DISCUSSION PAPER PI-1001

Bayesian Stochastic Mortality Modelling for Two Populations

Andrew J.G. Cairns, David Blake, Kevin Dowd,
Guy D. Coughlan, and Marwa Khalaf-Allah

January 2010

DISCLAIMER

Bayesian Stochastic Mortality Modelling for Two Populations

Andrew J.G. Cairns[a], David Blake[b], Kevin Dowd[b],
Guy D. Coughlan[c], and Marwa Khalaf-Allah[c]

1 January, 2010

Abstract

The paper introduces a new framework for modelling the joint development over time of mortality rates in a pair of related populations by combining a number of recent and novel developments in stochastic mortality modelling. First, we develop an underlying stochastic 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 in line with the principles of biological reasonableness. 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. This procedure employs Markov chain Monte Carlo (MCMC) techniques, permitting 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 attributable to small populations.

Mortality rates arising from this framework provide consistent forecasts for the two populations. Further, estimated correlations based on the simulated mortality improvement factors for two populations are consistent with historical data.

The framework is illustrated using two-population extensions of the Age-Period-Cohort and Lee-Carter models on the following populations: England & Wales national and CMI assured lives males and females, and US and California males. The approach is designed for large populations coupled with a small sub-population, but is easily adaptable to other combinations.

A key application of the modelling framework would be to allow longevity risk hedgers to model the basis risk that exists between mortality rates in two populations in the case where the hedger wishes to hedge the risk in one population using a hdging instrument based on the second population.

Keywords: Stochastic mortality, two populations, small sub-populations, mortality spreads, age effect, period effect, cohort effect, basis risk, Markov chain Monte Carlo, parameter uncertainty.

[a]Maxwell Institute for Mathematical Sciences, and Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, EH14 4AS, UK. E: A.Cairns@ma.hw.ac.uk;

[b]Pensions Institute, Cass Business School, City University, 106 Bunhill Row, London, EC1Y 8TZ, UK.

[c]Pension ALM Group, JPMorgan Chase Bank, 125 London Wall, London, EC2Y 5AJ, UK.

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, Currie et al. (2004), Renshaw and Haberman (2003, 2006), Cairns et al. (2006a,b, 2008b, 2009)). Most work has focused on stochastic mortality models for single populations. 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 projections of two related populations. As an example, consider Figure 1. This shows forecasts of mortality rates for England & Wales males and females aged 65, using data from 1961 to 2005. Forecasts are based on the LC model fitted to each population separately. We see that after 40 or so years, the central projection for the female mortality rates crosses over that of males, a consequence of the lower improvement rates over the last 45 years for females. To some, this cross-over might seem unreasonable: it seems more plausible that the relationship between male and female mortality rates remains approximately stable, with rates for females remaining below male rates at most ages. This issue can be resolved through the use of some subjective interventions such as calibrating the models over different ranges of years, or by directly making manual adjustments to the future mortality improvement rates using in making forecasts. This paper seeks to address the problem of inconsistent projections in related populations using more objective means.

Second, we need to consider the correlation and long-term cointegration between mortality improvements in different populations in order to have a thorough assessment of longevity risk, building on the non-parametric, historical analysis of Coughlan et al. (2009b). To illustrate, a typical life assurer covers assured lives and annuitants, two populations with potentially different patterns of mortality improvements. As another example, a large pension plan might wish to model blue- and white-collar workers separately, before aggregating longevity risk to the plan level. Other examples include smokers and non-smokers, and multinational assurers and pension plans.

Third, we might be concerned with mortality improvements in a small population. Data availability might be limited for such a population, for example, due to a small numbers of deaths, a limited number of calendar years of data or age range, or, simply, poor quality data. Any model fitted solely to this data will have highly unreliable parameter estimates, whereas a model fitted jointly to the small population and a much larger, linked population (for example, national population mortality) allows the small-population mortality forecasts to be linked to and be consistent with those of the larger population.
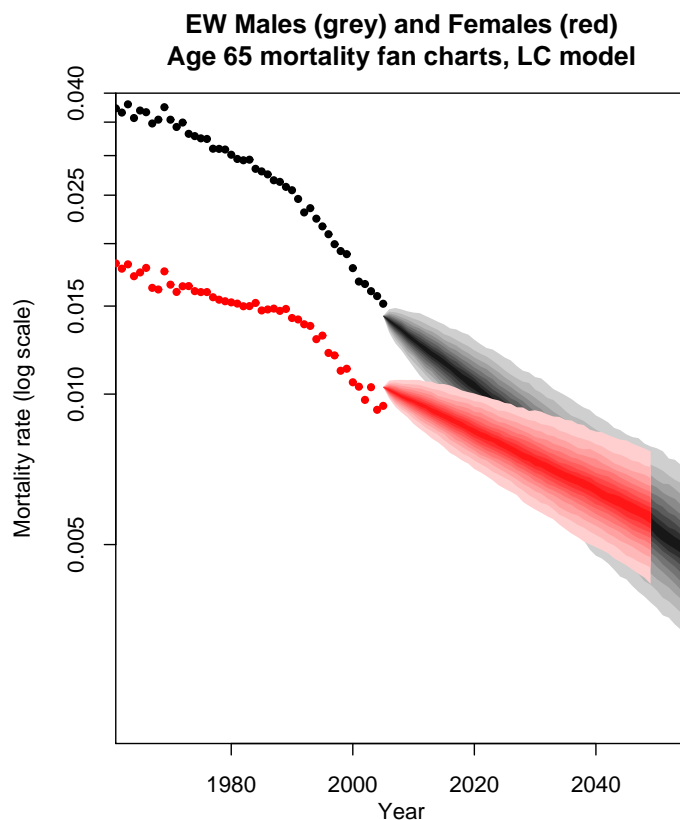
Figure 1: Single population projections for England & Wales males (grey) and females (red) mortality at age 65 using the Lee-Carter model. Both projections use data from 1961 to 2005 and ages 60 to 89. Dots: historical crude mortality rates. Fans: prediction intervals showing 5%, 10%, ..., 90%, 95% quantiles for future mortality rates.

