (i.e., different arbitrary identifiability constraints) without changing the historical fit to data, whilst using well-identified projection processes for the period and cohort parameters means that the arbitrary choice of identifiability constraints will not affect the projected mortality rates. We therefore propose the following approach.

First, we fit the model as in Section 2, imposing the constraints in Equations 5, 6 and 7. These constraints are convenient when fitting the model as they are simple to apply (by regressing the cohort parameters on the relevant

deterministic trends) and do not depend upon what time series process we subsequently use to project the period and cohort parameters.

Second, we select an appropriate time series process for the cohort parameters, working backwards from our desired demographic significance for the parameters and the need for the process to be well-identified, as discussed in Hunt and Blake (2015b). For illustrative purposes, we select the AR(1) around quadratic drift process discussed above.

Third, we fit an AR(1) around quadratic drift to the fitted cohort parameters. In doing so, we find $\beta = \big(0.74, \ -2.45 \times 10^{-4}, \ 1.16 \times 10^{-5}\big)$. Numerically, these regression coefficients are small, however it is important to note that they are not equal to zero. We observe that there is a constant level for the cohort parameters and, in the long run, the small quadratic trend in the cohort parameters will result in the projected cohort parameters diverging significantly from zero, which conflicts with our desired demographic significance.

One might be tempted to test β for statistical significant and potentially set it to be zero on these grounds. However, the magnitude of β is entirely dependent upon the identifiability constraints used, i.e., even if β is small, we see from Equation 10 that $\hat{\beta} = \beta + A$ can be arbitrarily large depending upon the value of A. Therefore, any decision to ignore β would also be entirely dependent upon the arbitrary identifiability constraints. Thus, we are unable to test β and set it to zero if it proves statistically insignificant, since the results of any statistical tests on them would also depend upon the arbitrary identifiability constraint. Hence, the choice of time series to use for γ_y cannot be motivated by arguments based on statistical significance or goodness of fit, but must be determined by the identifiability issues present in the model, in order to avoid generating poorly-identified projections of mortality rates that depend on the arbitrary constraints imposed when fitting the model.

However, since the value of β depends upon the identifiability constraints, we can impose $\beta = 0$ by choosing a new set of identifiability constraints. To do this, we use the transformations in Equation 8, with $A = -\beta$ found above. This gives an equivalent set of historical parameters, with the original constraints in Equations 5, 6 and 7 over-ridden by the new constraint, $\hat{\beta} = 0$ by construction. Imposing $\beta = 0$ in this fashion does not change our fitted mor-

tality rates (since it merely involves using the invariant transformations), nor does it affect the projected mortality rates, since all the time series processes used for the period and cohort parameters are well-identified. However, it will ensure that our projected cohort parameters have the subjective demographic significance we desire for them from Hunt and Blake (2015d), namely that they lack deterministic trends.

The identifiability constraint $\beta = 0$ could not have been imposed when fitting the model to data, since it depends on knowing which time series process we would use to project the cohort parameters a priori.[6] It therefore makes sense - and is certainly more convenient - to use the original set of identifiability constraints (Equations 5, 6 and 7), to fit the model to data and analyse the fitted cohort parameters. Once we have done this and chosen an appropriate time series process to project the cohort parameters, the fitting constraints can be revisited and we can switch to the more convenient set of identifiability constraints for projecting the model. Because all sets of fitted parameters give the same fitted mortality rates, and because using well-identified projection methods for both the period and cohort parameters means that, when we project any of these sets of parameters, we obtain the same projected mortality rates, we are free to switch between them at any stage of the analysis depending on which set of identifiability constraints is most convenient at the time. This is discussed in depth in Hunt and Blake (2015b).

[6]In principle, if the final time series processes are known in advance or determined by a trial two-step sequential estimation of the model and time series processes, it is possible to fit the model and time series processes to data jointly in a one step process. This can be done either using maximum likelihood techniques, as in Dowd et al. (2011), or Bayesian Markov chain Monte Carlo techniques, as in Pedroza (2006). However, such techniques are complicated to implement and so are not practical when using sophisticated mortality models or if the model is intended to be used for different datasets, where different time series processes might be appropriate.

4 A Bayesian approach for projecting the cohort parameters

We must be careful when allowing for the uncertainty in the cohort parameters, since our estimates to date will be based only on incomplete information. In attempting to allow for this uncertainty, it therefore makes sense to develop a process that is consistent with the nature of our observations of each cohort.

We do this using a Bayesian technique, since Bayesian methods are well suited to the situation where there is inherent uncertainty in parameter estimates based on partial information, but there are prior views regarding the process generating the data. Bayesian methods have been used extensively to fit various mortality models to data, for instance in Pedroza (2006), Cairns et al. (2006), Reichmuth and Sarferaz (2008) and Mavros et al. (2014), often using Markov chain Monte Carlo (MCMC) techniques. However, they have not been used to model the underlying processes generating the cohort parameters. To do this, we construct a Bayesian framework for the cohort parameters from the ground up, starting by specifying the underlying data generating process of each individual cohort parameter and then incorporating a (well-identified) time series process governing the evolution of the cohort parameters across years of birth.

4.1 The data generating process

We start by noting that our dataset gives us a limited number of observations for each cohort, each of which gives us a small amount of information regarding the mortality effects specific to that cohort. We also note that the value of each observation is proportional to the fraction of the cohort which dies at that age, with ages with many deaths providing relatively more insight than ages experiencing few deaths. We formalise this intuition as follows.

