

# BAYESIAN STOCHASTIC MORTALITY MODELLING FOR TWO POPULATIONS

BY

ANDREW J.G. CAIRNS, DAVID BLAKE, KEVIN DOWD,  
GUY D. COUGHLAN, AND MARWA KHALAF-ALLAH\*

## ABSTRACT

This paper introduces a new framework for modelling the joint development over time of mortality rates in a pair of related populations with the primary aim of producing consistent mortality forecasts for the two populations. The primary aim is achieved by combining a number of recent and novel developments in stochastic mortality modelling, but these, additionally, provide us with a number of side benefits and insights for stochastic mortality modelling. By way of example, we propose an Age-Period-Cohort model which incorporates a mean-reverting stochastic spread that allows for different trends in mortality improvement rates in the short-run, but parallel improvements in the long run. Second, we fit the model using a Bayesian framework that allows us to combine estimation of the unobservable state variables and the parameters of the stochastic processes driving them into a single procedure. Key benefits of this include dampening down of the impact of Poisson variation in death counts, full allowance for parameter uncertainty, and the flexibility to deal with missing data. The framework is designed for large populations coupled with a small sub-population and is applied to the England & Wales national and Continuous Mortality Investigation assured lives males populations. We compare and contrast results based on the two-population approach with single-population results.

## KEYWORDS

Small sub-populations, age effect, period effect, cohort effect, Markov chain Monte Carlo, parameter uncertainty, missing data.

---

\* This report has been partially prepared by the Pension Advisory group, and not by any research department, of JPMorgan Chase & Co. and its subsidiaries (“JPMorgan”). Information herein is obtained from sources believed to be reliable but JPMorgan does not warrant its completeness or accuracy. Opinions and estimates constitute JPMorgan’s judgment and are subject to change without notice. Past performance is not indicative of future results. This material is provided for informational purposes only and is not intended as a recommendation or an offer or solicitation for the purchase or sale of any security or financial instrument.

## 1. INTRODUCTION

Recent years have seen considerable developments in the modelling and forecasting of mortality rates. Pioneering work by Lee and Carter (LC, 1992) has been supplemented by a variety of alternatives that might be considered improvements on the single-factor LC model according to a variety of criteria (see, for example, Brouhns et al. 2002; Booth et al. 2002a,b; Currie et al. 2004; Renshaw and Haberman 2003, 2006; Cairns et al. 2006a,b, 2009, 2011a; Hyndman and Ullah 2007; Li et al. 2009).

Most work has focused on stochastic mortality models for single populations, but, for a variety of reasons, however, it is important to be able to model two or more populations simultaneously. First, we might simply want to impose consistency between forecasts for two populations. For example, governments will want to have consistent forecasts of mortality improvements between males and females (see, for example, Carter and Lee, 1992). But, if forecasts for the two populations are made in isolation, then there is the possibility that they cross over or diverge over time in an unreasonable way that only becomes apparent when the two populations are placed side by side. Second, the use of a good stochastic mortality model is important in a number of financial applications (see, for example, Blake et al. 2006; Olivieri and Pitacco 2009; and Pitacco et al. 2009). In a number of cases, the application requires the use of a model for mortality in two (or more) populations, with a critical factor being the need to investigate the degree of correlation in mortality improvements between the two populations over different time horizons (see, for example, Cairns et al. 2006a, 2008; Loeys et al. 2007; Dahl et al. 2008, 2009; Coughlan et al. 2007a,b, 2011; Coughlan 2009; Jarner and Kryger 2011; Li and Hardy 2009; and Plat, 2009).

A good two-population mortality model might be an essential element in making significant financial decisions relating, for example, to longevity risk. A number of authors have considered multi-country comparisons. Oeppen and Vaupel (2002) chart the progress of period life expectancy (PLE) in developed countries over the last 100 or more years. The headline observation is the near linear growth of the maximum PLE over time (maximised over the countries considered). From time to time, one country drops out of the lead position and another comes in when it manages to stumble upon the current optimal combination of lifestyle factors, healthcare and medical advances. Macdonald et al. (1998), Tuljapurkar et al. (2000) and Booth et al. (2006) make some qualitative comparisons of various countries using single-population models, with the latter focusing on the robustness of conclusions based on the LC model applied to different countries. Li et al. (2004) consider countries that have limited data availability and introduce the important idea that parameter estimates might be imported from countries with better quality data.

Full joint modelling of two or more populations has been considered by Li and Lee (2005), Jarner and Kryger (2011) and Biatat and Currie (2010) (Plat 2009 also considers two populations with the second measuring mortality rates by amounts rather than lives). Li and Lee (2005) extend the LC model

by introducing the idea of a global improvement process plus idiosyncratic variations for each country that are mean reverting. In the long run, the global improvement process dominates, resulting in consistent long-term developments in different countries. Järner and Kryger (2011) focus on modelling a small national population's mortality (Denmark) alongside a much larger supranational (Europe-wide) population. The concept of modelling two populations jointly can also be seen in the much earlier paper by Carter and Lee (1992) where bivariate models for the two populations' period effects are briefly touched upon. Biatat and Currie (2010), building on earlier ideas of Currie et al. (2004), introduce the idea of *similarity* between two populations, using P-splines. Populations that are *similar* involve one standard two-dimensional P-splines surface (that is, one that is relatively rich in terms of detail) for death rates underpinning the two (or more) populations, plus a further much more parsimonious P-splines surface describing the relationship between the two populations. Their method has, so far, been applied in situations involving populations of equal status, but it could be easily adapted to situations (as we have in this paper) where we have one large population and a smaller second population with modest amounts of data.

### 1.1. Bayesian framework

A key element of the proposed framework is our single-stage approach to model fitting and process parameter estimation.

In much of the existing stochastic-mortality literature (see, for example, the detailed account in Pitacco et al. 2009), a two-stage approach is taken to model fitting. In the first stage, the underlying state variables are estimated without reference to their assumed dynamic properties. The second stage then fits a time-series model to the stochastic period and cohort effects. These two stages can be combined in either a likelihood-based setting or by adopting the Bayesian paradigm. Both of these single-stage approaches result in improved (i.e., more consistent) estimates of the unobservable (latent) period and cohort effects. The benefits of this improved consistency are greatest for small populations where the single-stage approach dampens the impact of small population noise in the crude mortality data.

Of the two, we choose to adopt the single-stage Bayesian approach for three reasons. First, it helps us to take account of parameter uncertainty in a natural and coherent way. Second, the careful specification of a limited number of prior distributions helps us to avoid unreasonable model parametrisations: an issue that we discuss in more detail later on. Third, the Bayesian setting allows us to deal simply and effectively with small populations, possibly with substantial quantities of missing data. For example, if the larger population 1 has data from 1961 to 2007, while the smaller population 2 only has data from 1991 to 2005, the Bayesian approach allows us to use the full period from 1961 to 2007, treating the years 1961 to 1990 and 2006 to 2007 as missing data for population 2.

We implement the Bayesian approach using Markov chain Monte Carlo (MCMC).

The use of Bayesian methods is not new in this general context. Czado et al. (2005) and Pedroza (2006) provided the first Bayesian analyses using MCMC of the LC model, with further work by Kogure et al. (2009), and Kogure and Kurachi (2010). Prior to this, Bray (2002) used the Age-Period-Cohort (APC) model in a medical statistics context with an ARIMA(0,2,0) model underpinning each of the age, period and cohort effects. More recently, Reichmuth and Sarferaz (2008) have applied MCMC to a version of the Renshaw and Haberman (2006) model, while Girosi and King (2008) have developed models in a Bayesian setting that incorporate covariates, and analysis by cause of death. All of these studies considered the modelling of a single population. So far as we are aware, this paper represents the first attempt to model jointly two populations within a Bayesian setting.

## 1.2. The Age-Period-Cohort model

In this paper we develop, by way of example, a two-population version of the Age-Period-Cohort (APC) model (see, Osmond and Gardner 1982; Osmond 1985; Bray 2002; Jacobsen et al. 2002; Renshaw and Haberman 2006; Cairns et al. 2009). The relative simplicity of this model allows us to focus on the key contribution of this paper, namely combining Bayesian methods, smoothing, and coupling of two populations that are subject to stochastic period and cohort effects. The series of papers by Cairns et al. (2009, 2011a) and Dowd et al. (2010 a, b) found that other models might be preferred, depending on the dataset being considered and the criteria used for model selection. However, the effort in dealing with the additional factors in these models might cause too much distraction from the key contributions mentioned above, so we have chosen to avoid this.

## 1.3. Outline of the paper

The remainder of this paper is as follows. Section 2 outlines the England & Wales and Continuous Mortality Investigation (CMI) assured lives datasets that we will use to investigate two-population mortality modelling. In Section 3, we outline the core hypothesis concerning non-divergence of death rates. In Section 4, we outline the two-population APC model, and this is followed, in Section 5, by a detailed account of the Bayesian estimation approach used to fit this model. In Section 6, we fit the model to the EW and CMI datasets, and discuss aspects of the initial MCMC output before going on to analyse forecasting results incorporating full parameter uncertainty. Section 7 concludes.

## 2. DATA

In general terms, our datasets will have a three-dimensional structure: two populations,  $n_y$  calendar years of observation, and  $n_a$  ages. The age range will

be  $x_0, \dots, x_1 = x_0 + n_a - 1$ , and the range of years covered will be  $y_0, \dots, y_1 = y_0 + n_y - 1$ . The corresponding cohort years of birth are  $c_0, \dots, c_1$ , where  $c_0 = y_0 - x_1$  and  $c_1 = y_1 - x_0$ .

Our data will consist of deaths,  $D_i(t, x)$ , and (central) exposures,  $E_i(t, x)$ , for each population,  $i$ , calendar year,  $t$ , and age,  $x$  last birthday. From this, we derive the crude (central) death rates,  $\hat{m}_i(t, x) = D_i(t, x)/E_i(t, x)$ .

## 2.1. Missing data

For the MCMC approach described in this paper, we have the option of allowing for missing data in certain circumstances. Specifically, we allow for partially or completely missing calendar years or cohorts in either of the populations' data. Individual cells might be deemed missing if data are considered to be unreliable (as is the case for the EW 1886 cohort, see Cairns et al., 2009) or unrecorded (for example, data for one population might be available for fewer years than the other population).

The possibility to record some cells as missing data allows us to make greater use of other data either in the same population or in the second population: data that otherwise might have to be excluded entirely. In some cases, being able to use this additional data allows us to improve, refine or make more accurate forecasts of mortality.

## 2.2. Specific datasets

We focus on this paper on data from 1961 to 2005 for ages 60 to 89 for England and Wales (EW) males (186 million life-years) and the UK Continuous Mortality Investigation's (CMI) assured lives, males (21 million). The second population is (mostly) a sub-population of the larger and about 10% in size. The CMI assured lives datasets are made up of people who are willing and able to buy life assurance. These will generally be wealthier and healthier than the typical EW male or female. The methodology has also been applied successfully to EW and CMI females, to USA and California males, and to EW and Scotland males.

## 2.3. Age definitions

Most of the national datasets with which we might work, report deaths during a calendar year grouped by *age last birthday*. Similarly, exposures are also recorded by age last birthday. In contrast, the CMI dataset records deaths and exposures according to *age nearest birthday*. Strictly, therefore, CMI death rates for age  $x$  should be compared with, for example, the average of the EW death rates at ages  $x$  and  $x + 1$ . In acknowledging this difference, we note that it does not cause us any problems, because (a) of the non-parametric nature of the model, and (b) we treat the forecast and historical death rates in a consistent way within each population and between populations.