Fourth, entities seeking to hedge their exposure to longevity risk using longevity index-based hedging instruments need to be able to determine hedge ratios and to assess the degree of basis risk between their own population and the reference population underlying the longevity index. A good model will help identify situations where basis risk is considered too high, resulting in a need for alternative index-based instruments with lower basis risk or bespoke over-the-counter agreements that come much closer to the full indemnification of longevity risk, as in traditional reinsurance contracts.

From the hedger's point of view, basis risk[1] can take a number of forms. The core source of basis risk comes from the fact that the hedging instrument's reference

---

[1]Here we interpret basis risk as being the difference between the value of the liability being hedged and the value of the hedge portfolio.

population might be imperfectly correlated with the hedger's own population (see, for example, Dahl et al (2008, 2009), Jarner and Kryger (2009) and Li and Hardy (2009)). However, there are additional, secondary sources of basis risk to bear in mind:

- Actual numbers of deaths differ to some extent from expected numbers (Dahl et al., 2008, 2009). Death rates are, therefore, random and will be different from the true underlying death rate due to sampling variation.[2] Where a mortality-linked derivative is related to the underlying death rate, this type of basis risk will be of greater significance to small populations than large.

- Life assurance mortality data sometimes contains information about amounts assured, as well as deaths and exposures. Currie et al. (2004) analysed the evolution of death rates using both simple counts of lives and weighting by amounts, and found differences in the pattern of improvements. The use of a derivative based solely on lives-only data would therefore lead to basis risk relative to liabilities based on both lives and amounts. To get round this problem, Currie et al. proposed a simple method to model lives- and amounts-based data jointly by introducing a time-invariant, age-specific spread.

- Basis risk is increased by factors such as inadequate liquidity, transaction costs, market incompleteness, data frequency, and time lags in data availability.

Basis risk is discussed further and analysed by, amongst others, Blake et al. (2006), Cairns et al. (2006a, 2008a), Coughlan et al. (2007a,b, 2009b), Dahl et al. (2008, 2009), Jarner and Kryger (2009), Loeys et al. (2007) and Li and Hardy (2009). A key element in the consideration of basis risk is hedge effectiveness (Coughlan et al., 2004, 2007a, 2009a). When potential hedgers are considering the relative merits of potential longevity index-based hedging instruments versus, say, over-the-counter agreements, the way in which an index-based instrument is used is just as important as its basic design. A poorly thought-out hedging strategy might have low hedge effectiveness causing a rational hedger to favour an over-the-counter solution, whereas a well-designed, practical-to-implement and robust strategy might tip the balance in favour of the index-based instrument.

One of the key aims of this paper is to investigate the first form of basis risk, namely imperfect correlation between underlying mortality rates in different populations. The framework developed in this paper allows hedgers to conduct a comprehensive analysis of basis risk between two populations' future mortality rates, covering both uncertainty in the parameter estimates of the state variable dynamic processes and uncertainty in the future projections of the state variables.

---

[2]Often this will be referred to as sampling risk, but its inclusion here is consistent with our broad definition of basis 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, suggesting, perhaps, some form of linear growth of the PLE with periodic overtaking by a new healthier population. This is a plausible hypothesis, with the healthiest representing 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. For two similar, but closely linked populations, Currie et al. (2004) propose a *piggyback* model that treats the second population (e.g. one with poorer quality data) as piggybacking off the data for the first population.

Full joint modelling of two or more populations has been considered by Li and Lee (2005) and Jarner and Kryger (2009). The former 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. The latter focus on modelling a small population's mortality (Denmark) alongside a much larger supranational (Europe-wide) population. The large population is assumed to follow some form of deterministic trend over time. The relationship between the large and the small population log mortality rates is described by a multifactor, mean-reverting spread (a VAR(1) process). The mean reversion level is set at zero, reflecting the assumed similar nature of the two populations (that is, there is no *a priori* reason why Danish mortality should be higher or lower than supranational mortality in the long run).[3]

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 (building, for example, on

---

[3]It can be argued that mean reversion to zero is not necessarily appropriate, even at the national level. For example, Richards (2008) shows that regional differences in mortality within the UK are primarily due to socio-economic and lifestyle differences (for example, differences in the percentages of professional and unskilled workers). Where the social mixtures of different countries is expected to persist, we would therefore expect differences to persist in aggregate mortality: that is, non-zero mean reversion.

Brouhns et al., 2002), 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 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 two reasons. First, it helps us to take account of parameter uncertainty in a more natural way. Second, the careful specification of a limited number of prior distributions helps us to avoid unreasonable model parametrisations. We implement the Bayesian approach using Markov chain Monte Carlo (MCMC).

Countering these advantages, it is much easier under the two-stage approach to try out different time series models for random period and cohort effects. This can be done in the single-stage approach but it would require considerably more programming effort. In this paper, the choice of time series models has been heavily influenced by what we have learned from our extensive modelling work on single populations using the two-stage approach.