Consider a cohort born in year y where a proportion, d_x, of the total cohort dies at age x (assuming ages in the range $[1, X]$ and no other decrements from the population other than death, such as migration). For simplicity, d_x

is assumed to be the same for all cohorts.[7] Therefore, by the time the cohort has reached age x, we have seen a proportion, $D_x = \sum_{\xi=1}^{x} d_\xi$, of the cohort die. Trivially, $D_X = \sum_{\xi=1}^{X} d_\xi = 1$.

We begin by assuming that each observation of cohort y at age x gives us a "packet of information", γ_y^x, relating to the cohort-specific mortality effects. We assume

$$\gamma_y^x | \Gamma_y, \sigma^2 \sim N\left(\Gamma_y, \frac{\sigma^2}{d_x}\right) \tag{11}$$

where Γ_y is the common mean of the information packets for year of birth y. We assume that the information packets are conditionally independent of each other, apart from sharing a common mean. This implies that an observation of a cohort at age 50 only depends upon the observation of the same cohort when it was aged 40 via the mean, Γ_y, and so observations of the γ_y^x can be used to estimate this unknown variable. We will assume a prior distribution for Γ_y based on the time series structure for the cohort parameters considered in Section 4.2.

What we are primarily interested in, however, is the "ultimate" cohort parameter, γ_y. This is the lifelong mortality effect experienced by the cohort, and is constructed from the packets of information observed at each age. Because the ultimate cohort parameter is a lifelong effect, it will only be known fully at the extinction of the cohort (i.e., at time $y + X$), and will be unobservable at any time before this. We assume that the ultimate cohort parameter is given by the weighted sum of the information packets, with the weights given by the schedule of deaths for the cohort, i.e.,

$$\gamma_y = \sum_{x=1}^{X} d_x \gamma_y^x \tag{12}$$

From this, we find the distribution of the ultimate cohort parameter, assuming we have observed no information packets to date (e.g., for cohorts which have yet to be born)

$$\gamma_y | \Gamma_y, \sigma^2 \sim N(\Gamma_y, \sigma^2) \tag{13}$$

[7] In practice, we take d_x to be given by the fitted mortality rates in the final year of the data. However, the results are relatively insensitive to the choice of d_x as long as these reflect a plausible pattern of deaths from a cohort across different ages.

Thus, Γ_y is also the mean of the ultimate cohort parameter, as well as the mean of the information packets. Note that the packets are all a lot more variable than the ultimate cohort parameter, since d_x will tend to be small (since typically less than 5% of people in a cohort die at each age).

As stated previously, before the extinction of the cohort, γ_y is unobservable. However, we will have partial information regarding its value, based on the packets of information observed to date. The challenge, therefore, is to find the distribution of the ultimate cohort parameter given the partial information we have at time t. We will typically assume that t is fixed at the current year of observation (i.e., the last year of the dataset).[8] At this time, we have received the first $t-y$ packets of information, i.e., $\gamma_y^x,\ x \in [1, t-y]$. We, therefore, define the partial sum of the packets, $\underline{\gamma}_y(t) = \sum_{x=1}^{t-y} d_x \gamma_y^x$. The distribution of this partial sum in the absence of any observations of the cohort is given by

$$\underline{\gamma}_y(t) | \Gamma_y, \sigma^2 \sim N\left(D_{t-y}\Gamma_y, D_{t-y}\sigma^2\right) \tag{14}$$

Unlike the individual information packets, γ_y^x, the partial sums, $\underline{\gamma}_y(t)$, are, in principle, observable at time t and could be found from the available data. However, they are not the same as the estimated cohort parameters found when fitting a mortality model to the available data at time t. This is because the expected value of the partial sums depends upon D_{t-y}, i.e., the proportion of the cohort expected to have died to date, and so we observe very small values of $\underline{\gamma}_y(t)$ for cohorts which have just been born, but considerably larger values for older cohorts (for fixed Γ_y). This is inconsistent with the assumption, implicit in the majority of APC mortality models, that the cohort parameters have the same scale.[9]

To deal with this, we define "interim" cohort parameters, $\overline{\gamma}_y(t) = \frac{1}{D_{t-y}}\underline{\gamma}_y(t)$.

[8] In Hunt and Blake (2016), this is relaxed and the year of observation is allowed to change to reflect the impact of new observations on the previously estimated cohort parameters.

[9] This is a consequence of having a simplified age/cohort structure and setting $\beta_x^{(0)} = 1$, discussed in Hunt and Blake (2015d).

From Equation 14, we see that the $\overline{\gamma}_y(t)$ have distribution

$$\overline{\gamma}_y(t)|\Gamma_y, \sigma^2 \sim N\left(\Gamma_y, \frac{1}{D_{t-y}}\sigma^2\right) \tag{15}$$

Not only do the $\overline{\gamma}_y(t)$ have means independent of D_{t-y}, they have variances which are inversely proportional to D_{t-y}, the number of deaths expected from the cohort to date.

Accordingly, we identify the interim cohort parameters, $\overline{\gamma}_y(t)$, with the cohort parameters estimated by the model in Section 2 and shown in Figure 1. Hence, we are able to obtain values of $\overline{\gamma}_y(t)$ by fitting the APC model to data. The interim cohort parameters, $\overline{\gamma}_y(t)$, are therefore assumed to be known at time t, as opposed to having the distribution in Equation 15, and similarly the partial sums, $\underline{\gamma}_y(t)$, are also assumed to be known at time t. It is trivial to move between the fitted $\overline{\gamma}_y(t)$ and the partial sums, $\underline{\gamma}_y(t)$, which are more fundamental in the analysis.