### 3. TWO-POPULATION MODELLING: CORE HYPOTHESIS

Before we focus on a two-population analysis of a specific dataset, we will introduce the key idea that underpins two-population modelling. We have two populations  $i = 1, 2$ . Let  $m_i(t, x)$  be the underlying death rate at age  $x$  in calendar year  $t$  for population  $i$ .

We know from numerous papers and analyses that genuine differences exist between populations. Often, one population has significantly lower mortality than another. For example, the CMI assured lives dataset consists of a subset of the UK population that, as previously mentioned, is, on average, wealthier and healthier than the UK average. We expect that this sub-population to remain wealthier and healthier in the future. It seems reasonable, therefore, to assume that the CMI mortality (and, also, central forecasts) will remain correspondingly lower in the future than that of the national population. Developing this idea to be applied in a wider context, we would expect the death rates in two related populations not to diverge over time (see, for example, Li and Lee (2005), and Jarner and Kryger (2011)). We translate this qualitative expectation into the following mathematical hypothesis: *for each age  $x$ , the ratio  $m_1(t, x)/m_2(t, x)$  will not diverge as  $t \rightarrow \infty$ .*

The above hypothesis allows the CMI death rates to remain at a steady level below the EW rates in the long term, while at the same time allowing for random fluctuations. However, whatever the model for random fluctuations is, it needs to involve some form of mean reversion.

### 4. THE TWO-POPULATION AGE-PERIOD-COHORT MODEL

In this paper we use a two-population version of the APC model:

$$\begin{aligned}\log m_1(t, x) &= \beta_x^{(11)} + n_a^{-1} \kappa_t^{(21)} + n_a^{-1} \gamma_{t-x}^{(31)} \\ \log m_2(t, x) &= \beta_x^{(12)} + n_a^{-1} \kappa_t^{(22)} + n_a^{-1} \gamma_{t-x}^{(32)}\end{aligned}$$

for given age, period and cohort effects,  $\beta_x^{(1i)}$ ,  $\kappa_t^{(2i)}$  and  $\gamma_{t-x}^{(3i)}$ .  $x$  is the age,  $t$  the calendar year of observation,  $t-x$  is the cohort year of birth and  $n_a$  the number of ages. Although cohort effects do not appear to be significant in some countries, they are a well established feature in some populations such as England & Wales. So the inclusion of a cohort effect is necessary in some cases and, in the present context, enriches the two-population modelling problem.

We do not claim that this model is necessarily the best model for the datasets to be considered in this paper based on a particular model selection criterion: other models such as “M7” in Cairns et al. (2009) or the Renshaw and Haberman (2006) (“M2”) model might fit better. Instead, the objective here is to illustrate the process of developing and fitting a two-population model,

and it was felt that the greater complexity of either the Cairns et al. (2009) or Renshaw and Haberman (2006) models would significantly complicate the main message that is trying to be conveyed in this paper. We choose to use the APC model because of its relative simplicity on the one hand, while on the other it is known to compete reasonably well alongside other models (Cairns et al., 2009) as well as incorporate a cohort effect. In a similar vein, the detailed elements of the stochastic model laid out in equations (1) to (4) could be generalised, but again that would result in a loss of clarity in the discussion of the key messages.

To satisfy the core hypothesis, it is sufficient to assume that both  $\kappa_t^{(21)} - \kappa_t^{(22)}$  and  $\gamma_{t-x}^{(31)} - \gamma_{t-x}^{(32)}$  are mean reverting.

### 4.1. Empirical findings

Our experiences in modelling mortality point to the following empirical findings where we have two closely linked populations (see, for example, Figure 1):

- Period effects contain significant year-on-year randomness reflecting current environmental fluctuations. The random effects in linked populations have significant positive correlation.
- Cohort effects are relatively smooth processes (compared to period effects) around a random trend. This smoothness reflects a gradual build up of lifestyle, medical and environmental factors over time resulting in high correlation between adjacent cohorts. Over longer time horizons we see strong correlation between two populations' cohort effects (Figure 1, right).
- Where we use a two-stage approach to model fitting with single populations, estimated cohort effects for small populations contain significant noise resulting from Poisson randomness in death counts (Figure 1, right).
- Correlation between mortality improvements in the two populations rises with the time horizon.

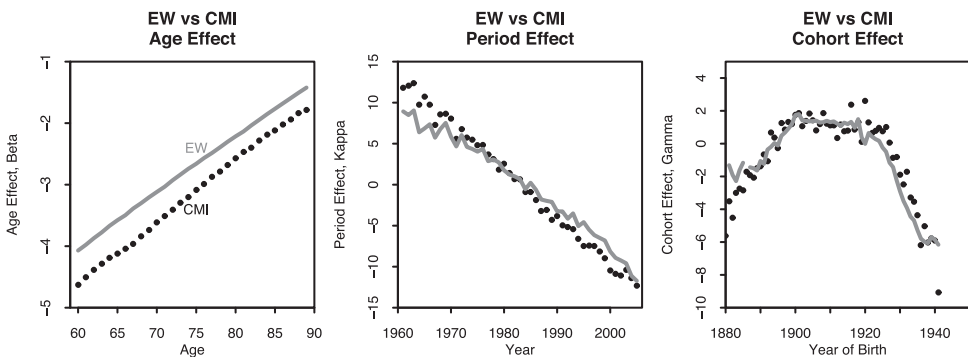


FIGURE 1: Single population estimates for age (left), period (centre) and cohort (right) effects for England & Wales males (lines) and CMI males (dots) using the single-population APC model discussed by Cairns et al. (2009). Data from 1961 to 2005 and ages 60 to 89.

## 4.2. Desirable criteria

In this section, we list potential criteria to add to our core hypothesis in Section 3 which seem reasonable from a biological/environmental point of view. These criteria are subjective in nature and are only partly supported by our empirical findings:

- C1:  $\kappa_t^{(21)}$  and  $\kappa_t^{(22)}$  should have similar conditional 1-year-ahead variances.
- C2:  $\gamma_c^{(31)}$  and  $\gamma_c^{(32)}$  should have similar conditional 1-year-ahead variances.
- C3:  $\gamma_c^{(31)}$  and  $\gamma_c^{(32)}$  (both covariance-stationary processes) should have similar unconditional variances.
- C4: In the long-run,  $\gamma_c^{(31)}$  and  $\gamma_c^{(32)}$  should be positively correlated.

Each of these needs some justification. C1-C3 are included on the basis of subjective judgement. From a biological and environmental perspective it is difficult, under normal circumstances, to envisage how the level of variability in the *underlying* death rates and period effects could be substantially different in two populations.

Substantially different levels of short-term variability in two populations would mean that one population was somehow much more vulnerable to short term shocks. The degree to which this is possible depends on the detailed characteristics of the two populations, but where the characteristics are broadly comparable, especially in terms of geography and access to medical facilities, we would not expect to see substantial differences in variability in the period effect (C1). A similar argument applies to the cohort effects (C2), particularly where the two populations have some important shared characteristics.

Looking, now, to the longer term: period effects will be underpinned by a random-walk model so it is difficult to specify any particular long-term relationship between the variances of  $\kappa_t^{(21)}$  and  $\kappa_t^{(22)}$ . However, we can remark that for the datasets considered the absence of a long-term criterion was not found to cause any difficulty. In contrast, the cohort effects (might) involve some element of mean reversion, so long term variance is a meaningful quantity to consider (C3). Again, we might consider under what circumstances the long-term variance of two population's cohort effects might differ significantly. And again we might take the view that significant differences would be difficult to justify where the two populations share key characteristics. If the two populations differ in terms of their socio-economic background then we might see some differences: differing exposures over individual lifetimes to various medical, environmental and lifestyle factors might result in different accumulations of the benefits or adverse effects of these factors. However, again we would not expect these differences to be very large. Criterion C4 also concerns the long term. In part, this is supported by empirical evidence (Figure 1, right). However, it is backed up by similar subjective arguments to those concerning long term variability. Typically, equivalent cohorts in the two populations will be exposed to a similar range of "random" medical, environmental and lifestyle factors over their lifetimes and so we would expect the combination of



these factors to result in positive correlation between the two populations' cohort effects.

These criteria are incorporated into our model fitting process through the use of enhanced priors, as discussed in Section 5.4.

**4.3. Alternative cohort effect hypothesis**

Instead of having two distinct, but highly-correlated cohort effects the right-hand plot in Figure 1 suggests that it might be possible to have just a single common cohort effect. For these particular datasets, imposing the EW cohort effect on the CMI population was considered, and we found that there was only a small deterioration in the quality of fit, leading to a conclusion that the cohort effects were not significantly different. However, there is no *a priori* reason why the two populations' cohort effects should be identical and, therefore, we chose, in the present work, to keep the two cohort effects as distinct processes.

**4.4. Population 1 dominant**

The approach adopted in this paper is influenced by a typical scenario where we wish to model, say, a pension fund's mortality alongside the national population. We focus, therefore, on situations where one population (say population 1) is much larger than the other (population 2). Hence, we choose to model population 1 using a standard one-population model, and then tackle the second population by modelling the spreads between it and population 1. We, therefore, define

$$R_2(t) = \kappa_t^{(21)}, S_2(t) = \kappa_t^{(21)} - \kappa_t^{(22)}, R_3(c) = \gamma_c^{(31)}, \text{ and } S_3(c) = \gamma_c^{(31)} - \gamma_c^{(32)}.$$

Our core hypothesis in Section 3 indicates that we require the spreads,  $S_2(t)$  and  $S_3(c)$ , to be mean reverting.

The models used will be as follows:

- $R_2(t)$  is modelled as a random walk.  $S_2(t)$  is modelled as an AR(1) time series. Innovations for  $R_2(t)$  and  $S_2(t)$  are modelled as i.i.d. bivariate normal from one year to the next, allowing for non-zero correlation.
- $R_3(c)$  is modelled as an AR(2) process around a deterministic linear trend.  $S_3(c)$  is modelled as a mean-reverting AR(2) time series. Innovations for  $(R_3(c), S_3(c))$  are modelled as i.i.d. bivariate normal from one year to the next, allowing for non-zero correlation.

The random walk for the central period effect,  $R_2(t)$ , mimics what has been done elsewhere (for example, Lee and Carter, 1992, and Cairns et al., 2011a). For the central cohort effect,  $R_3(c)$ , Renshaw and Haberman (2006) and Cairns et al. (2011a) previously used an ARIMA(1,1,0) model. However, the AR(2)

model around a linear trend has been found to work just as well, and, indeed, incorporates the ARIMA(1,1,0) model as a limiting case. From a qualitative point of view, also, an AR(2) model with autoregressive coefficients that are relatively large in magnitude can produce results that mimic the large-scale patterns that we see in the data (e.g. Figure 1, right), with relative smoothness in the short term and occasional shifts in the trend. This smoothness was considered a desirable property in our section on empirical findings. For the spread between the period effects,  $S_2(t)$ , the AR(1) model is a pragmatic choice that works well when applied to the single-population period effects using the two-stage approach. The AR(1) model, of course, also incorporates mean reversion. The AR(2) model for the spread between the cohort effects,  $S_3(c)$ , is again a choice that works well when applied to the single-population period effects using the two-stage approach. However, choosing AR(2) rather than AR(1) (which in any event is a special case of AR(2)) allows us to model the central cohort effect,  $R_3(c)$ , and the spread,  $S_3(c)$ , in a more consistent way.