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). 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. 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 Models considered in this paper

Amongst the single-population models analysed previously by Cairns et al. (2009), we have chosen to focus on two of the simpler models: the APC model (labelled M3 by Cairns et al. (2009); see also Osmond and Gardner (1982), Osmond (1985), Bray (2002) and Jacobsen et al. (2002)) and the LC (1992) model (labelled M1). The relative simplicity of these two models allows us to focus on the key contribution of this paper, namely combining Bayesian methods, smoothing, and coupling of two populations. The series of papers by Cairns et al. (2008b, 2009) and Dowd et al. (2008 a, b) found that other models might be preferred to LC or APC, 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 datasets that we will use to investigate two-population mortality modelling. We focus mainly on England & Wales (EW) versus Continuous Mortality Investigation (CMI) assured lives data (both males and females), before turning to US versus California males mortality data. In Section 3, we outline the core hypothesis concerning non-divergence of mortality 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.[4] In Sections 6 to 12, we fit the model to the various datasets, and discuss aspects of the initial MCMC output before going on to analyse forecasting results incorporating full parameter uncertainty. Section 13 concludes.

A glossary is provided at the end which lists the main notation used throughout the paper.

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 $x_0, \ldots, x_1 = x_0 + n_a - 1$, and the range of years covered will be $y_0, \ldots, y_1 = y_0 + n_y - 1$. The corresponding cohort years of birth are $c_0, \ldots, c_1$, 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 exposures, $E_i(t, x)$, for each population, i, year, t, and age, x. From this, we derive the crude 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.[5]

---

[4]Section 5 is more technical in nature, and some readers might choose to skip forward to Section 6.

[5]Missing data means that the exposures and deaths are recorded as 0's. Cells that have non-zero exposures, but zero deaths are quite acceptable, and should not be recorded as missing 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[6]).[7]

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[8] allows us to improve, refine or make more accurate forecasts of mortality. This line of reasoning is discussed further in Section 10.

2.2   Specific datasets

We consider three pairs of datasets:

- EW and CMI assured lives males from 1961 to 2005, ages 60 to 89.

- EW and CMI assured lives females from 1983 to 2003, ages 60 to 89.

- USA and California males from 1980 to 2003, ages 60 to 84 (California data are only available with sufficient detail from 1980).

We also consider augmentation of the EW/CMI datasets listed above, where additional years of data exist for the EW population.

The first pair has 186 million life-years for EW males and 21 million for CMI males. However, this masks the fact that the CMI exposures have fallen substantially over the last 15 years reflecting a decline in demand in the UK for traditional life assurance products. The second pair has 124 million life-years for EW females and only 3.0 million for CMI females. In contrast to the CMI males' data, the females exposures were rising up to 2000 with a small decrease since then. The third pair of populations has a total exposure of 404 million life-years for the US and 41 million for California.

In each case, the second population is a sub-population of the larger and about 10% in size or less. 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

---

[6]For example, at the time of writing, CMI data are currently available for males up to 2005, whereas EW population data are available up to 2007. The MCMC approach allows us to include years 2006 and 2007, but treat CMI data for those years as missing. The inclusion of the 2006/7 EW data allows us to improve the forecasts for both populations.

[7]The current version of the model does not allow for complete missing ages in one or other of the populations.

[8]For example, the EW data for 2006 and 2007, against missing CMI data for the same two years.

the typical EW male or female. Similarly, the Californian population is, in general, qualitatively different from the rest of the US: it is wealthier, healthier, fitter, enjoys better weather and has a different pattern of immigration by age than much of the rest of the US.

We would argue that if our methodology works on these population pairs, this will give us good grounds for thinking it will work well on other datasets.

2.3 Age definitions

Most of the datasets with which we 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, as we treat the forecast and historical death rates in a consistent way within each population and between populations.

When we receive the output forecasts, however, we must remember to use the simulated death rates in the correct way. For example, for a male aged *exactly* 65 at the start of 2015, the probability of survival to the end of 2015 is:

- EW: $\exp[-m_{EW}(2015, 65)]$

- CMI: $\exp\left[-(m_{CMI}(2015, 64) + m_{CMI}(2015, 65))/2\right]$

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 $q_i(t, x)$ be the underlying mortality 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. At the very least, we would expect the mortality rates in two related populations not to diverge over time (see, for example, Li and Lee

(2005), and Jarner and Kryger (2009)). We translate this qualitative expectation into the following mathematical hypothesis: *for each age x, the ratio $q_1(t, x)/q_2(t, x)$ will not diverge as $t \to \infty$*. These qualitative and quantitative criteria fall under the heading of biological reasonableness in the sense discussed by Cairns et al. (2006a, 2008a, b).

The above hypothesis allows the CMI mortality rates to remain 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.

Model	formula
M1	$\log m(t, x) = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)}$
M2	$\log m(t, x) = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \beta_x^{(3)} \gamma_{t-x}^{(3)}$
M3	$\log m(t, x) = \beta_x^{(1)} + n_a^{-1} \kappa_t^{(2)} + n_a^{-1} \gamma_{t-x}^{(3)}$
M5	$\operatorname{logit} q(t, x) = \kappa_t^{(1)} + \kappa_t^{(2)} [x - \bar{x}]$
M6	$\operatorname{logit} q(t, x) = \kappa_t^{(1)} + \kappa_t^{(2)} [x - \bar{x}] + \gamma_{t-x}^{(3)}$
M7	$\operatorname{logit} q(t, x) = \kappa_t^{(1)} + \kappa_t^{(2)} [x - \bar{x}] + \kappa_t^{(3)} \left[(x - \bar{x})^2 - \hat{\sigma}_x^2 \right] + \gamma_{t-x}^{(4)}$

Table 1: Formulae for the six mortality models considered by Cairns et al. (2008b): The functions $\beta_x^{(k)}$, $\kappa_t^{(k)}$, and $\gamma_{t-x}^{(k)}$ are age, period and cohort effects, respectively. $\bar{x}$ is the mean age over the range of ages being used in the analysis. $\hat{\sigma}_x^2$ is the mean value of $[x - \bar{x}]^2$. n_a is the number of ages.