We can use the knowledge of $\overline{\gamma}_y(t)$ (and $\underline{\gamma}_y(t)$) to update the distribution for the ultimate cohort parameter, γ_y, by conditioning on the partial information we have to time t. To do this, we note that, for times in the interval $y \leq t < y + X$

$$\begin{aligned}\gamma_y &= \sum_{x=1}^{t-y} d_x \gamma_y^x + \sum_{x=t-y+1}^{X} d_x \gamma_y^x \\ &= \underline{\gamma}_y(t) + \sum_{x=t-y+1}^{X} d_x \gamma_y^x \end{aligned} \tag{16}$$

Therefore, from Equation 11, we find

$$\gamma_y|\underline{\gamma}_y(t), \Gamma_y, \sigma^2 \sim N(\underline{\gamma}_y(t) + (1 - D_{t-y})\Gamma_y, (1 - D_{t-y})\sigma^2) \tag{17}$$

Thus, we have found the distribution of the ultimate cohort parameters for year of birth, y, conditional on our observations of the cohort to date and its prior expected value. However, we have not made any assumptions regarding the form that this prior expectation should take and, in particular, how this expected value relates to the values for neighbouring cohorts.

4.2 Time series dynamics

The dependence of the ultimate cohort parameters, γ_y, upon the preceding cohorts is given by the time series process driving the dynamics of the cohort parameters. These assumed time series dynamics act as a prior distribution in the Bayesian approach. Working backwards from our desired demographic significance for the cohort parameters, we said, in Section 3, that an AR(1) process around a quadratic drift can provide relatively parsimonious projections in line with our desire for stationary but persistent cohort parameters. Writing the AR(1) process around a quadratic drift in distributional terms gives

$$\gamma_y|\gamma_{y-1},\beta,\rho,\sigma^2 \sim N\left(\beta X_y + \rho(\gamma_{y-1} - \beta X_{y-1}), \sigma^2\right) \tag{18}$$

Comparing this with Equation 13, we see that using the AR(1) process around a quadratic drift is equivalent to setting $\Gamma_y = \beta X_y + \rho(\gamma_{y-1} - \beta X_{y-1})$.[10] This choice for Γ_y also feeds through into the distributions both of the partial sums, $\underline{\gamma}_y(t)$, in Equation 14 to give

$$\underline{\gamma}_y(t)|\gamma_{y-1},\beta,\rho,\sigma^2 \sim N\left(D_{t-y}(\beta X_y + \rho(\gamma_{y-1} - \beta X_{y-1})), D_{t-y}\sigma^2\right) \tag{19}$$

and of the information packets, γ_y^x, in Equation 11 to give[11]

$$\gamma_y^x|\gamma_{y-1},\beta,\rho,\sigma^2 \sim N\left(\beta X_y + \rho(\gamma_{y-1} - \beta X_{y-1}), \frac{\sigma^2}{d_x}\right) \tag{20}$$

To incorporate both sources of information regarding the ultimate cohort parameter, γ_y (i.e., the partial information observed to date for the cohort and that from the cohort parameter for the previous year of birth using the time series structure), we substitute the expression for Γ_y into Equation 17, to obtain

$$\begin{aligned}\gamma_y|\underline{\gamma}_y(t),&\gamma_{y-1},\beta,\rho,\sigma^2 \sim \\ &N\left(\underline{\gamma}_y(t) + (1 - D_{t-y})(\beta X_y + \rho(\gamma_{y-1} - \beta X_{y-1})), (1 - D_{t-y})\sigma^2\right)\end{aligned} \tag{21}$$

[10] The model could, theoretically, be extended to allow for more lags and an AR(p) structure via a different choice for Γ_y.

[11] While the distribution for γ_y^x is not used here, it is necessary when updating the estimates of the cohort parameters for additional data, as done in Hunt and Blake (2016).

This expression gives the distribution of the ultimate cohort parameter for cohort, y, given our observations of the cohort parameter to date and the previous ultimate cohort parameter, γ_{y-1}. It can, therefore, be considered as the posterior distribution in the Bayesian approach, since it takes the prior distribution given by the time series dynamics in Equation 18 and updates it by incorporating the information observable in $\underline{\gamma}_y(t)$. This posterior distribution can be used for simulation purposes, especially when it is rewritten in the form

$$\begin{aligned} \gamma_y &= \underline{\gamma}_y(t) + (1 - D_{t-y})(\beta X_y + \rho(\gamma_{y-1} - \beta X_{y-1})) + \epsilon_y \\ \epsilon_y &\sim N(0, (1 - D_{t-y})\sigma^2) \end{aligned} \tag{22}$$

We refer to this as the "updating equation", which we can use to simulate sample paths for the ultimate cohort parameters, γ_y, over the range $t - X < y < Y$ (where Y is the last cohort in the data for which we have estimated a cohort parameter).