Previous papers dealing with the APC model (Renshaw and Haberman 2006; Cairns et al. 2009, 2011a) have discussed the need to incorporate identifiability constraints. Here, we use constraints that are equivalent, but nevertheless different in concept and that have been developed to facilitate convergence of the MCMC algorithm: namely, that

- $R_2(1) = 0$ ,
- $S_2(t)$  is mean reverting to zero
- $R_3(c)$  is AR(2) around zero, and
- $S_3(c)$  is AR(2) around zero.

All of these constraints can be achieved by shifting and tilting  $R_2(t)$ ,  $\beta_x^{(11)}$  and  $\beta_x^{(12)}$ , without having an impact on the Poisson log-likelihood function for deaths (Section 5.2). For example, we indicated that  $R_3(c)$  should be modelled as an AR(2) model around a deterministic linear trend. However, the linear trend can be subtracted from  $R_3(c)$ , with compensating adjustments to  $R_2(t)$  and  $\beta_x^{(11)}$ . It is important also to remark that the particular choice of constraints does not impact in any way on the forecast dynamics of future death rates.

All of the above equates to the following mathematical statement of the model:

$$R_2(t + 1) = R_2(t) + \mu_{R2} + C_{211}Z_{21}(t + 1) \tag{1}$$

$$S_2(t + 1) = \mu_{S2} + \psi_{S2}(S_2(t) - \mu_{S2}) + C_{221}Z_{21}(t + 1) + C_{222}Z_{22}(t + 1) \tag{2}$$

$$\tilde{R}_3(c) = R_3(c) - \mu_{R3} - \delta_{R3}(c - \bar{c})$$

$$\tilde{S}_3(c) = S_3(c) - \mu_{S3}$$

$$\tilde{R}_3(c + 1) = (\phi_{R31} + \phi_{R32})\tilde{R}_3(c) - \phi_{R31}\phi_{R32}\tilde{R}_3(c - 1) + C_{311}Z_{31}(c + 1) \tag{3}$$

$$\begin{aligned} \tilde{S}_3(c+1) &= (\phi_{S31} + \phi_{S32})\tilde{S}_3(c) - \phi_{S31}\phi_{S32}\tilde{S}_3(c-1) + C_{321}Z_{31}(c+1) \\ &\quad + C_{322}Z_{32}(c+1). \end{aligned} \tag{4}$$

The details of these equations are as follows:

- $\mu_{R2}$  is the drift in the random walk  $R_2(t)$ .
- $\mu_{S2}$  and  $\psi_{S2}$  are the mean-reversion level and the AR(1) parameter respectively of the period-effect spread,  $S_2(t)$ . For the process to be stationary (mean reverting), we require  $-1 < \psi_{S2} < 1$ .
- Define  $C^{(2)} = \begin{pmatrix} C_{211} & 0 \\ C_{221} & C_{222} \end{pmatrix}$ , and  $V^{(2)} = C^{(2)}C^{(2)'}$ .  $V^{(2)}$  is the 1-year-ahead conditional covariance matrix of  $(R_2(t), S_2(t))'$ .
- $\bar{c}$  is defined as  $(c_0 + c_1 + 2)/2$ , where  $(c_0, c_1)$  is the complete range of years of birth covered in the dataset.
- $\mu_{R3} + \delta_{R3}(c - \bar{c})$  is the linear trend, which  $R_3(c)$  is reverting to.
- $\mu_{S3}$  is the mean-reversion level of the cohort-effect spread,  $S_3(c)$ .
- $\tilde{R}_3(c)$  and  $\tilde{S}_3(c)$  are AR(2) processes that are mean reverting to 0.
- Define  $C^{(3)} = \begin{pmatrix} C_{311} & 0 \\ C_{321} & C_{322} \end{pmatrix}$ , and  $V^{(3)} = C^{(3)}C^{(3)'}$ .  $V^{(3)}$  is the 1-year-ahead conditional covariance matrix of  $(R_3(t), S_3(t))'$  (and of  $(\tilde{R}_3(t), \tilde{S}_3(t))'$ ).
- $\phi_{R31}, \phi_{R32}, \phi_{S31}$  and  $\phi_{S32}$  are the AR(2) parameters for the processes  $\tilde{R}_3(c)$  and  $\tilde{S}_3(c)$ . For the processes to be stationary, we require each of  $\phi_{R31}, \phi_{R32}, \phi_{S31}$  and  $\phi_{S32}$  to lie between  $-1$  and  $+1$ .
- The identifiability constraints used in our specific MCMC algorithm require that  $\mu_{S2} = \mu_{R3} = \mu_{S3} = \delta_{R3} = 0$ .

Death rates and death counts are then modelled as follows. First, reconstruct the period and cohort effects:

$$\kappa_t^{(21)} = R_2(t), \kappa_t^{(22)} = R_2(t) - S_2(t), \gamma_t^{(31)} = R_3(t), \gamma_t^{(32)} = R_3(t) - S_3(t).$$

Second, calculate the underlying death rates:

$$\begin{aligned} m_1(t, x) &= \exp\left[\beta_x^{(11)} + n_a^{-1}\kappa_t^{(21)} + n_a^{-1}\gamma_{t-x}^{(31)}\right] \\ m_2(t, x) &= \exp\left[\beta_x^{(12)} + n_a^{-1}\kappa_t^{(22)} + n_a^{-1}\gamma_{t-x}^{(32)}\right], \end{aligned}$$

where  $\beta_x^{(11)}$  and  $\beta_x^{(12)}$  are the populations 1 and 2 age effects. Third, given the matrices of exposures,  $E_k(t, x)$ , actual numbers of deaths,  $D_k(t, x)$ , at age  $x$  last birthday during year  $t$  are assumed to be independent Poisson random variables

with mean  $m_k(t, x)E_k(t, x)$  (Brouhns et al. 2002). (For alternatives to the Poisson assumption, see, for example, Lee and Carter 1992; Li et al. 2009; and Pitacco et al. 2009).

#### 4.5. Equal populations

This paper focuses on the case described above where population 2 can be considered to be a small sub-population of or subsidiary to population 1. In other applications, we might have two populations which carry equal status, such as males and females, or two different national populations. In this case, we suggest adapting the model above as follows. In contrast to the previous section,  $R_2(t)$  is redefined as the mean of  $\kappa_t^{(21)}$  and  $\kappa_t^{(22)}$ . Similarly,  $R_3(t)$  is redefined as the mean of  $\gamma_c^{(31)}$  and  $\gamma_c^{(32)}$ .  $S_2(t)$  and  $S_3(c)$  retain the same definitions as being the difference between the period and cohort effects. We then have

$$\begin{aligned}\kappa_t^{(21)} &= R_2(t) + \frac{1}{2}S_2(t), & \kappa_t^{(22)} &= R_2(t) - \frac{1}{2}S_2(t), \\ \gamma_t^{(31)} &= R_3(t) + \frac{1}{2}S_3(c), & \gamma_t^{(32)} &= R_3(t) - \frac{1}{2}S_3(c).\end{aligned}$$

This redefinition acknowledges the equal status of the two populations by imposing symmetry in the relationship between, for example,  $R_2(t)$  and  $S_2(t)$  on the one hand and  $\kappa_t^{(21)}$  and  $\kappa_t^{(22)}$  on the other. This is just one of a number of variants that could be considered, but it is one that would require only modest changes to the estimation method (and programs) that is discussed in the next section. A different variant could model: (a) the vector  $(\kappa_t^{(21)}, \kappa_t^{(22)})$  as a vector autoregressive (VAR) process, integrated of order 1, with the additional constraint that the spread is autoregressive of order 0; and (b)  $(\gamma_c^{(31)}, \gamma_c^{(32)})$  is VAR of order 0.

### 5. ESTIMATION METHOD

In previous work (see, for example, Lee-Carter 1992; Brouhns et al. 2002; Booth et al. 2005; Cairns et al. 2009, 2011a), most researchers have employed a two-stage, non-Bayesian approach to modelling: stage 1 estimating age, period and cohort effects without reference to their underlying dynamics; and stage 2 fitting a suitable stochastic process to the period and cohort effects. A good general account of this including iterative schemes for parameter estimation can be found in Pitacco et al. (2009). This approach only works well if the population size is large. More recently, some authors (e.g., Bray 2002; Czado et al. 2005; Reichmuth and Sarferaz 2009; Kogure and Kurachi 2010), when considering single populations, have sought to combine these two stages in a Bayesian setting.

For smaller populations, sample variation affects death counts, which, in turn, can have a non-negligible impact on estimates of age, period and cohort effects, with significant noise obscuring the true signal (as, for example, in Figure 1, right).

This provides one motivation for combining stages 1 and 2. A likelihood-based approach, therefore, would combine the Poisson likelihood for the death counts with the ARIMA likelihood functions for the latent random period and cohort effects. With a large population, the Poisson component will dominate, so that the impact of combining stages 1 and 2 will have little impact. For a smaller population, the ARIMA likelihood functions will compete with the Poisson likelihood to produce estimates for the latent period and cohort effects that look more like the proposed ARIMA ( $p, d, q$ ) models.

The combined fitting procedure allows us to include cohorts with only one observation (a problem with the two-stage approach: see Cairns et al. 2009), since the low level of information provided by one data point will be balanced by the ARIMA likelihood for that observation. In the Bayesian setting, estimates for cohorts with fewer observations will have wider posterior distributions. For the youngest cohorts, use of the limited amount of data gives us some precious information about the most recent values for the two populations' cohort effects that would otherwise be unavailable to us. This in turn helps us to make improved forecasts of what will happen to the cohort effects in the future.

**5.1. Markov chain Monte Carlo (MCMC)**

Bayesian statistics and MCMC methods (specifically the Metropolis-Hastings, MH, algorithm) provide a framework within which we can tackle the estimation problem in a single stage (see, for example, Gilks et al. 1996).

- MCMC will produce a Bayesian posterior distribution for the forecasting model parameters and for the latent age, period and cohort effects.
- The method can deal effectively with missing data in our mortality dataset, or, for example, the removal of data points that are considered to be unreliable or out of line for some unknown reason. The MCMC output will allow us to derive a posterior distribution for the relevant parameters and for the underlying death rate for the missing cells.

**5.2. Likelihood, prior and posterior**

The complete parameter vector is denoted by  $\theta$ , and consists of subvectors for the period and cohort effect process parameters, and subvectors for each of the latent age, period and cohort effects. The log-likelihood function is made up of several components:

- $l(\theta) = l_1(\theta) + l_{21}(\theta) + l_{22}(\theta) + l_{31}(\theta) + l_{32}(\theta)$
- $l_1(\theta) =$  Poisson log-likelihood for the observed deaths given the  $\beta_x^{(1i)}, \kappa_i^{(2i)}, \gamma_c^{(3i)}$  vectors. Cells,  $(i, t, x)$ , with missing data are given a weight of zero,  $W_i(t, x) = 0$ , otherwise  $W_i(t, x) = 1$ .

- $l_{21}(\theta)$  = unconditional log-likelihood for  $(R_2(1), S_2(1)) = (0, S_2(1))$ .
- $l_{22}(\theta)$  = conditional log-likelihood for  $(R_2(t), S_2(t))$  for  $t = 2, \dots, n_y$ .
- $l_{31}(\theta)$  = unconditional log-likelihood for  $X = (R_3(2), S_3(2), R_3(1), S_3(1))'$ .
- $l_{32}(\theta)$  = conditional log-likelihood for  $(R_3(c), S_3(c))$  for  $c = 3, \dots, n_c$ .