This hypothesis has different consequences depending on the stochastic mortality model being considered. Consider, for example, the six models considered by Cairns et al. (2008b) and listed in Table 1. We generalise the notation in this table to include reference to the population: thus, the $\beta_x^{(ki)}$ are population i age effects, the $\kappa_t^{(ki)}$ are population i period effects, and the $\gamma_{t-x}^{(ki)}$ are population i cohort effects. For non-divergence in a stochastic model of the age x mortality rates, we require the following for each model:

- M1 (Lee-Carter, 1992): $\beta_x^{(21)} = \beta_x^{(22)}$ for each x, and $\kappa_t^{(21)} - \kappa_t^{(22)}$ must be mean reverting. (See, for example, the discussion in Li and Lee (2005).)

- M2 (Renshaw-Haberman, 2006): As M1 plus $\beta_x^{(31)} = \beta_x^{(32)}$ for each x, and

$\gamma_c^{(31)} - \gamma_c^{(32)}$ must be mean reverting in some way as we simulate forward.[9]

- M3 (Age-Period-Cohort model): $\kappa_t^{(21)} - \kappa_t^{(22)}$ and $\gamma_c^{(31)} - \gamma_c^{(32)}$ must both be mean-reverting.

- M5 (Cairns et al., 2006b, 2009, CBD1): $\kappa_t^{(11)} - \kappa_t^{(12)}$ and $\kappa_t^{(21)} - \kappa_t^{(22)}$ must be mean reverting.

- M6 (Cairns et al., 2009, CBD2): As M5 plus $\gamma_c^{(31)} - \gamma_c^{(32)}$ must be mean-reverting.

- M7 (Cairns et al., 2009, CBD3): As M5 plus $\kappa_t^{(31)} - \kappa_t^{(32)}$ and $\gamma_c^{(41)} - \gamma_c^{(42)}$ must both be mean-reverting.

Aside from longer-term mean reversion, we might also expect to see some correlation between the year-on-year changes in both the period and the cohort effects.[10] For example, if population 2 is a subset of population 1, we might expect that environmental fluctuations (e.g., flu epidemics, extreme weather, etc.) will affect both populations in similar ways. The accumulation of environmental factors during a cohort's lifetime (e.g., exposure in early life to certain viruses, changes in diet, government health policy) might not result in high short-term correlations between the cohort effects of the two populations, but we might expect to see higher correlations over longer horizons.[11]

4 The Two-Population Age-Period-Cohort Model

We will focus on the APC (M3) Model, in Table 1. For the two populations, we have underlying death rates modelled as follows

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

for given age, period and cohort effects, $\beta_x^{(1i)}$, $\kappa_t^{(2i)}$ and $\gamma_{t-x}^{(3i)}$.

---

[9]For M1 and M2 only, the non-divergence requirements mean that we must fit the model simultaneously to both populations to estimate the common ages effects $\beta_x^{(2i)}$ and $\beta_x^{(3i)}$. For all other models, we can choose whether or not to fit the model to each population separately as a first step.

[10]This hypothesis builds on the empirical results of Coughlan et al. (2009b). In their analysis, mortality rates in each calendar year were first subjected to non-parametric smoothing. They then observed that annual changes in historical mortality rates at specific ages in two populations were positively correlated at most, although not all ages.

[11]One possibility is that the cohort effect might actually be a random process that is relatively smooth in the short term.
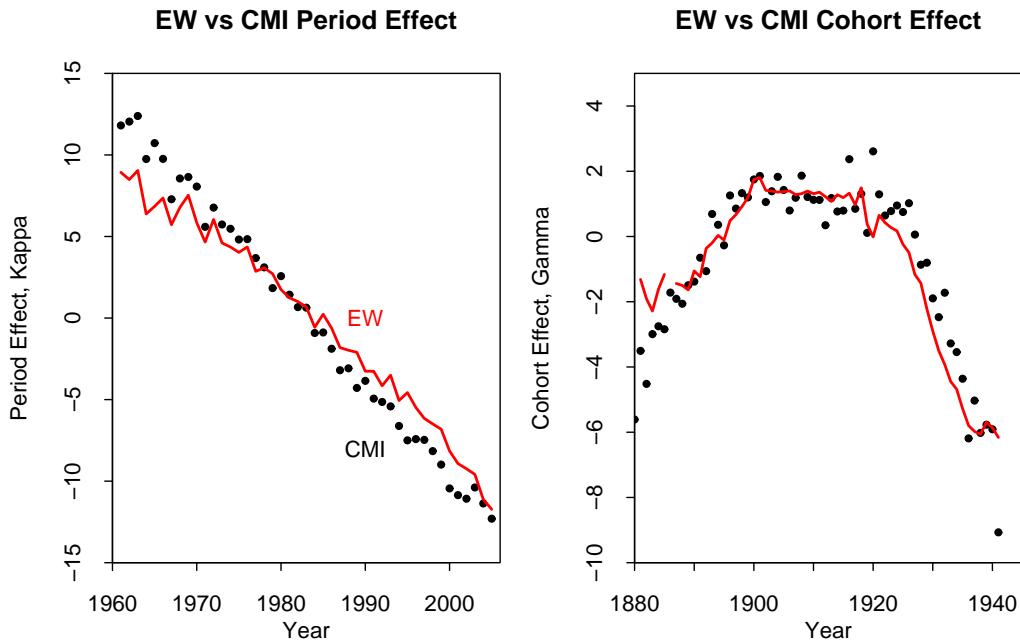
Figure 2: Single population estimates for period (left) and cohort (right) effects for England & Wales males (dots) and CMI males (lines) using the Age-Period-Cohort model (M3) discussed by Cairns et al. (2009). Data from 1961 to 2005 and ages 60 to 89.