If we were to write Equation 22 using the interim cohort parameters, $\overline{\gamma}_y(t)$, estimated by the model, instead of the partial sums, $\underline{\gamma}_y(t)$, we can see that the expectation of the ultimate cohort parameter is of the form of a weighted sum of the fitted parameter based on observations of the cohort to time t and the expected value from the time series dynamics

$$\mathbb{E}\gamma_y|\underline{\gamma}_y(t), \gamma_{y-1}, \beta, \rho, \sigma^2 = D_{t-1}\overline{\gamma}_y(t) + (1 - D_{t-y})(\beta X_y + \rho(\gamma_{y-1} - \beta X_{y-1}))$$

In this form, the approach can be compared to a "credibility analysis" of the cohort parameters as discussed in Chapter 7 of Kaas et al. (2001), since our estimate of the true parameter is formed as a weighted average of our observed parameter and what would be predicted by the time series. These weights, i.e., the proportion of each cohort expected to have died by the observation date, are shown in Figure 2. We can see that we place a high degree of confidence in our estimates of the cohort parameters before c. 1930 (i.e., individuals currently aged around 80), but this falls rapidly for younger cohorts. For these, the second term in Equation 22 will dominate.

While useful for simulation purposes, Equation 21 is not the end of the story, since it is still conditional on knowing the previous ultimate cohort parameter, γ_{y-1}. However, for the majority of cohort parameters, the previous ultimate cohort parameter will also be unknown at time t. Nevertheless,

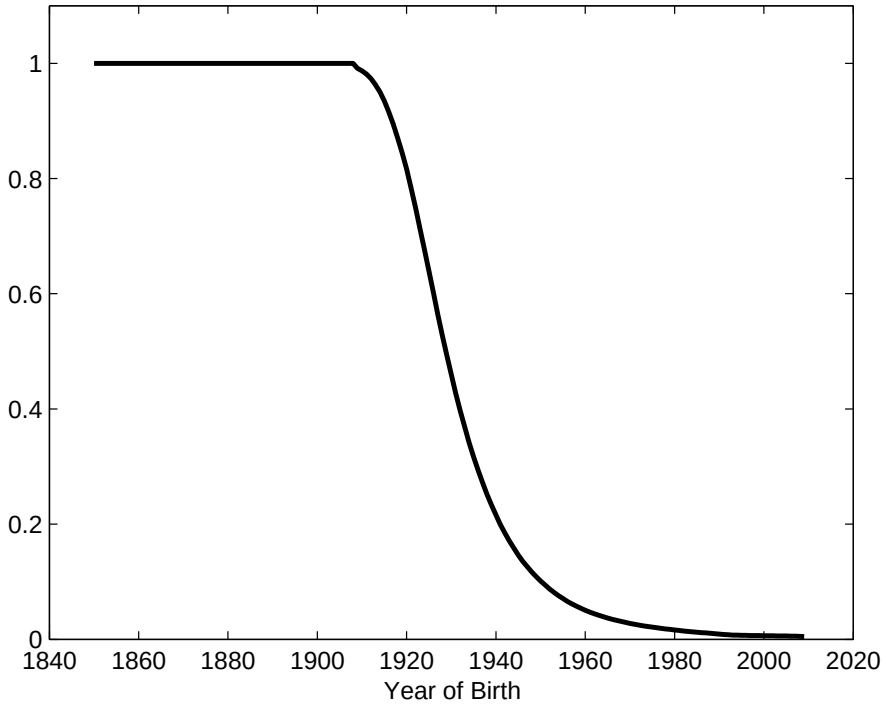

Figure 2: Deceased proportion of cohort, D_y

it is possible to solve Equation 21 iteratively to remove the dependence on γ_{y-1} and obtain the distribution for the cohort parameter γ_y at time t, based solely on the observations made to date. We do this by writing

$$\gamma_y|\mathcal{F}_{t,y}, \beta, \rho, \sigma^2 \sim N(M(y,t), V(y,t)) \tag{23}$$

where $\mathcal{F}_{t,y}$ represents the sum total of information known at time t about cohorts up to and including year of birth y, i.e., $\{\underline{\gamma}_\upsilon(t)\ \upsilon \leq y\}$, and $M(y,t)$ and $V(y,t)$ are the mean and variance functions, respectively. From Equation 21 and Bayes Theorem, we work backwards to give

$$\begin{aligned}\gamma_y|\mathcal{F}_{t,y}, \beta, \rho, \sigma^2 \sim N\Big(&\underline{\gamma}_y(t) + (1-D_{t-y})(\beta X_y + \rho(M(y-1,t) - \beta X_{y-1})),\\ &(1-D_{t-y})\sigma^2 + (1-D_{t-y})^2\rho^2 V(y-1,t)\Big)\\ \Rightarrow M(y,t) = &\underline{\gamma}_y(t) + (1-D_{t-y})(\beta X_y + \rho(M(y-1,t) - \beta X_{y-1}) \qquad (24)\\ V(y,t) = &(1-D_{t-y})\sigma^2 + (1-D_{t-y})^2\rho^2 V(y-1,t) \qquad (25)\end{aligned}$$

This gives us iterative equations for the mean and variances functions, respectively, for the ultimate cohort parameters, based on the information observed to date. This can be solved to give

$$M(y,t) = \sum_{s=0}^{\infty} \left[\prod_{r=0}^{s-1} (1 - D_{t-y+r}) \right] \rho^s \left[\underline{\gamma}_{y-s}(t) + (1 - D_{t-y+s}) \beta (X_{y-s} - \rho X_{y-s-1}) \right] \tag{26}$$

$$V(y,t) = \sum_{s=0}^{\infty} \left[\prod_{r=0}^{s-1} (1 - D_{t-y+r})^2 \right] (1 - D_{t-y+s}) \rho^{2s} \sigma^2 \tag{27}$$

in closed form. We adopt the convention that empty products equal unity (i.e., $\prod_{r=0}^{s-1}(1 - D_{t-y+r}) = 1$ for $s = 0$). It is also important to note that, although these are written as infinite sums, they will in fact terminate, since $D_X = 1$.