More specifically:

$$l_1(\theta) = \sum_{i,t,x} W_i(t,x) \{ D_i(t,x) \log m_i(t,x) - m_i(t,x) E_i(t,x) \} + \text{constant}$$

where  $m_i(t,x) = \exp[\beta_x^{(1i)} + n_a^{-1} \kappa_i^{(2i)} + n_a^{-1} \gamma_{t-x}^{(3i)}]$ .

$$l_{21}(\theta) = -\frac{1}{2} \log(V_{22}^{(2)} / (1 - \psi_{S2}^2)) - \frac{1}{2} S_2(1)^2 (1 - \psi_{S2}^2) / V_{22}^{(2)} + \text{constant}.$$

$$l_{22}(\theta) = -\frac{n_y - 1}{2} \log|V^{(2)}| - \frac{1}{2} \sum_{t=2}^{n_y} Y_2(t)' V^{(2)-1} Y_2(t) + \text{constant},$$

where  $Y_2(t) = (R_2(t) - R_2(t-1) - \mu_{R2}, S_2(t) - \psi_{S2} S_2(t-1))'$

$$l_{31}(\theta) = -\frac{1}{2} \log|\Omega| - \frac{1}{2} X' \Omega^{-1} X,$$

where  $X = (R_3(2), S_3(2), R_3(1), S_3(1))'$  and  $\Omega$  is the solution to the equation

$$\Omega = A\Omega A' + \tilde{V}^{(3)} \tag{5}$$

$$\text{where } A = \begin{pmatrix} \phi_{R31} + \phi_{R32} & 0 & -\phi_{R31} & \phi_{R32} & 0 \\ 0 & \phi_{S31} + \phi_{S32} & 0 & 0 & -\phi_{S31} & \phi_{S32} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\text{and } \tilde{V}^{(3)} = \begin{pmatrix} V_{11}^{(3)} & V_{12}^{(3)} & 0 & 0 \\ V_{21}^{(3)} & V_{22}^{(3)} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

While it is not possible to solve equation (5) using standard matrix algebra, the equation is linear in all elements of  $\Omega$  and can be solved by writing  $\Omega$  as a  $16 \times 1$  vector. Finally,

$$l_{32}(\theta) = -\frac{n_c - 2}{2} \log|V^{(3)}| - \frac{1}{2} \sum_{t=3}^{n_c} Y_3(c)' V^{(3)-1} Y_3(c) + \text{constant},$$

where  $Y_3(c) = (Y_{31}(c), Y_{32}(c))'$ ,

$$Y_{31}(c) = R_3(c) - (\phi_{R31} + \phi_{R32})R_3(c - 1) + \phi_{R31}\phi_{R32}R_3(c - 2),$$

$$Y_{32}(c) = S_3(c) - (\phi_{S31} + \phi_{S32})S_3(c - 1) + \phi_{S31}\phi_{S32}S_3(c - 2).$$

### 5.3. The prior distribution

In general, we aim to use prior distributions in our study that are as uninformative as possible, and therefore allow the data to speak for themselves. Therefore, unless otherwise stated below, all parameters (for example, the  $\beta_x^{(11)}$ ) have improper uniform prior distributions.

Our basic prior distribution assumes:

- $V^{(2)}$  has a relatively uninformative inverse-Wishart prior with density proportional to  $|V|^{-2.5} \exp[-\text{trace}(\Psi V^{-1})]$  where the inverse scale matrix  $\Psi = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$ .
- The prior for  $V^{(3)}$  is made up of two components, the second being described in more detail in Section 5.4. The first component is also a relatively uninformative inverse-Wishart prior with density proportional to  $|V|^{-2.5} \exp[-\text{trace}(\Psi V^{-1})]$  where the inverse scale matrix  $\Psi = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.02 \end{pmatrix}$ .
- $\mu_{R2} \sim N(-0.9, 0.9^2)$ ;
- $\text{logit}(\phi_{R31}) \sim N(2, 0.5^2)$  distribution.
- $\text{logit}(\phi_{R32}) \sim N(2, 0.5^2)$  distribution.
- $\text{logit}(\phi_{S31}) \sim N(2, 0.5^2)$  distribution.
- $\text{logit}(\phi_{S32}) \sim N(2, 0.5^2)$  distribution.
- $\text{logit}(\psi_{S2}) \sim \text{Gumbel}(2, 0.5)$  distribution.<sup>1</sup>

For  $V^{(2)}$  and  $V^{(3)}$ , the choice of prior means that the conditional posterior distribution is approximately inverse Wishart, and this is used as the candidate distribution in the MH algorithm to good effect (see Appendix A). The use of a non-zero, inverse scale matrix  $\Psi$  was found to be necessary to avoid singularities in the log-posterior distribution at  $V^{(3)} = 0$ . (The fact that the cohort effects are not directly observable means that the optimiser might achieve an infinite maximum likelihood if  $R_3(c)$  and  $S_3(c)$  are linear and  $V^{(3)} = 0$ .) A zero scale matrix for the inverse Wishart prior for  $V^{(2)}$  was not found to cause any problems for this pair of datasets. However, a slightly stronger, but still relatively uninformative prior might be needed for  $V^{(2)}$  for other pairs of populations.

A normal prior for  $\mu_{R2}$  ensures a normal conditional posterior. Again the prior is weak, but not completely uninformative. Specifically, we wish to avoid

<sup>1</sup> Let  $y = \text{logit}(\psi_{S2})$ . The density for the Gumbel( $\mu, \sigma$ ) distribution is  $\exp[-(y - \mu) / \sigma] \exp[-\exp[-(y - \mu) / \sigma]]$ .

$\mu_{R2}$  being too negative since this might result in decreasing cohort death rates over time (that is, decreasing  $m(t, x + t)$ ): something that we regard as being biologically unreasonable in the long run. In practice,  $\beta_x^{(1i)}$  increases roughly linearly at higher ages with a gradient of around 0.1 in most developed countries. It follows that we aim to avoid  $\mu_{RS}/n_a$  being less than  $-0.1$ , with  $n_a = 30$  as we have later on. The prior of  $N(-0.9, 0.9^2)$  assigns only a small prior probability of around 0.01 that  $\mu_{R2}/n_a < -0.1$ . As will be seen in Section 6, this prior does not seem to have a strong influence on the posterior distribution for  $\mu_{R2}$ , which has a significantly different mean and a substantially smaller standard deviation.

As noted in Appendix A, the exponentials of the age effects,  $\beta_x^{(1i)}$ , all have a gamma conditional posterior distribution, since the  $\beta_x^{(1i)}$  have no time series structure and a uniform prior. The same would be true for the period and cohort effects. However, the interplay of the exponential-gamma structure with the time series structure of these effects means that there is no analytical form for the conditional posterior distribution. Fortunately, the exponential-gamma structure is well approximated by a multivariate normal resulting in a multivariate normal being a good approximation overall to the conditional posteriors for the period and cohort effects. This again is used to good effect as the proposal distribution for the period effects and the cohort effects in turn.

The priors for the mean reversion parameters  $\psi_{S2}$ ,  $\phi_{R31}$ ,  $\phi_{R32}$ ,  $\phi_{S31}$  and  $\phi_{S32}$  all required some experimentation. All needed moderately informative priors: with limited time-series data and uninformative priors, the Markov chain typically spent too much time close to 1 (no mean reversion) to be comfortable with the core hypothesis in this paper. The normal priors for the logit( $\phi$ )'s were found to solve this problem without being too prescriptive apart from avoiding values close to 1.

The double exponential in the Gumbel prior density for  $\psi_{S2}$  was required to provide a stronger push away from  $\psi_{S2} = 1$  than the logit-normal priors used for the other parameters, but otherwise the Gumbel prior is not too strong through the choice of 2 and 0.5 for the Gumbel parameters.

In practical applications, the impact of these moderate priors for the autoregressive parameters tends to be modest except for long time horizons. For shorter time horizons, it is the correlations embedded in  $V^{(2)}$  and  $V^{(3)}$  that matter.

#### 5.4. Enhanced priors

Recall from Section 4.2 that it would be desirable to have similar short- and long-term volatility in period and cohort effects and to have long-term correlation in the cohort effects. With the basic priors above we found that the short- and long-term volatility of the cohort effects in the two populations were not consistent with these criteria. As a consequence we strengthened the priors by multiplying the basic prior density by further prior density functions as follows:



- A Gamma(100,100) prior for the ratio of the conditional 1-step-ahead variance of  $\gamma_c^{(32)}$  to the conditional 1-step-ahead variance of  $\gamma_c^{(31)}$ .
- A Gamma(100,100) prior for the ratio of the unconditional (i.e. long-term) variance of  $\gamma_c^{(32)}$  to the unconditional variance of  $\gamma_c^{(31)}$ .
- A Beta(20,2) prior (scaled to cover the interval  $(-1, +1)$ ) for the unconditional correlation between  $\gamma_c^{(32)}$  and  $\gamma_c^{(31)}$ .

These might seem relatively strong, but their inclusion or exclusion was not found to have a significant impact on headline outputs such as the distribution (central trend and spread) of future death rates. For each prior, the important elements are the mean and standard deviation of the prior, and the domain. Thus, for example, the Gamma priors properly restrict variances to be positive real numbers, and have a mean of 1 and standard deviation of 0.1. The log-normal as a prior with the same mean and standard deviation is almost identical to the Gamma and would give similar results. The scaled Beta properly restricts the unconditional correlation to the range  $(-1, +1)$ , and has mean of 0.82 and standard deviation of 0.06.

These particular parameterisations for the basic and enhanced priors might need to be adjusted to suit the specific characteristics of a given pair of populations. However, all pairs of populations considered so far in this paper (EW versus CMI males) and elsewhere (EW versus CMI females, and EW versus Scottish males) work with the same priors as listed.

## 5.5. Metropolis-Hastings algorithm

Details of how the MH algorithm is implemented are given in Appendix A. As before, we consider  $\theta$  to be the complete vector of process parameters, and age, period and cohort effects. The algorithm generates a sequence of values for the parameter vector  $\theta(1), \theta(2), \dots, \theta(t)$  (the Markov chain). With a properly implemented MH algorithm the empirical distribution of the  $\theta(i)$  will converge to the full posterior distribution of the unknown  $\theta$ .

## 6. RESULTS AND DISCUSSION

Our analysis focuses on EW versus CMI males using data from 1961 to 2005 and ages 60 to 89. This analysis includes the enhanced priors discussed in Section 5.4.

Let  $\theta(i, j)$  be the  $j$ 'th element of the Markov chain  $\theta(i)$  after  $i$  iterations. The *burn-in* period is the initial phase of the iterative scheme where we move from the initial  $\theta(0)$  towards the posterior distribution of  $\theta$ . After  $i_B$  iterations, we consider the burn-in period to be complete and that further iterations are cycling around the posterior distribution. After  $i_B$ , we record every 50th iteration  $\theta(i_B + 50), \theta(i_B + 100), \dots$ . Taking every 50th observation results in efficient use of memory, but it also reduces substantially the degree of autocorrelation between successive recorded observations.

When we wish to simulate one future sample path of the two-population APC model we choose one of the  $\theta(i_B + 50k)$  at random and then use this to specify the process parameters and historical state variables for simulating that sample path. Further details are given in Appendix B.