In previous work on single populations (Cairns et al. 2008b), we modelled each period effect, $\kappa_t^{(2i)}$, as a random walk and each cohort effect, $\gamma_{t-x}^{(3i)}$, as an ARIMA(1,1,0) model. However, both features might result in divergence of the death rates in the two populations over time. To prevent this, we require that the following are both mean reverting: $\kappa_t^{(21)} - \kappa_t^{(22)}$ and $\gamma_{t-x}^{(31)} - \gamma_{t-x}^{(32)}$.

4.1 Stylised facts

A number of basic analyses looking at various populations were conducted before we proceeded to develop the full MCMC approach. These involved analysing each population separately (as in Cairns et al., 2008b) and then considering the relationship between the two populations' period and cohort effects. These investigations resulted in the emergence of the following stylised facts:

- Year-on-year innovations in the two populations' period effects, $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$, are significantly positively correlated: see, for example, Figure 2, left hand plot for the case of EW and CMI males.

- Demonstrating the existence, statistically, of mean reversion in the spread between the period effects, $\kappa_t^{(21)} - \kappa_t^{(22)}$, is typically difficult due to the relatively modest number of years of data available.

- Correlations between year-on-year innovations in the two populations' cohort effects, $\gamma_c^{(31)}$ and $\gamma_c^{(32)}$, where $c = t - x$, are relatively small and not necessarily significant.

- For some pairs of populations, the medium- and longer-term shape of the two fitted cohort effects are relatively similar. See, for example, Figure 2, right hand plot for EW and CMI males: the similarity between the underlying shapes of the respective cohort effects is particularly striking.

- Poisson randomness in the death counts for small populations results in noticeably more 'noisy' estimates compared with large populations for the age and cohort effects: that is, the fitted age and cohort effects are much less smooth than typical fitted age and cohort effects for large populations.

These final observations lead us to conclude that, unlike the period effects, the cohort effects are likely to be relatively smooth functions of cohort year of birth. Also, the Poisson noise that affects our estimates of the cohort effect results in the lack of correlation that we observe in year-on-year innovations between the two cohort effects. However, our stylised facts also point to stronger correlation in the medium and longer terms between the two smooth cohort effects, indicative of local smoothness around a random trend.

Our interpretation of the two populations period effects is different. As noted above, we observe significant positive correlations between innovations in the stage-1 fitted period effects. This can only happen if the true (rather than the fitted) period effects themselves also incorporate significant randomness. In contrast, if the true period effects were smooth, then Poisson randomness in the death count would dominate, estimates for small population period effects would look much more noisy, and innovations in the two populations' fitted period effects would not be significantly correlated.

In addition to the above, Coughlan et al. (2009b) carried out an model-independent analysis of the correlation between mortality improvement factors in two populations. They found that correlation rises with the time horizon (in a way that is consistent with our core hypothesis) and, as a stylised fact, we would expect to see a similar pattern in model-based forecasts of improvement factors.

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 point of view. These criteria are subjective

in nature and are not typically supported by stylised facts:

- $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$ should have similar conditional 1-year-ahead variances.

- $\gamma_c^{(31)}$ and $\gamma_c^{(32)}$ should have similar conditional 1-year-ahead variances.

- $\gamma_c^{(31)}$ and $\gamma_c^{(32)}$ should have similar unconditional variances.[12]

- In the long-run, $\gamma_c^{(31)}$ and $\gamma_c^{(32)}$ should be positively correlated.

In our initial analysis of the EW and CMI males populations (Section 6), we do not incorporate prior distributions for these criteria. However, analysis of this initial set of results reveals that, at least for this pair of populations, the criteria are not satisfied for the two populations' cohort effects. We use this, then, as motivation for incorporating suitable priors that are consistent with the above criteria in our second and subsequent analyses (Sections 7 onwards).[13]

4.3 Population 1 dominant

As a motivation for the approach taken in this paper, we have the following typical scenario. A pension fund might be considering using a mortality-linked security or derivative to help hedge its longevity-linked liabilities. The market in such contracts is currently developing, and one strand involves index contracts that are linked to national population mortality rates, leading to basis risk between the pension fund's own liabilities and the index being used to help hedge those liabilities. Our aim is to develop a methodology that will help us to analyse this risk.

In this paper, therefore, we focus on situations where one population (say population 1) is much larger than the other (population 2). We therefore 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

$$
\begin{aligned}
R_2(t) &= \kappa_t^{(21)} \\
S_2(t) &= \kappa_t^{(21)} - \kappa_t^{(22)} \\
R_3(c) &= \gamma_c^{(31)} \\
S_3(c) &= \gamma_c^{(31)} - \gamma_c^{(32)}.
\end{aligned}
$$

---

[12]$\gamma_c^{(31)}$ and $\gamma_c^{(32)}$ are both mean-reverting processes, and so, in a long run, have a limiting maximum or unconditional variance. Biological reasonableness suggests that these limiting variances should be similar. In contrast, $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$ are random walks with infinite unconditional variances; however, biological reasonableness requires similar conditional 1-year-ahead variances.

[13]The priors apply to the true but unobservable processes. Actual variances will be higher for small populations because of Poisson noise in the death counts.

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.

- The 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.

- We allow the innovations for $R_3(c)$ and $S_3(c)$ to be correlated, although our discussion of the stylised facts suggests that this correlation might be small. The bivariate innovations for $(R_3(c), S_3(c))$ will be modelled as i.i.d. bivariate normal from one year to the next.

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., 2008b). For the central cohort effect, $R_3(c)$, we previously used an ARIMA(1,1,0) model (Cairns et al., 2008b). 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.[14] 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.