So far, this analysis has assumed that we know the parameters of the underlying time series dynamics, i.e., Equation 23 is conditional on knowing the values of β, ρ and σ^2. In practice, these parameters have to be estimated from the fitted cohort parameters, once we find the predictive distribution for $\underline{\gamma}_y(t)|\mathcal{F}_{t,y-1}$, i.e., the observed $\underline{\gamma}_y(t)$, given all previous $\underline{\gamma}_\upsilon(t)$. This can be calculated using Bayes Theorem and Equation 19 to give

$$\begin{aligned} \underline{\gamma}_y(t)|\mathcal{F}_{t,y-1}, \beta, \rho, \sigma^2 \sim N \big(& D_{t-y}(\beta X_y + \rho(M(y-1,t) - \beta X_{y-1})), \\ & D_{t-y}\sigma^2 + \rho^2 D_{t-y}^2 V(y-1,t) \big) \end{aligned} \tag{28}$$

This predictive distribution gives us the distribution of an observable quantity, $\underline{\gamma}_y(t)$, in terms other observable quantities, $\underline{\gamma}_\upsilon(t)$ for $\upsilon < y$ (in $M(y-1,t)$), and the unknown time series parameters. This means that we can use quasi-maximum likelihood methods to estimate β, ρ and σ^2. As discussed in Section 3, in general, we will observe non-zero values for β, which is undesirable given our demographic significance for the cohort parameters. We, therefore, use the invariant transformations in Equations 2, 3 and 4 to set $\beta = 0$, as discussed in Section 3. This also has the benefit of simplifying both the expression for $M(y,t)$ in Equation 26 and the projections of the cohort parameters considerably.

So far, we have only considered the situation where we have two sources of information for each cohort, the observations to date and the time series

structure. In order to project the cohort parameters into the future (i.e., beyond year of birth Y), we do not have any observations to date and therefore we simply use the AR(1) structure to generate projections. To project beyond the last fitted cohort parameter (assumed to be known for the time being), the AR(1) process gives

$$\gamma_{Y+\eta}|\gamma_Y, \rho, \sigma^2 \sim N\left(\rho^\eta \gamma_Y, \frac{1-\rho^{2\eta}}{1-\rho^2}\sigma^2\right)$$

To remove the dependence on γ_Y, which will be unknown in practice, we use Bayes Theorem to obtain

$$\gamma_{Y+\eta}|\mathcal{F}_{t,Y} \sim N\left(\rho^\eta M(Y,t), \frac{1-\rho^{2\eta}}{1-\rho^2}\sigma^2 + \rho^{2\eta}V(Y,t)\right) \tag{29}$$

The variance in Equation 29 contains two parts. First, the variability from projecting the time series, which increases to a constant $\sigma^2(1-\rho^2)^{-1}$ as $\eta \to \infty$ as expected. Second, there is the variability from the fact that our initial value γ_Y is unknown: this source of variability decays exponentially. However, as $V(Y,t) < \sigma^2(1-\rho^2)^{-1}$,[12] this means that our confidence intervals for $\gamma_{Y+\eta}$ increase with time towards a limit.

As with Equation 21, it is helpful to rewrite Equation 29 in the form of an updating equation

$$\begin{aligned}\gamma_{Y+\eta} &= \rho\gamma_{Y+\eta-1} + \varepsilon_y \\ \varepsilon_y &\sim N(0, \sigma^2)\end{aligned}$$

which can be used for generating sample paths. Again, we see that this is simply the time series process for an AR(1) process and is similar to Equation 22, but with $D_{t-y} = 0$ and $\beta = 0$, i.e., we are forecasting cohorts for which there have been no observed deaths to date.

Figure 3 shows a fan chart of the values of the cohort parameters using this method, with the fitted parameters indicated by a dotted line for comparison. We note that the cohort parameters have three regimes:

[12]Mathematically, this is a consequence of $D_{t-Y} > 0$. More intuitively, it can be seen that $\sigma^2(1-\rho^2)^{-1}$ is the variability of a cohort parameter under the prior distribution from the AR(1) time series without any additional information from the data to refine the parameter estimate.