We now illustrate the results in a series of figures, and comment as follows:

- Figure 2 provides fan charts (with 5% quantile bands) for historical and forecast mortality at ages 65, 75 and 85 for EW and CMI males. For the years 1961 to 2005, the outer limits of the fans provide us with 90% credibility intervals for the underlying mortality rate,  $q_i(t, x) = 1 - \exp[-m_i(t, x)]$ , in each year. The 90% credibility interval is bounded by the 5% and 95% quantiles of the marginal posterior distribution for each  $q_i(t, x)$ .

The left-hand plots show forecasts generated using the original single-population APC model, fitted using the two-stage approach and with no parameter uncertainty (PC), as in Cairns et al. (2011a). The right-hand plots show forecasts generated using the new joint 2-population model with full parameter uncertainty (PU).

In contrast to the original two-stage approach to model fitting and projection, the underlying historical  $q_i(t, x)$  are no longer simply point estimates. Instead, we can see the degree of uncertainty that is associated with each  $q_i(t, x)$  between 1961 and 2005. In particular, we can see that the credibility intervals for the CMI data up to 2005 are significantly wider than for the EW data, reflecting the smaller size of the CMI dataset.

For age 85, the EW fan widens out slightly in the 1960's. This reflects the fact that we do not have data for ages 85-89 for 1961-1970, and so what we see here is a backwards extrapolation to age 85 that learns from EW ages 85-89 after 1970 and from ages 60-84 between 1961 and 1970.

Looking beyond 2005, the fans are based on Monte Carlo simulations and spread out reflecting growing future uncertainty. We can see that the CMI fans (especially at age 65) move closer to the EW fan, but the spread between the EW and CMI populations is maintained and stabilises after 30 years. The average gap in 2050 (right-hand end) is similar to what it was in 1961 (left-hand end).

- In Figure 3, we compare fan charts produced using the MCMC two-population model with fan charts produced individually for the two populations using the old two-stage approach, with no allowance for parameter uncertainty, and an ARIMA(1,1,0) model for the cohort effect.

The left-hand plots (EW top; CMI bottom) show fans for age 75 using the MCMC approach. We can see the relatively smooth central forecasts in each case. The left hand plots illustrate the impact of including parameter uncertainty (PU). The parameter certain (PC) case (B: upper, narrower fans) takes the means of all process parameters and state variables from the MCMC output and takes these as point estimates for conducting simulations. Central forecasts are about the same, but the PU fans (A) are significantly wider.

In the right-hand plots, the underlying fans A and B are the same as on the left. These now have superimposed on them fans (C) produced using the

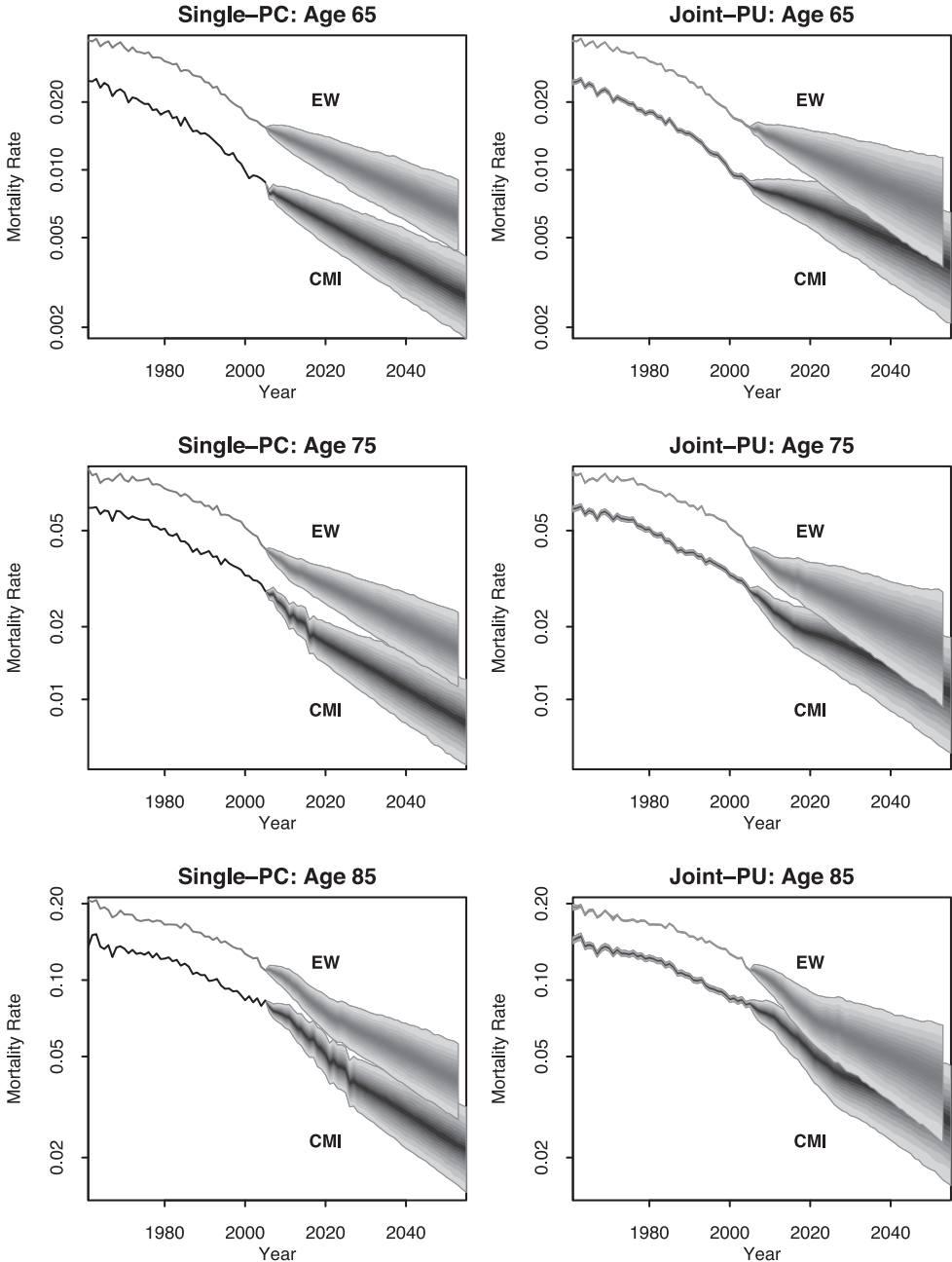


FIGURE 2: Mortality fan charts for ages 65, 75 and 85 for EW males (upper fans) and CMI males (lower fans). Left-hand plots: fan charts constructed using the single-population models with no parameter uncertainty (PC). Right-hand plots: fan charts constructed using the joint-population model (MCMC) with parameter uncertainty (PU).

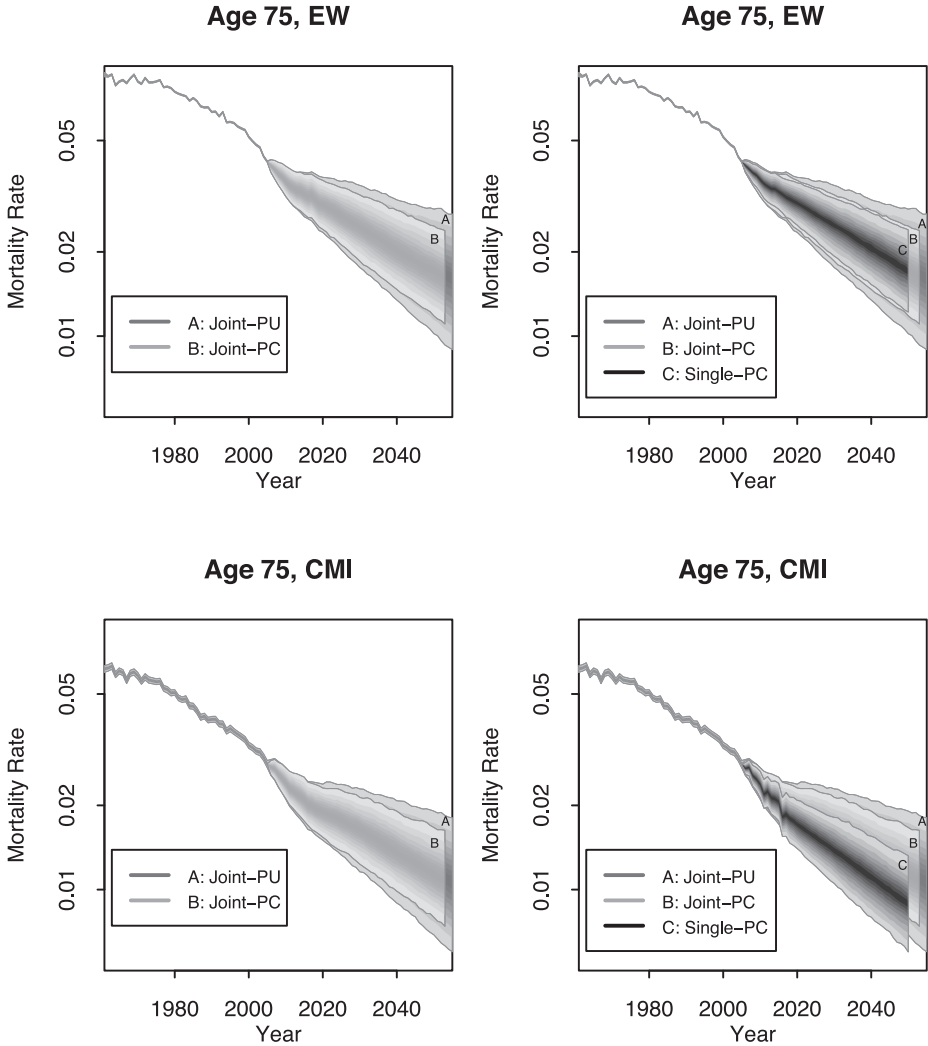


FIGURE 3: Comparison of fan charts based on A: the new MCMC algorithm with full parameter uncertainty (PU) (rear fans), B: the new MCMC algorithm with parameters certain (PC) (upper (left hand plots) or middle (right hand plots) fans), C: the original two-stage method with parameters certain using ARIMA(1,1,0) models for the cohort effect (upper fans in the right-hand plots). Left-hand plots compare A and B. Right-hand plots compare A, B and C.

single-population models (Cairns et al., 2011a). For the EW data, fan A is reasonably smooth and the results are reasonably consistent with the single-population EW forecasts. However, fan A is wider, reflecting the allowance for parameter uncertainty (parameter uncertainty being a by-product of the MCMC output). For the EW data, both of the PC cases (B and C) have fans of similar width.

The CMI plot (bottom right) differs in three important ways. First, the single-population fan C is much less smooth. For age 75, the first 15 years of projections include values for the cohort effect that have been inferred from the historical data. The small size of the CMI population results in noisy estimates of the cohort effect and this feeds through, in an unreasonable way, to the forecasts. Fans A and B, by contrast (bottom left), are smoother and much more plausible, and reflect the use of a single-stage rather than a two-stage estimation process. Second, for the CMI data there is a greater difference between the central trends in fans A and C. This reflects mean reversion in the spread between the two populations when modelled simultaneously. The greater size of the EW population means that the mean-reversion has a greater impact on the CMI projections. With the single-population projections, CMI mortality rates with a faster improvement rate were gradually diverging from EW rates, whereas, here, mean reversion pulls the CMI forecasts back towards the EW population. Third, consider the width of the parameter certain fans (B and C). For EW, these were about the same width, while, for the CMI data, the fan B is slightly narrower. In Figure 1, we saw that the single-population model resulted in rather noisy estimates for the cohort effect for the CMI data. This fed through to greater noise in the forecasts. A key contribution of the approach to modelling two populations using a single-stage estimation procedure is that it substantially dampens the noise observed in Figure 1 which results in narrower fans and hence more “confident” predictions.