In previous papers (Cairns et al. 2008b, 2009), we 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

- $R_2(1) = 0$,

---

[14]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, with relative smoothness in the short term and occasional shifts in the trend. This smoothness was considered a desirable property in our section on stylised facts.

- $S_2(t)$ 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.[15] 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 mortality rates.

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

$$
\begin{aligned}
R_2(t+1) &= R_2(t) + \mu_{R2} + C_{211}Z_{21}(t+1) \\
S_2(t+1) &= \mu_{S2} + \psi_{S2}(S_2(t) - \mu_{S2}) + C_{221}Z_{21}(t+1) + C_{222}Z_{22}(t+1) \\
\tilde{R}_3(c) &= R_3(c) - \mu_{R3} - \delta_{R3}(c - \bar{c}) \\
\tilde{S}_3(c) &= S_3(c) - \mu_{S3} \\
\tilde{R}_3(c+1) &= (\phi_{R31} + \phi_{R32})\tilde{R}_3(c) - \phi_{R31}\phi_{R32}\tilde{R}_3(c-1) + C_{311}Z_{31}(c+1) \\
\tilde{S}_3(c+1) &= (\phi_{S31} + \phi_{S32})\tilde{S}_3(c) - \phi_{S31}\phi_{S32}\tilde{S}_3(c-1) + C_{321}Z_{31}(c+1) \\
&\qquad\qquad + C_{322}Z_{32}(c+1).
\end{aligned}
$$

In the third equation above, $\bar{c}$ 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. The identifiability constraints used in our specific MCMC algorithm require that $\mu_{S2} = \mu_{R3} = \mu_{S3} = \delta_{R3} = 0$. Stationarity of $\tilde{R}_3(c)$ and $\tilde{S}_3(c)$ requires each of ϕ_{R31}, ϕ_{R32}, ϕ_{S31} and ϕ_{S32} to lie between -1 and $+1$.

4.4 Population 2 carries non-zero weight

In the previous subsection, population 2 was subsidiary to population 1 and carried no weight in dictating the main trends $R_2(t)$ and $R_3(c)$. Where the two populations are of more equal size (for example, males and females, or two national populations), we might define $R_2(t)$ and $R_3(c)$ to be weighted averages of the two-populations' period and cohort effects. The methodology that we develop in this paper can be readily generalised to incorporate this possibility. We have chosen not to investigate this further in this paper in order to keep the length of the paper in check and to keep the focus on the basis risk between a large national population and a small population seeking to hedge its longevity risk.

---

[15]See section 5.2.

5 Estimation Method

In previous work (see, for example, Cairns et al. (2008b, 2009), Lee-Carter (1992), Brouhns et al. (2002), Booth et al. (2005)), most researchers have employed a two-stage, non-Bayesian approach to modelling:

- Stage 1: estimate age, period and cohort effects without reference to their underlying dynamics.

- Stage 2: fit a suitable stochastic process to the period and cohort effects.

This is a satisfactory approach, provided the population size is large. More recently, some authors (e.g., Bray (2002), Czado et al. (2005), and Reichmuth and Sarferaz (2009)) have sought to combine these two stages in a Bayesian setting when considering single populations.

For smaller populations, sample variation affects death counts, which, in turn, can have a non-negligible impact on crude death rates and estimates of age, period and cohort effects. Consider the LC model (M1), for example, and suppose the period effect is a random walk. With a small population, the stage-1 fitted period effect will turn out to be more like a random walk with additional noise. The variance of this noise will be proportional to the population size, and this might feed through in stage 2 to an estimate for the random-walk variance that is on the high side.

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 function 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 function 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. For example, as noted above, if the stage-1 fitted period effects look like a random walk plus noise,[16] then a combined fitting procedure will rein in at least some of the noise to produce a series that looks closer to a true random walk.

A second motivation concerns the fitting of the cohort effect. Our earlier approach (Cairns et al., 2009) excluded cohorts with four or fewer observations to avoid overfitting. The combined fitting procedure allows us to include cohorts with only one observation, 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

---

[16]If there is sufficient noise on top of a random walk, then we will observe negative lag-1 auto-correlations in the fitted random-walk innovations.

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 provide a framework within which we can tackle the estimation problem in a single stage:[17]

- 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.[18] The MCMC output will allow us to derive a posterior distribution for the relevant parameters and for the underlying mortality rate for the missing cells.

A popular approach to this uses the Metropolis-Hastings (MH) algorithm:

- Vector $\theta(i)$ = current set of parameter and latent variable values

- D = observed data.

- $p(\theta|D)$ = posterior density for θ. The posterior distribution is sufficiently complex that direct simulation from $p(\theta|D)$ is impossible.

- Step $i + 1$. In a series of substeps, $j = 1, 2, ...$

  - Substep j updates a single element or a block of the vector θ.
  - $\tilde{\theta}$ = latest θ including accepted substep updates.
  - Generate a candidate $\hat{\theta}$ from a distribution with density $f(\hat{\theta}|\tilde{\theta})$
  - Accept $\hat{\theta}$ with probability

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

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

The 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 $i = 1, \ldots, N$ will be a good approximation to the true posterior.

---

[17]For a general introduction to MCMC, see, for example, Gilks et al. (1996).

[18]See, for example, the comments about the 1886 cohort in the EW males data in Section 2.1.

5.2   Likelihood, prior and posterior

The complete parameter vector is denoted by θ, and consists of:

- Subvector $\theta^{(2)}$ containing the process parameters for the period effects

  - μ_{R2} = random-walk drift for $R_2(t)$
  - ψ_{S2} = AR(1) parameter for $S_2(t)$
  - $V^{(2)}$ = variance-covariance matrix for the $(R_2(t), S_2(t))$ innovations.