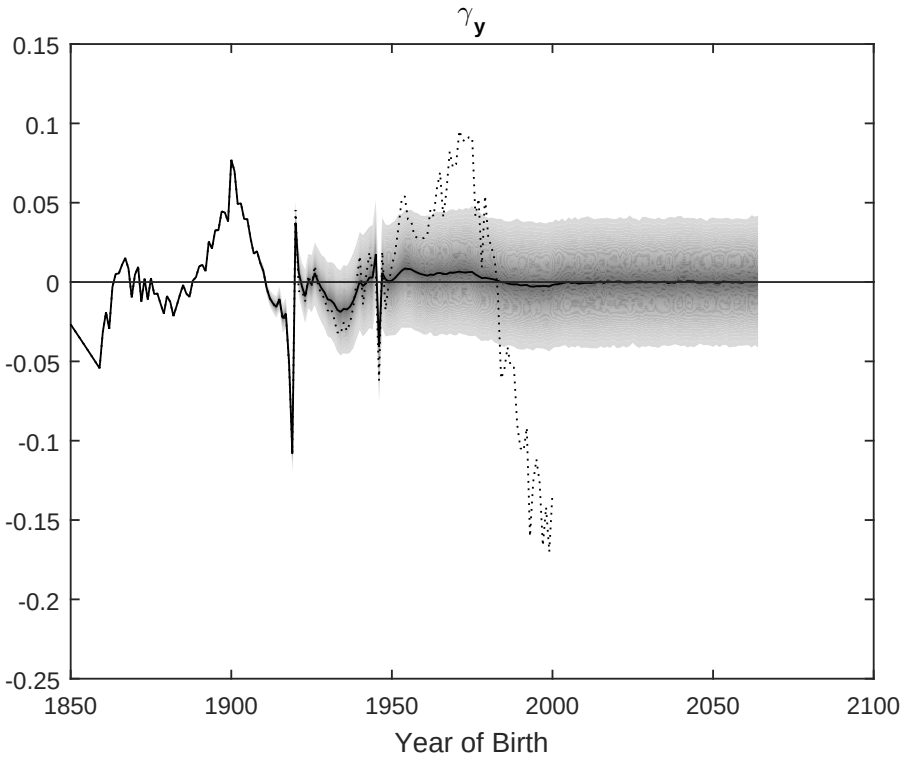

Figure 3: 95% fan chart of the projected cohort parameters using the Bayesian approach

1. $y \leq t - X$ (i.e., $y \leq 1909$): our data has a complete set of observations regarding the cohort and therefore we do not have any uncertainty in the cohort parameters (i.e., $\gamma_y = \underline{\gamma}_y(t) = \overline{\gamma}_y(t)$).

2. $t - X < y \leq Y$ (i.e., $y \in [1910, 1999]$): we have partial observations for each cohort and, therefore, γ_y is not known with certainty but is constructed from the observations to date and the time series dynamics. However, older cohorts are considerably less variable as we have a greater number of observations for these years of birth (and observations including ages where a larger proportion of the cohort is expected to die). In contrast, the uncertainty in the parameter estimates grows rapidly for more recent cohorts.

3. $Y < y$ (i.e., $y \geq 2000$): we have no observations for these years of birth and so the projected cohort parameters are based solely upon the time series dynamics assumed.

It is important to note that, despite the qualitative differences between these three regimes, the confidence interval showing the uncertainty in the parameters blends smoothly between the fitted and the projected parameters, with no sharp discontinuity at the regime boundary. This is in contrast to classical approaches, which have the uncertainty of the cohort parameters increasing sharply at the boundary between estimated cohort parameters, $y \leq Y$ (assumed known) and projected cohort parameters, $y > Y$ (projected using the time series). This is important in many applications, such as projecting annuity values for valuing longevity-linked securities, as discussed in Hunt and Blake (2015c).

We also note from Figure 3 that the expectation of the ultimate cohort parameter, $M(y, t)$ (given by the centre of the confidence interval in Figure 3), can be significantly different from the cohort parameters estimated from data to time t, $\overline{\gamma}_y(t)$. For instance, the interim cohort parameters for years of birth after 1950 often lie outside the 95% prediction interval for the ultimate cohort parameters. This should not concern us unduly, however, since these most recent interim parameters are estimated on the basis of relatively little historical data and so we give them very little weight in our analysis, as shown in Figure 2. Therefore, we are not surprised if the ultimate cohort parameter, revealed once the cohort is fully extinct, is significantly different from this initial estimate. In contrast, classical approaches assume that the unusual

behaviour exhibited by the cohort parameters for the most recent years of birth is genuine. This may give projections of mortality rates which are not biologically reasonable.[13] Indeed, it is a virtue of the Bayesian approach that it can balance the evidence presented by these limited observations of the cohort with the time series process generating the cohort parameters to give ultimate cohort parameters that agree with our demographic significance.

However, since the interim cohort parameters were fitted (along with the other parameters in the model) on the basis of maximising the goodness of fit to data, using the Bayesian approach will have fitted mortality rates that give a worse fit to the historical data. However, the reduction in the goodness of fit is relatively marginal,[14] as the difference between the two is only significant for the most recent cohorts, for which we have relatively little data to fit the model. However, this worsening of the goodness of fit is more than compensated by the more plausible projections and increased allowance for uncertainty in these parameter estimates. In addition, the use of the Bayesian approach for the cohort parameters may appear inconsistent with the use of the other fitted age and period functions in the model. However, these other parameters are estimated over a wide range of years of birth and so are not significantly affected by the changes to the most recent years of birth caused by using the Bayesian approach for the cohort parameters.[15]

Finally, we also see that the pattern of the fitted cohort parameters shown in Figure 1 after 1950 (i.e., a rapid increase and then decrease in cohort mortality relative to the baseline) is smoothed out, since it is not based on sufficient observations to be credible. Therefore, using the Bayesian approach will tend to avoid the issues found in Cairns et al. (2011), where distinctive patterns in the most recent cohort parameters lead to projected mortality rates which are not biologically reasonable.

[13]Introduced in Cairns et al. (2006) and defined as "*a method of reasoning used to establish a causal association (or relationship) between two factors that is consistent with existing medical knowledge*".

[14]We find log-likelihoods of -3.09×10^{-4} using the estimated parameters and -3.25×10^{-4} using the expectation of the ultimate parameters, which is mainly due to worsening the fit to mortality data at age zero. This may indicate that the most recent fitted cohort parameters attempt to overfit data at this unusual age, rather than capturing genuine lifelong mortality effects.