- In Figure 4, we have plotted fans (90% credibility intervals) for age, period and cohort effects for the two populations. Before plotting, outputs from the MCMC program were adjusted to satisfy the following identifiability constraints:  $\sum_c R_3(c) = 0$ ,  $\sum_c R_3(c)(c - \bar{c}) = 0$ ,  $\sum_c S_3(c) = 0$ ,  $\sum_t R_2(t) = 0$  and  $\sum_t S_2(t) = 0$ . This involves shifting and tilting relevant outputs, and also making corresponding adjustments to the random process parameters. For example, shifting and tilting  $R_3(c)$  to satisfy the first two constraints means that we must make a corresponding tilt to  $R_2(t)$ , an identical shift and tilt to  $\beta_x^{(11)}$  and  $\beta_x^{(12)}$ , and adjustments to  $\mu_{R2}$ ,  $\mu_{R3}$  and  $\delta_{R3}$ . For each of the age, period and cohort effects, we can see that the CMI credibility intervals are rather wider, reflecting the smaller size of the CMI population.

For the cohort effects, we see that both the EW and CMI fans widen out towards both ends. This reflects the number of cells available for estimating a given cohort effect: for example, we have just one cell linked to the 1945 birth cohort, compared with 30 cells for the 1915 cohort. The EW fan widens out more on the left than on the right, reflecting the missing data for ages 85-89 in years 1961-1970.

In the bottom right plot, we consider the central cohort effect,  $R_3(c)$ , and its linear trend,  $\mu_{R3} + \delta_{R3}(c - \bar{c})$ . The upper fan provides credibility intervals for  $R_3(c)$  (a repeat from the bottom left plot). The wide fan in the background provides credibility intervals for the trend,  $\mu_{R3} + \delta_{R3}(c - \bar{c})$ . Clearly,

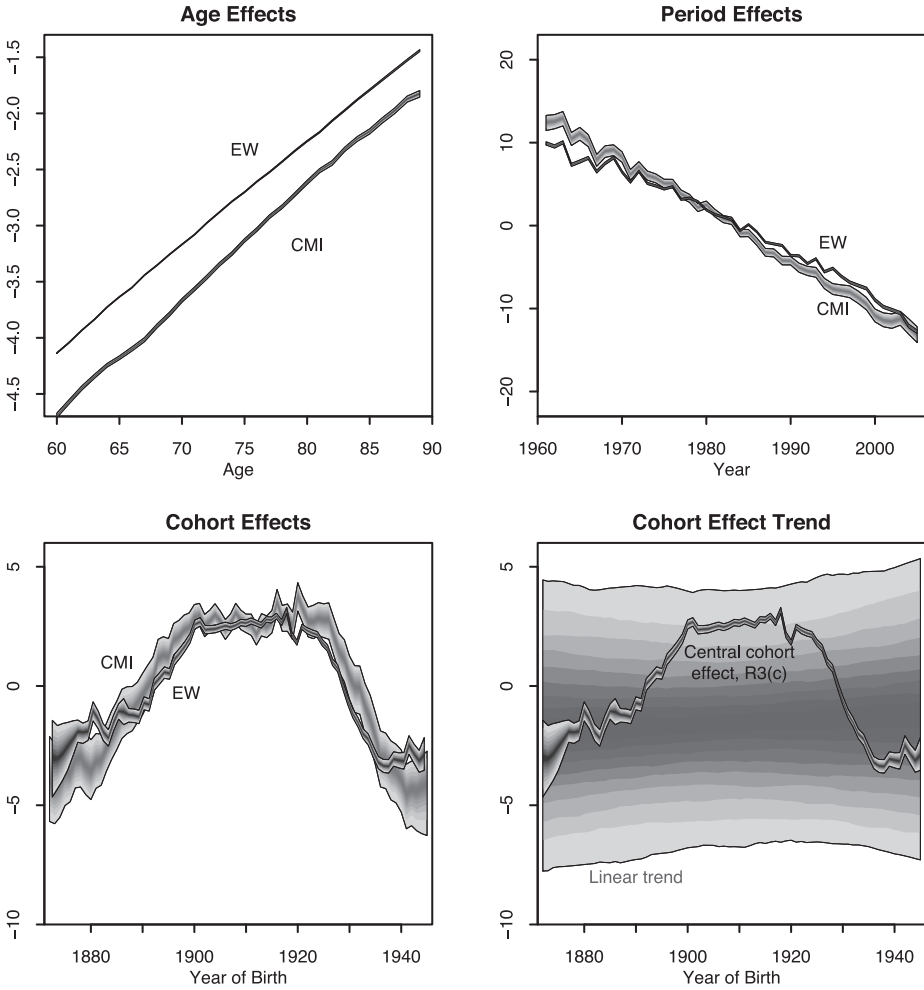


FIGURE 4: Age, period and cohort effects for EW males (upper fans) and CMI males (lower fans). Bottom right: EW cohort effect (upper, narrow fan) and its underlying linear trend (rear, wide fan). Age, period and cohort effects have been adjusted to satisfy identifiability constraints:  $\sum_c R_3(c) = 0$ ,  $\sum_c R_3(c)(c - \bar{c}) = 0$ ,  $\sum_c S_3(c) = 0$ ,  $\sum_t R_2(t) = 0$ ,  $\sum_t S_2(t) = 0$ .  $R_3(c)$  is modelled as an AR(2) process around the linear trend  $\mu_{R_3} + \delta_{R_3}(c - \bar{c})$ .

there is considerable uncertainty in this trend, reflecting the relatively modest number of (latent) observations of  $R_3(c)$ . In the short run, this does not cause significant problems as there is only gentle mean reversion to this long-term linear trend. In the long run (say, 40 or 50 years), this will result in some additional uncertainty in the overall level of mortality.

- In Figure 5, we plot the empirical correlation between the simulated improvement factors at ages 65, 75 and 85 as a function of the time horizon.

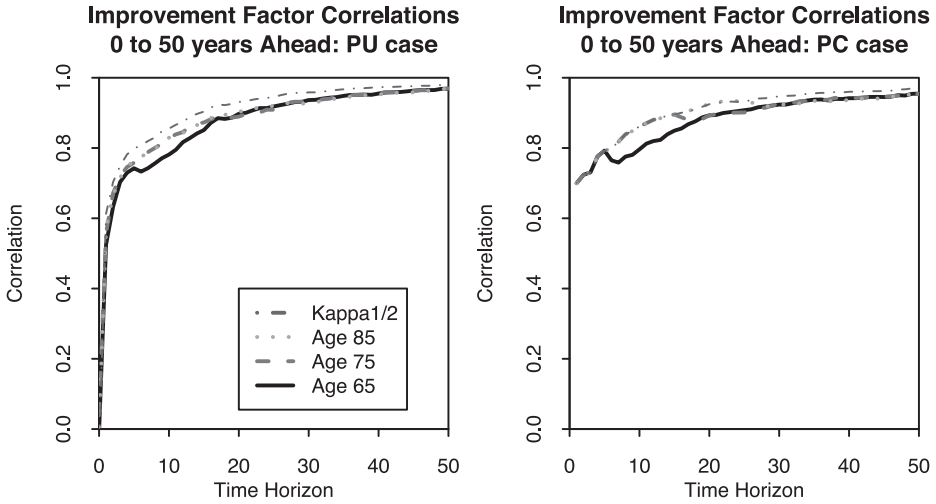


FIGURE 5: Correlation between simulated improvement factors in EW males and CMI males mortality rates as a function of the time horizon beyond 2005 – Age 65 (solid line), 75 (dashed line) and 85 (dotted line) – and correlation between period effects  $\kappa_t^{(21)}$  and  $\kappa_t^{(22)}$  as a function of the time horizon (dot-dashed line). Left: simulations incorporating parameter uncertainty (PU). Right: simulations with no parameter uncertainty (PC).

For reference, we also plot the correlation between the period effects,  $\kappa_t^{(21)}$  and  $\kappa_t^{(22)}$  (the uppermost line in both plots).

Correlations that reflect full allowance in the simulations for parameter uncertainty are given in the left-hand plot of Figure 5. To help understand the structure in the left-hand plot, however, it is more straightforward to consider, first, the PC case (right-hand plot). In this case, we took the MCMC output and used the mean of each process parameter and also the mean of each of the latent effects. For time horizons of up to 5 years, the age 65, 75 and 85 correlations are all equal: since randomness in each depends only on randomness in the period effects. After 5 years, the age 65 mortality rate includes randomness in the cohort effect that requires simulation beyond the cohorts in our historical dataset. This additional randomness results in a different and, here, lower correlation for age 65 compared with ages 75 and 85. (The additional randomness contributed by the cohort effect could push the overall correlation being measured here up or down. Here, it goes down because the short-term correlation between the EW and CMI cohort effects is lower than the short-term correlation between the respective period effects). After 15 years, the age 75 mortality also includes simulated cohort effects, so the age 75 and 85 correlations also diverge. In the long run, the correlation between  $\kappa_t^{(21)}$  and  $\kappa_t^{(22)}$  dominates as mean reversion in the cohort effects reduces their relative impact over time. Finally, mean reversion in the spread between  $\kappa_t^{(21)}$  and  $\kappa_t^{(22)}$  means that the correlation between the simulated improvement factors will tend to 1 as the forecast time horizon increases.

The right-hand plot also shows us that the correlations follow quite closely the correlations between  $\kappa_i^{(21)}$  and  $\kappa_i^{(22)}$ , with deviations only when the cohort effect is uncertain.

The PU case allows for uncertainty in both the process parameters and in the values of the historical latent period and cohort effects (left-hand plot). For the final year of birth in our historical dataset, we only have one observation (age 60 in 2005 for the 1945 cohort), leading to a relatively large amount of uncertainty in the estimate of the 1945 cohort effect. In more general terms, there is growing uncertainty in the cohort effect as we approach 1945 (recall Figure 4). The correlation between estimates of the latent cohort effect for these years is quite small and has the immediate effect of dragging down the age 65 correlation plot (Figure 5, left) relative to its PC counterpart (Figure 5, right). In the longer run, this uncertainty in estimates of historical latent effects is replaced by uncertainty in the process parameters such as  $\mu_{R2}$ . This can push correlation up or down. Here, the parameter uncertainty pushes correlation down initially, but for longer maturities, the correlation is slightly higher in the PU plot reflecting the common dependence of  $\kappa_i^{(21)}$  and  $\kappa_i^{(22)}$  on the uncertain random-walk drift  $\mu_{R2}$ .

A significant difference between the left- and right-hand plots in Figure 5 is that on the left the correlation between  $\kappa_i^{(21)}$  and  $\kappa_i^{(22)}$  is substantially above the age 65, 75 and 85 correlations. This tells us that uncertainty in estimates of the historical age and cohort effects has a significant downwards effect on correlation.

The unambiguous conclusion from this plot is that correlations are rising with the time horizon. Additionally, though, the shape of the curve and the values that we see here, based on our very specific model, are consistent with the model-free empirical findings of Coughlan et al. (2011).