- Subvector $\theta^{(3)}$ containing the process parameters for the cohort effects

  - ϕ_{R31}, ϕ_{R32} parameters for the $R_3(c)$ AR(2) process
  - ϕ_{S31}, ϕ_{S32} parameters for the $S_3(c)$ AR(2) process
  - $V^{(3)}$ = variance-covariance matrix for the $(R_3(t), S_3(t))$ innovations.

- Latent age, period and cohort effects and their spreads

  - Vectors $\beta_x^{(11)}$, $\beta_x^{(12)}$, $R_2(t)$, $S_2(t)$, $R_3(c)$ and $S_3(c)$.

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 $\beta_x^{(1i)}$, $\kappa_t^{(2i)}$, $\gamma_c^{(3i)}$ vectors.

- $l_{21}(\theta)$ = 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 $t = 2, \ldots, n_y$.

- $l_{31}(\theta)$ = 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 $c = 3, \ldots, n_c$.

Specifically,

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

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

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

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

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

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

where $X = (R_3(2), S_3(2), R_3(1), S_3(1))'$ and Ω is the solution to the equation[19]

$$\Omega = A\Omega A' + \tilde{V}^{(3)} \tag{2}$$

$$A = \begin{pmatrix} \alpha_{R31} & 0 & \alpha_{R32} & 0 \\ 0 & \alpha_{S31} & 0 & \alpha_{S32} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\alpha_{R31} = \phi_{R31} + \phi_{R32}$$
$$\alpha_{R32} = -\phi_{R31}\phi_{R32}$$
$$\alpha_{S31} = \phi_{S31} + \phi_{S32}$$
$$\alpha_{S32} = -\phi_{S31}\phi_{S32}$$

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

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

$$\text{where } Y_3(c) = (Y_{31}(c), Y_{32}(c))',$$
$$Y_{31}(c) = R_3(c) - \alpha_{R31}R_3(t-1) - \alpha_{R32}R_3(c-2),$$
$$Y_{32}(c) = S_3(c) - \alpha_{S31}S_3(t-1) - \alpha_{S32}S_3(c-2).$$

---

[19]While it is not possible to solve (2) below using standard matrix algebra, the equation is linear in all elements of Ω and can be solved by writing Ω as a 16×1 vector.

5.3 Directed acyclic graph

Alongside the formal specification of the model and the likelihoods, we can present the model in less formal way using a directed acyclic graph, Figure 3.[20]

5.4 The Gibbs sampler

The Gibbs sampler is a special case of the MH algorithm under which the candidate distribution is 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, α (see equation 1), 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.

Typically, the conditional posterior distribution of a given parameter or parameters depends upon a subset of the full set of parameters. For example, the normal posterior for μ_{R2} depends on the current set of $(R_2(1), \ldots, R_2(n_y), S_2(1), \ldots, S_2(n_y))$, and on the process parameters $V^{(2)}$, and ψ_{S2}. A standard MH step is still required for $\psi_{S2}, \phi_{R31}, \phi_{R32}, \phi_{S31}, \phi_{S32}$.

5.4.1 Pseudo-Gibbs sampler

Occasionally, the conditional posterior is not in a form that can be simulated easily (see Table 2). In this case, we can simulate from a simpler distribution that is still a good approximation to the true conditional posterior. The acceptance probability (equation 1) will now be different from 1, and so the full likelihood needs to be evauated. This is still worthwhile doing as the pseudo-Gibbs sampler still results in efficient mixing.

---

[20]The empty circle and absence of arrows in the upper left box reflect the fact that we have not specified a model for the age effects.
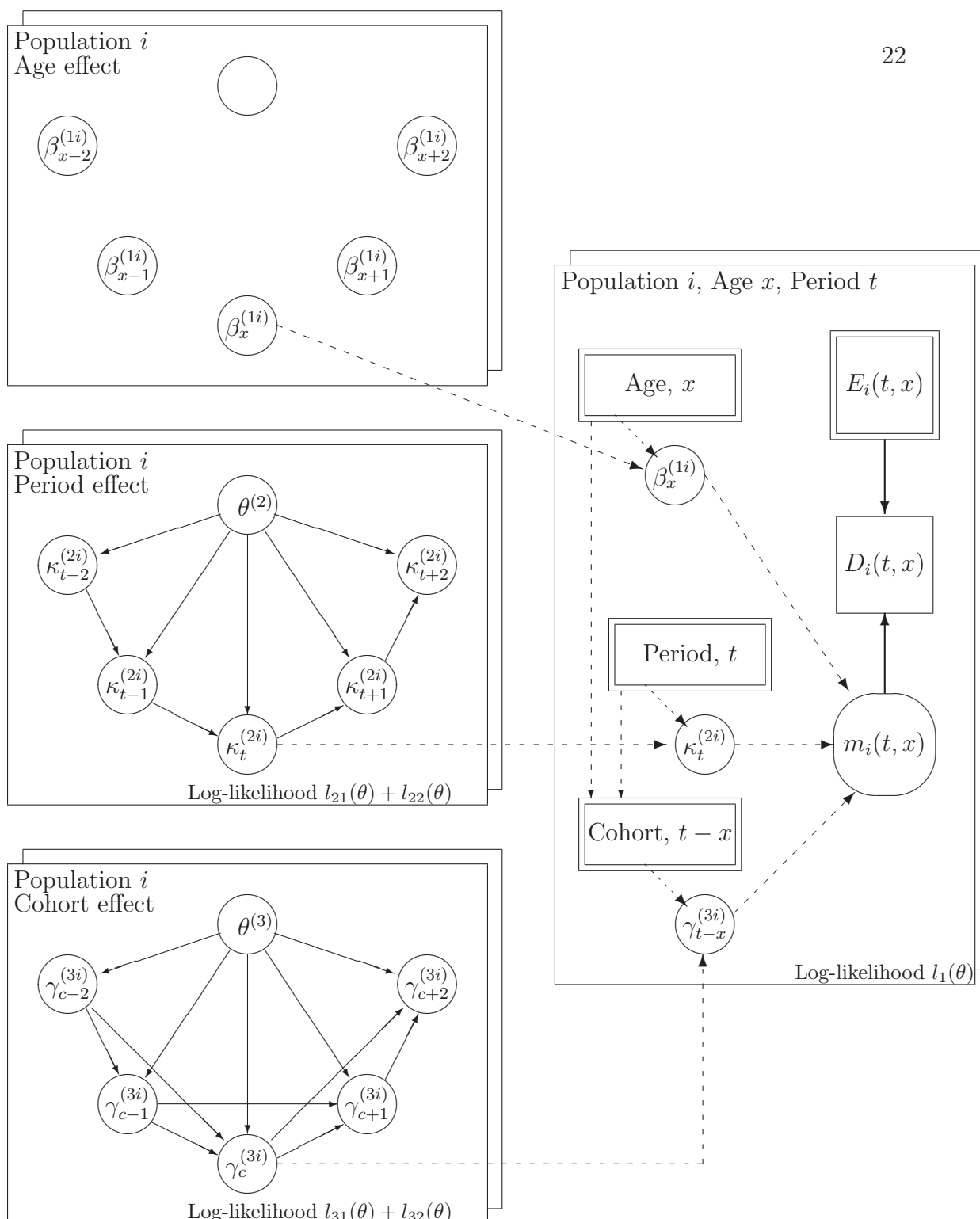
Figure 3: Directed acyclic graph for the two-population version of the Age-Period-Cohort model. Solid lines represent probabilistic dependencies. Dashed lines: functional (deterministic) relationships. Circles and ovals: unknown or unobserveable quantities. Rectangles: observed data. Double rectangles: quantities fixed by design. Stacked sheets: repetitive components.