[15]In principle, the other age/period terms in the model could be re-estimated subsequent to finding $M(y,t)$. In practice, however, this was not done in this study.

In summary, we propose a new Bayesian approach for projecting the cohort parameters, which involves updating a prior distribution for them based on assumed time series dynamics with the partial observations we have for each cohort from the available data. This is similar conceptually to a credibility analysis of the form familiar to actuaries. In addition, we have ensured that these projections are well-identified, in the sense that the projected mortality rates do not depend upon any arbitrary set of identifiability constraints imposed. Although this approach is complicated, it yields projections of the cohort parameters which we believe are more plausible and also allow for the uncertainty in the historical cohort parameters as we have only partial data regarding them.

5 Conclusions

Cohort parameters are, increasingly, an important component of mortality models. However, they are often difficult to estimate robustly from the historical data, due to the identifiability issues present in age/period/cohort mortality models and the incorrect specification of the age/period terms in these models at younger ages. Approaches for solving both of these issues are described in Hunt and Blake (2015b) and Hunt and Blake (2014), respectively, which go a long was to dealing with these problems in analysing the past. However, the techniques for projecting cohort parameters into the future often fail to take into account the issues encountered when fitting them to historical data, resulting in the biologically unreasonable projections of mortality rates shown in Cairns et al. (2011).

In this study, we have applied the results of Hunt and Blake (2015b) in the specific context of the model constructed in Hunt and Blake (2014) to ensure that the projections of the cohort parameters from the model do not depend upon the arbitrary identifiability constraints chosen by the user. Furthermore, this freedom means that we can change these constraints at will, choosing a convenient set of constraints when fitting the model to data, but revising this choice subsequently in order to obtain projections of the parameters which accord with our desired demographic significance for the parameters.

We then discussed the need to allow appropriately for the uncertainty that exists in the fitted cohort parameters and the importance of ensuring that there is no discontinuity in the level of this uncertainty between estimated and projected cohort parameters. To do this, we introduce a Bayesian approach for projecting the cohort parameters in Section 4. This uses an assumed time series process to act as a prior assumption for generating the "ultimate" cohort parameters that we would see on exhaustion of the cohorts in question, combined with the "interim" cohort parameters found by fitting the model to historical data. While this approach is introduced in the context of the model constructed in Hunt and Blake (2014), it can be easily applied to any APC mortality model and could be extended to allow for alternative prior assumptions for the time series prior for the ultimate cohort parameters.

A The Bayesian approach for multiple populations

In many circumstances, we are interested in projecting multiple populations simultaneously in a fashion which allows for the dependencies between them, including the cohort parameters. For instance, it is natural to believe that the cohort effects for men and women in the same population should show significant dependence in the historical data and, therefore, be projected in a fashion which allows for this dependence.

The Bayesian approach presented in Section 4 can be extended to allow for multiple populations by making appropriate adjustments to the data generating process and the prior distribution for the ultimate cohort parameters. To do this, we define

$$\boldsymbol{\gamma}_y = \left(\gamma_y^{(1)}, \quad \dots \quad \gamma_y^{(P)}\right)^\top$$

where $\gamma_y^{(p)}$, $p = 1, \dots, P$, are the ultimate cohort parameters for population p, and similarly for the packets of information, partial sums, interim cohort parameters, etc.

Based on this, we generalise Equation 12 for multiple populations

$$\boldsymbol{\gamma}_y = \sum_{x=1}^{X} \boldsymbol{d}_x \boldsymbol{\gamma}_y^x$$

where

$$\boldsymbol{d}_x = \begin{pmatrix} d_x^{(1)} & 0 & \dots \\ 0 & \ddots & \\ \vdots & & d_x^{(P)} \end{pmatrix}$$

is a diagonal matrix and we define $\boldsymbol{D}_x = \sum_{\xi=1}^{x} \boldsymbol{d}_x$ in a similar manner as in Section 4. By assuming

$$\boldsymbol{\gamma}_y^x | \boldsymbol{\Gamma}_y, \Sigma \sim N\left(\boldsymbol{\Gamma}_y, \Sigma \boldsymbol{d}_x^{-1}\right)$$

we obtain multi-population analogues of Equations 13, 14, 15 and 17 which define the multi-population data generating process.

Similar to Section 4.2, we set the prior distribution for the time series dynamics to be a well-identified, multi-variate AR(1) process

$$\boldsymbol{\gamma}_y | \boldsymbol{\gamma}_{y-1}, \beta, \boldsymbol{R}, \Sigma \sim N\left(\beta X_y + \boldsymbol{R}\left(\boldsymbol{\gamma}_{y-1} - \beta X_{y-1}\right), \Sigma\right) \tag{30}$$

as the analogue of Equation 18. Following a similar analysis to that performed in Section 4.2, we obtain the following results