- As a final remark, we compared results with and without the enhanced priors discussed in Section 5.4. For the plots discussed above, the use of the enhanced prior did not result in any significant changes, indicating that our conclusions about headline aspects of forecasting are robust relative to this feature.

However, if we plot relevant statistics linked to the enhanced priors (Figure 6) we can see that the enhanced priors do exactly what we intend them to do by pulling the short- and long-term variances of the cohort effect closer together. In the left-hand panel, we plot the CDFs of the ratio of the short-term (one-step-ahead conditional) variances ( $Var(\gamma_c^{(32)} | \mathcal{G}_{c-1}, V^{(3)}) / Var(\gamma_c^{(31)} | \mathcal{G}_{c-1}, V^{(3)})$ ), where  $\mathcal{G}_{c-1}$  is the history of  $R_3(u)$  and  $S_3(u)$  up to time  $c-1$ , and  $V^{(3)}$  is sampled at random from the posterior distribution of  $\theta$ ) of the population 1 and 2 cohort effects with and without the enhanced prior. Without the enhanced prior, population 2 conditional variances are much higher suggesting that Poisson sampling variation in population 2 is still having a strong influence on estimates of the underlying cohort effect (Figure 1, right). With the enhanced prior we can infer that this issue effectively disappears. In the middle panel, we plot the CDF of the ratio of the long-term



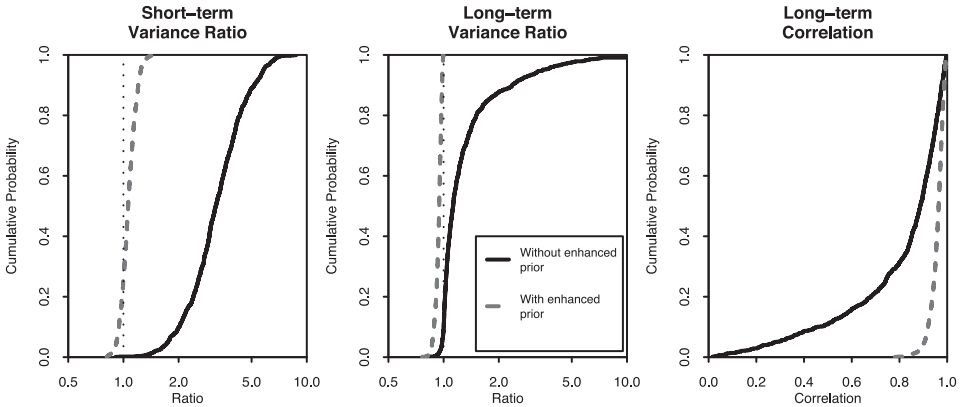


FIGURE 6: Left: Cumulative posterior distribution function (CPDF) of the ratio of the short-term volatility of  $\gamma_c^{(32)}$  to the short-term volatility of  $\gamma_c^{(31)}$ . Middle: CPDF of the ratio of the long-term (stationary) variance of  $\gamma_c^{(32)}$  to the long-term variance of  $\gamma_c^{(31)}$ . Right: Long-term correlation between  $\gamma_c^{(32)}$  and  $\gamma_c^{(31)}$ . Solid lines: without the use of the enhanced prior distributions. Dashed lines: with the use of the enhanced prior distributions.

(unconditional) variances ( $Var(\gamma_c^{(32)}) / Var(\gamma_c^{(31)})$ ) between the two cohort effects. The right-hand panel shows the corresponding long-term correlation ( $cor(\gamma_c^{(32)}, \gamma_c^{(31)})$ ). The inclusion of the enhanced prior can be seen to have the desired effect of stabilising the relationship between the variances in the two populations.

### 7. CONCLUSIONS

The Bayesian Markov chain Monte Carlo approach to estimating jointly the parameters of stochastic mortality models for two related populations has clear advantages over the individual modelling of the populations. First, the difference between the two populations’ death rates is modelled as a mean-reverting stochastic spread, which allows for different short-term trends in improvement rates, but parallel improvements in the long run, thereby preventing a biologically implausible long-term divergence in death rates. Second, the approach permits us to analyse uncertainty in the estimates of the historical age, period and cohort effects, and this helps us to smooth out noise in parameter estimates, particularly those relating to cohort effects, attributable to small populations: the framework is especially valuable when the population of interest is smaller with more volatile mortality than the other population. Third, the forecasts of mortality rates arising from this framework provide consistent central projections, as well as consistent distributions (fans) around these central projections. The bottom right-hand panel of Figure 3, for example, shows how the fan chart projections for the smaller population (in this case CMI males) is “pulled” towards that of the larger population (in this case EW males

– see top right-hand panel in the figure) without any increase in forecast uncertainty – the width of the fans in the two-population case is actually slightly less for both populations than the fans in the single-population case. Fourth, the correlations between the estimated mortality improvement factors for two populations are consistent with the historical data.

However, the approach remains sensitive to the underlying stochastic mortality models used and will compound any weaknesses in these models. For example, the approach might be sensitive to the model used to estimate the cohort effect. Finally, the approach is sensitive to the amount of data used, although the amount of data is a less significant factor than the fact that two related populations are being modelled jointly.

Some of the details internal to the model should not be regarded as cast in stone. The 0/1 weights in the spreads model might be varied if the two populations are more similar in size compared to those considered here. The prior distributions might be varied from those considered here if initial results produce results that are implausible in some way (for example, priors might be required to ensure that the age effects,  $\beta_x^{(li)}$ , are reasonably smooth). So users of the approach must always be vigilant, analyse results carefully, and not use the model as a black box.

Overall, we can conclude that the MCMC framework is a very useful one for modelling related populations, particularly the basis risk between them. This is especially important when we are interested in hedging the longevity risk in the smaller population using an index hedging contract related to the larger population. Such a hedge analysis will be the subject of future work (Cairns et al., 2011b).

#### ACKNOWLEDGMENTS

We are grateful to the Continuous Mortality Investigation in the UK for providing the assured lives data. AC wishes to thank Iain Currie and Stephen Richards for useful discussions on multi-population mortality. Finally, the authors wish to thank the referees of the paper for their helpful and insightful comments on the original submission.

#### REFERENCES

- BIATAT, V.D. and CURRIE, I.D. (2010) *Joint models for classification and comparison of mortality in different countries*. Proceedings of 25rd International Workshop on Statistical Modelling, Glasgow, 89-94.
- 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.
- BOOTH, H., MAINDONALD, J. and SMITH, L. (2002a) Applying Lee-Carter under conditions of variable mortality decline. *Population Studies*, **56**: 325-336.
- BOOTH, H., MAINDONALD, J. and SMITH, L. (2002b) *Age-time interactions in mortality projection: Applying Lee-Carter to Australia*. Working Papers in Demography, The Australian National University.

- BOOTH, H., HYNDMAN, R.J., TICKLE, L. and DE JONG, P. (2006) Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions. *Demographic Research*, **15**: 289-310.
- BOOTH, H. and TICKLE, L. (2008) Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, **3**: 3-43.
- BRAY, I. (2002) Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality. *Applied Statistics*, **51**: 151-164.
- 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., BLAKE, D. and DOWD, K. (2006a) Pricing death: Frameworks for the valuation and securitization of mortality risk. *ASTIN Bulletin*, **36**: 79-120.
- CAIRNS, A.J.G., BLAKE, D. and DOWD, K. (2006b) 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. and DOWD, K. (2008) Modelling and management of mortality risk: A review. *Scandinavian Actuarial Journal*, **2008(2-3)**: 79-113.
- 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., EPSTEIN, D. and KHALAF-ALLAH, M. (2011a) Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, **48**: 355-367.
- CAIRNS, A.J.G., BLAKE, D., DOWD, K. and COUGHLAN, G.D. (2011b) *Longevity hedge effectiveness: A decomposition*. Working paper, Heriot-Watt University.
- CARTER, L. and LEE, R.D. (1992) Modelling and forecasting US sex differentials in mortality. *International Journal of Forecasting*, **8**: 393-411.
- COUGHLAN, G., EPSTEIN, D., ONG, A., SINHA, A., HEVIA-PORTOCARRERO, J., GINGRICH, E., KHALAF-ALLAH, M. and JOSEPH, P. (2007a) *LifeMetrics: A toolkit for measuring and managing longevity and mortality risks*. Technical document. Available at [www.lifemetrics.com](http://www.lifemetrics.com).
- COUGHLAN, G., EPSTEIN, D., SINHA, A. and HONIG, P. (2007b) *q-Forwards: Derivatives for transferring longevity and mortality risk*. Available at [www.lifemetrics.com](http://www.lifemetrics.com).
- COUGHLAN, G.D. (2009). Longevity risk transfer: Indices and capital market solutions. In Barrieu, P.M. and Albertini, L. (eds), *The Handbook of Insurance Linked Securities*, Wiley, London.
- COUGHLAN, G.D., KHALAF-ALLAH, M., YE, Y., KUMAR, S., CAIRNS, A.J.G., BLAKE, D. and DOWD, K. (2011) Longevity hedging 101: A framework for longevity basis risk analysis and hedge effectiveness. To appear in *North American Actuarial Journal*.
- CURRIE, I.D., DURBAN, M. and EILERS, P.H.C. (2004) Smoothing and forecasting mortality rates. *Statistical Modelling*, **4**: 279-298.
- CZADO, C., DELWARDE, A. and DENUIT, M. (2005) Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics*, **36**: 260-284.
- DAHL, M., MELCHIOR, M. and MØLLER, T. (2008) On systematic mortality risk and risk minimisation with survivor swaps. *Scandinavian Actuarial Journal*, **2008(2-3)**: 114-146.
- DAHL, M., GLAR, S. and MØLLER, T. (2011) Mixed dynamic and static risk minimization with an application to survivor swaps. To appear in *European Actuarial Journal*.
- DOWD, K., CAIRNS, A.J.G., BLAKE, D., COUGHLAN, G.D., EPSTEIN, D. and KHALAF-ALLAH, M. (2010a). Evaluating the goodness of fit of stochastic mortality models. *Insurance: Mathematics and Economics*, **47**: 255-265.
- DOWD, K., CAIRNS, A.J.G., BLAKE, D., COUGHLAN, G.D., EPSTEIN, D. and KHALAF-ALLAH, M. (2010b). Backtesting stochastic mortality models: An ex-post evaluation of multi-period-ahead density forecasts. *North American Actuarial Journal*, **14**: 281-298.
- GILKS, W.R., RICHARDSON, S. and SPIEGELHALTER, S.J. (1996) *Markov chain Monte Carlo in Practice*. Chapman and Hall, New York.
- GIROSI, F. and KING, G. (2008) *Demographic Forecasting*. Princeton University Press, Princeton.
- 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.
- JACOBSEN, R., KEIDING, N. and LYNGE, E. (2002) Long-term mortality trends behind low life expectancy of Danish women. *J. Epidemiol. Community Health*, **56**: 205-208.

- JARNER, S.F. and KRYGER, E.M. (2011) *Modelling adult mortality in small populations: The SAINT model*. To appear in *ASTIN Bulletin*.
- KOGURE, A., KITSUKAWA, K. and KURACHI, Y. (2009) A Bayesian comparison of models for changing mortalities toward evaluating longevity risk in Japan. *Asia Pacific Journal of Risk and Insurance*, **3**(2): 1-21.
- KOGURE, A. and KURACHI, Y. (2010) A Bayesian approach to pricing longevity risk based on risk-neutral predictive distributions. *Insurance: Mathematics and Economics*, **46**: 162-172.
- LI, J.S.-H. and HARDY, M.R. (2009) *Measuring basis risk involved in longevity hedges*. Working paper, University of Waterloo.
- 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.
- LI, N., LEE, R. and TULJAPURKAR, S. (2004) Using the Lee-Carter method to forecast mortality for populations with limited data. *International Statistical Review*, **72**: 19-36.
- LI, J.S.-H., HARDY, M.R. and TAN, K.S. (2009) Uncertainty in mortality forecasting: An extension to the classic Lee-Carter approach. *ASTIN Bulletin*, **39**: 137-164.
- LOEYS, J., PANIGIRTZOGLOU, N. and RIBEIRO, R.M. (2007) *Longevity: A market in the making*. Available at [www.lifemetrics.com](http://www.lifemetrics.com).
- MACDONALD, A.S., CAIRNS, A.J.G., GWILT, P.L. and MILLER, K.A., (1998) An international comparison of recent trends in population mortality. *British Actuarial Journal* **4**: 3-141.
- OEPPE, J. and VAUPEL, J.W. (2002) Broken limits to life expectancy. *Science*, **296**: 1029-1030.
- OLIVIERI, A. and PITACCO, E. (2009) Stochastic mortality: the impact on target capital. *ASTIN Bulletin*, **39**: 541-563.
- OSMOND, C. (1985) Using age, period and cohort models to estimate future mortality rates. *International Journal of Epidemiology*, **14**: 124-129.
- OSMOND, C. and GARDNER, M.J. (1982) Age, period and cohort models applied to cancer mortality rates. *Statistics in Medicine*, **1**: 245-259.
- PEDROZA, C. (2006) A Bayesian forecasting model: Predicting U.S. male mortality. *Biostatistics*, **7**: 530-550.
- PITACCO, E., DENUIT, M., HABERMAN, S. and OLIVIERI, A. (2009) *Modelling longevity dynamics for pensions and annuity business*. Oxford University Press, Oxford.
- PLAT, R. (2009) Stochastic portfolio specific mortality and the quantification of mortality basis risk. *Insurance: Mathematics and Economics*, **45**: 123-132.
- REICHMUTH, W. and SARFERAZ, S. (2008) *Bayesian demographic modelling and forecasting: An application to US mortality*. SFB 649 Discussion paper 2008-052.
- RENSHAW, A.E. and HABERMAN, S. (2003) Lee-Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, **33**: 255-272.
- RENSHAW, A.E. and HABERMAN, S. (2006) A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, **38**: 556-570.
- TULJAPURKAR, S., LI, N. and BOE, C. (2000) A universal pattern of mortality change in G7 countries. *Nature* **405**: 789-792.

## A. THE METROPOLIS-HASTINGS ALGORITHM

A popular approach to tackling Bayesian model fitting problems uses the Metropolis-Hastings (MH) algorithm:

- Vector  $\theta(i)$  = current set of parameter and latent variable values after  $i$  iterations.
- $D$  = observed data.
- $p(\theta|D)$  = posterior density for  $\theta$ . The posterior distribution is sufficiently complex that direct simulation from  $p(\theta|D)$  is impossible.

- Iteration  $i + 1$  proceeds in a series of substeps,  $j = 1, 2, \dots$ 
  - Substep  $j$  updates a single element or a block of the vector  $\theta$ .
  - $\tilde{\theta}$  = latest  $\theta$  including accepted substep updates.
  - Generate a candidate  $\hat{\theta}$  from a candidate distribution with density  $f(\theta|\tilde{\theta})$
  - Accept the candidate  $\hat{\theta}$  to replace the current  $\tilde{\theta}$  with probability

$$\alpha = \min \left\{ \frac{p(\hat{\theta}|D) f(\tilde{\theta}|\hat{\theta})}{p(\tilde{\theta}|D) f(\hat{\theta}|\tilde{\theta})}, 1 \right\}, \tag{6}$$

otherwise stick with  $\tilde{\theta}$ .

- At the end of this cycle of substeps, record the updated  $\theta(i + 1)$ .

The MH algorithm is such that although the  $\theta(i)$  are highly autocorrelated, their stationary distribution is equal to the posterior distribution  $p(\theta|D)$ . It follows that if we run the MH algorithm for a long time, then the empirical distribution of the observed  $\theta(i)$  for  $i = 1, \dots, N$  will be a good approximation to the true posterior.

### A.1. The Gibbs sampler

The Gibbs sampler is a special case of the MH algorithm under which the candidate distribution is exactly equal to the conditional posterior distribution for a subset of the parameters, conditional on the current values of all other parameters in the model. Typically, it is not possible for us to know, or at least to be able to sample from the full posterior distribution (if we could, we would not need to use the MH algorithm). However, the conditional posterior for subsets of parameters is often a standard distribution from which we can simulate.

Under the Gibbs sampler, the acceptance probability,  $\alpha$  (see equation 6), is always equal to 1. This is an advantage if it is computationally expensive to compute the full log-likelihood function: if it is known that the acceptance probability is 1, then the log-likelihood does not need to be computed. A bigger advantage of the Gibbs sampler in this study, though, is that the Markov chain mixes more quickly through the full posterior distribution. We can, therefore, obtain a more reliable sample from the posterior in less time.

### A.2. Pseudo-Gibbs sampler

Often, the conditional posterior is not in a form that can be matched to a standard distribution that can be easily simulated from. In some of these cases, however, we can simulate from a simpler distribution that is still a good approximation to the true conditional posterior. The acceptance probability (equation 6) will now be different from 1, and so the full likelihood needs to be evaluated. If such an approximation can be found then it often results in efficient mixing.

### A.3. Outline of the Metropolis-Hastings (MH) implementation

A full run of the MH algorithm consists of a large number of iterations (typically 50,000 or more). As described above, within each iteration we carry out a number of substeps to update, or leave unchanged, the various parameters. In our description of the substeps below we will label each step as *Gibbs* or *pseudo-Gibbs* or *MH* according to whether the candidate distribution is, respectively, the exact conditional posterior distribution, an approximation to the conditional posterior distribution, or a simpler candidate distribution.

1. Update the vectors  $\beta_x^{(11)}$  and  $\beta_x^{(12)}$  (Gibbs). In both cases, the  $\exp(\beta_x^{(1k)})$  can be shown to have a Gamma conditional posterior distribution. This is simple to generate, and so we can use the Gibbs sampler to update the  $\beta_x^{(11)}$  and  $\beta_x^{(12)}$  with each candidate vector having an acceptance probability that is always exactly 1.
2. Update simultaneously the vectors  $R_2(t)$  and  $S_2(t)$  (pseudo-Gibbs). Due to the influence of the Poisson likelihood, the conditional posterior distribution cannot be identified exactly. However, we can derive a good multivariate approximation to the conditional posterior and we use this to generate candidate vectors for  $R_2(t)$  and  $S_2(t)$ . Since this is a pseudo-Gibbs step an acceptance probability must be calculated and a random decision made to accept or reject the candidate.
3. Update simultaneously the vectors  $R_3(c)$  and  $S_3(c)$  (pseudo-Gibbs). The same remarks for  $R_2(t)$  and  $S_2(t)$  apply.
4. Update individually and in sequence  $\psi_{S2}$ ,  $\phi_{R31}$ ,  $\phi_{R32}$ ,  $\phi_{S31}$  and  $\phi_{S32}$  (MH). None of these have a straightforward exact or approximate conditional posterior. For each, the candidate distribution is an independent normal distribution for the logit transform of the parameter centred on the current value.
5. Update  $V^{(2)}$  (pseudo-Gibbs). For the approximation we use the inverse-Wishart distribution based on the annual changes in  $(R_2(t), S_2(t))$ . The approximation results from the exclusion of the initial  $(R_2(1), S_2(1))$ .
6. Update  $\mu_{R2}$  (Gibbs). The conditional posterior is a normal distribution.
7. Update  $V^{(3)}$  (pseudo-Gibbs). For the approximation we use the inverse-Wishart distribution based on the annual changes in  $(R_3(c), S_3(c))$ . The approximation results from the exclusion of the initial  $(R_3(c), S_3(c))$  for  $c = 1, 2$ .
8. Apply a (small) random shift and tilt to the vector of  $R_3(c)$ 's with compensatory adjustments to  $R_2(t)$ ,  $\beta_x^{(11)}$  and  $\beta_x^{(12)}$  in such a way that the underlying death rates are unchanged (MH). The Poisson likelihood is unaltered, but there is an impact on the time-series likelihoods.
9. Apply a (small) random shift and tilt to the vector of  $S_3(c)$ 's with compensatory adjustments to  $S_2(t)$  and  $\beta_x^{(12)}$  in such a way that the underlying death rates are unchanged (MH). The Poisson likelihood is unaltered, but there is an impact on the time-series likelihoods.

The final two substeps, 8 and 9, are not necessary for the MH algorithm to work. However, it was found that these additional randomisations helped the

Markov chain to mix more quickly, and, therefore, for the results to converge more quickly to a satisfactory solution. Identifiability constraints mean that  $\mu_{S2} = \mu_{R3} = \delta_{R3} = 0$  remain fixed and do not need updating.

B. SIMULATION OF FUTURE SAMPLE PATHS

Let  $\tilde{\theta}(k)$ , for  $k = 1, \dots, N$  be the  $k$ 'th recorded value of the parameter vector  $\theta$  out of the original Markov chain  $\theta(i)$  (here, every 50th value was recorded after completion of the burn-in phase). Suppose that we wish to generate  $M$  sample paths of future death rates and mortality rates. For scenario  $j$  we proceed as follows:

- Select  $K(j)$  at random from the integers  $\{1, \dots, N\}$  independently of all other values of  $K(1), K(2), \dots$
- Let  $\theta_j = \tilde{\theta}(K(j))$ .
- Generate a random sample path for future values of  $(R_2(t), S_2(t))$  using values for  $\mu_{R2}, \mu_{S2}, \psi_{S2}, V^{(2)}$ , and historical values for  $(R_2(t), S_2(t))$  extracted from the relevant elements of  $\theta_j$ .
- Generate a random sample path for future values of  $(R_3(c), S_3(c))$  using values for  $\mu_{R3}, \delta_{R3}, \mu_{S3}, \phi_{R31}, \phi_{R32}, \phi_{S31}, \phi_{S32}, V^{(3)}$ , and historical values for  $(R_3(c), S_3(c))$  extracted from the relevant elements of  $\theta_j$ .
- Extract values for  $\beta_x^{(11)}$  and  $\beta_x^{(12)}$  from the relevant elements of  $\theta_j$ .
- Use the simulated and extracted values of the age, period and cohort effects to construct the future arrays of death rates  $m_1(t, x)$  and  $m_2(t, x)$  for scenario  $j$ .

ANDREW J.G. CAIRNS  
*Maxwell Institute for Mathematical Sciences, and  
 Department of Actuarial Mathematics and Statistics,  
 Heriot-Watt University,  
 Edinburgh, EH14 4AS, UK.  
 E-mail: A.Cairns@ma.hw.ac.uk*

DAVID BLAKE AND KEVIN DOWD  
*Pensions Institute,  
 Cass Business School,  
 City University,  
 106 Bunhill Row,  
 London, EC1Y 8TZ, UK.*

GUY D. COUGHLAN AND MARWA KHALAF-ALLAH  
*Pension Advisory Group,  
 JPMorgan Chase Bank,  
 125 London Wall,  
 London, EC2Y 5AJ, UK.*