$$\begin{aligned}
\boldsymbol{\gamma}_y | \underline{\boldsymbol{\gamma}}_y(t), \boldsymbol{\gamma}_{y-1}, \beta, \boldsymbol{R}, \Sigma \sim N\Big(&\underline{\boldsymbol{\gamma}}_y(t) + (\mathtt{I} - \boldsymbol{D}_{t-y})\left(\beta X_y + \boldsymbol{R}\left(\boldsymbol{\gamma}_{y-1} - \beta X_{y-1}\right)\right), \\
&(\mathtt{I} - \boldsymbol{D}_{t-y})\Sigma\Big) \qquad (31) \\
\boldsymbol{\gamma}_y | \mathcal{F}_{t,y}, \beta, \boldsymbol{R}, \Sigma \sim N&\left(\boldsymbol{M}(y,t), \boldsymbol{V}(y,t)\right) \\
\boldsymbol{M}(y,t) = \underline{\boldsymbol{\gamma}}_y(t) &+ (\mathtt{I} - \boldsymbol{D}_{t-y})(\beta X_y + \boldsymbol{R}(\boldsymbol{M}(y-1,t) - \beta X_{y-1}) \qquad (32) \\
\boldsymbol{V}(y,t) = (\mathtt{I} - \boldsymbol{D}_{t-y})\Sigma &+ (\mathtt{I} - \boldsymbol{D}_{t-y})\boldsymbol{R}\boldsymbol{V}(y-1,t)\boldsymbol{R}^\top(\mathtt{I} - \boldsymbol{D}_{t-y}) \qquad (33) \\
\underline{\boldsymbol{\gamma}}_y(t) | \mathcal{F}_{t,y-1}, \beta, \boldsymbol{R}, \Sigma \sim N\Big(&\boldsymbol{D}_{t-y}(\beta X_y + \boldsymbol{R}(\boldsymbol{M}(y-1,t) - \beta X_{y-1})), \\
&\boldsymbol{D}_{t-y}\Sigma + \boldsymbol{D}_{t-y}\boldsymbol{R}\boldsymbol{V}(y-1,t)\boldsymbol{R}^\top\boldsymbol{D}_{t-y}\Big) \qquad (34)
\end{aligned}$$

which extend Equations 22, 24, 25 and 28 into the multivariate setting. Using these, we can obtain estimates for the time series parameters β, $\boldsymbol{R}$ and Σ, closed forms for $\boldsymbol{M}(y,t)$ and $\boldsymbol{V}(y,t)$ and make stochastic projections of $\boldsymbol{\gamma}_y$ for multiple populations that are well-identified and allow fully for the uncertainty in the cohort parameters estimated from the historical data.

References

Cairns, A. J., Blake, D., Dowd, K., 2006. Pricing death: Frameworks for the valuation and securitization of mortality risk. ASTIN Bulletin 36 (1), 79–120.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Khalaf-Allah, M., 2011. Mortality density forecasts: An analysis of six stochastic mortality models. Insurance: Mathematics and Economics 48 (3), 355–367.

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

Cairns, A. J., Blake, D., Dowd, K., Kessler, A., 2015. Phantoms never die: Living with unreliable population data. Journal of the Royal Statistical Society: Series A (Statistics in Society), Forthcoming.

Dowd, K., Cairns, A. J., Blake, D., Coughlan, G. D., 2011. A gravity model of mortality rates for two related populations. North American Actuarial Journal 15 (2), 334–356.

Haberman, S., Renshaw, A., 2011. A comparative study of parametric mortality projection models. Insurance: Mathematics and Economics 48 (1), 35–55.

Human Mortality Database, 2014. Human Mortality Database. Tech. rep., University of California, Berkeley and Max Planck Institute for Demographic Research.
URL `www.mortality.org`

Hunt, A., Blake, D., 2014. A general procedure for constructing mortality models. North American Actuarial Journal 18 (1), 116–138.

Hunt, A., Blake, D., 2015a. Identifiability in age/period mortality models. Tech. rep., Pensions Institute PI 15-08, Cass Business School, City University of London.

Hunt, A., Blake, D., 2015b. Identifiability in age/period/cohort mortality models. Tech. rep., Pensions Institute PI 15-09, Cass Business School, City University of London.

Hunt, A., Blake, D., 2015c. Modelling longevity bonds: Analysing the Swiss Re Kortis Bond. Insurance: Mathematics and Economics 63, 12–29.

Hunt, A., Blake, D., 2015d. On the structure and classification of mortality models. Tech. rep., Pensions Institute PI 15-06, Cass Business School, City University of London.

Hunt, A., Blake, D., 2016. Forward mortality rates in discrete time II: Longevity risk and hedging strategies. Tech. rep., Pensions Institute PI 16-02, Cass Business School, City University of London.

Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M. M., 2001. Modern actuarial risk theory. Kluwer Academic Publishers.

Mavros, G., Cairns, A. J., Kleinow, T., Streftaris, G., 2014. A parsimonious approach to stochastic mortality modelling with dependent residuals. Tech. rep., Heriot-Watt University, Edinburgh.

Murphy, M., 2009. The "golden generations" in historical context. British Actuarial Journal 15 (S1), 151–184.

Murphy, M., 2010. Re-examining the dominance of birth cohort effects on mortality. Population and Development Review 36 (2), 365–90.

Pedroza, C., 2006. A Bayesian forecasting model: Predicting US male mortality. Biostatistics 7 (4), 530–550.

Plat, R., 2009. On stochastic mortality modeling. Insurance: Mathematics and Economics 45 (3), 393–404.

Reichmuth, W., Sarferaz, S., 2008. Bayesian demographic modeling and forecasting: An application to U.S. mortality. Tech. rep., Humbolt University, Berlin.

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

Willets, R., 1999. Mortality in the next millennium. Staple Inn Actuarial Society.

Willets, R., 2004. The cohort effect: Insights and explanations. British Actuarial Journal 10 (4), 833–877.