5.5 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.

The prior distribution assumes:

- $V^{(2)}$ has a relatively uninformative inverse Wishart prior with density proportional to $|V|^{-1.5}$;

- $V^{(3)}$ 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 $\Psi = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.02 \end{pmatrix}$.

- $\mu_{R2} \sim N(-0.9, 0.9^2)$;

- $\log(1 - \phi_{R31}) \sim N(2, 0.5^2)$ distribution;

- $\log(1 - \phi_{R32}) \sim N(2, 0.5^2)$ distribution;

- $\log(1 - \phi_{S31}) \sim N(2, 0.5^2)$ distribution;

- $\log(1 - \phi_{S32}) \sim N(2, 0.5^2)$ distribution;

- logit $(\psi_{S2}) \sim \text{Gumbel}(2, 0.5)$ distribution.[21]

For $V^{(2)}$ and $V^{(3)}$, the choice of prior means that the conditional posterior distribution is approximately inverse Wishart (Table 2), allowing effective use of the pseudo Gibbs sampler. The use of a non-zero, inverse scale matrix Ψ was found to be necessary to avoid singularities in the log-posterior distribution at $V^{(3)} = 0$.[22] 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 μ_{R2} ensures a normal conditional posterior. Again the prior is weak, but not completely uninformative. Specifically, we wish to avoid μ_{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

---

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

[22]The fact that the cohort effects are not directly observable means that the optimiser can achieve an infinite maximum likelihood if $R_3(c)$ and $S_3(c)$ are linear and $V^{(3)} = 0$.

Set of parameters	Conditional posterior
$V^{(2)}$	(approx) inverse Wishart
$V^{(3)}$	(approx) inverse Wishart
μ_{R2}	normal
$\exp(\beta_{x_1}^{(11)}, \ldots, \beta_{x_{n_a}}^{(11)}, \beta_{x_1}^{(12)}, \ldots, \beta_{x_{n_a}}^{(12)})$	multivariate gamma
$(R_2(1), \ldots, R_2(n_y), S_2(1), \ldots, S_2(n_y))$	(approx) multivariate normal
$(R_3(1), \ldots, R_3(n_c), S_3(1), \ldots, S_3(n_c))$	(approx) multivariate normal

Table 2: Conditional posterior distributions for various components of θ.

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 μ_{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 μ_{R2}, which has a significantly different mean and a substantially smaller standard deviation.

As noted in Table 2, the exponentials of the age effects, $\beta_x^{(1i)}$, all have a gamma 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 posterior. 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 ψ_{S2}, ϕ_{R31}, ϕ_{R32}, ϕ_{S31} and ϕ_{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 $\log(1-\phi)$'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 ψ_{S2} was required to provide a stronger push away from $\psi_{S2} = 1$, than the log-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.

These particular parameterisations 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 females) and elsewhere (EW versus Scottish males) work with the same priors as listed.

6 Analysis 1: EW/CMI Males, Ages 60-89, Years 1961-2005

We begin our analyses by looking at EW versus CMI males using data from 1961 to 2005 and ages 60 to 89. The benchmark Analysis 1 uses the APC model plus the priors described in the previous section.

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 θ. 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 $\theta(i_B + 50), \theta(i_B + 100), \ldots$. 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.

The MCMC output includes values for the log-posterior. If we were just considering a simulation from the underlying distribution, then, by analogy with the likelihood ratio statistic, the maximum log-density minus the log-density for the simulated value of the random parameter set will produce a random variable that is approximately 0.5 times a chi-squared random variable. The variance that we observe for the log-posterior is consistent with this asymptotic theory (it is higher, but of the right order of magnitude).

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

- Figure 4 plots output from the MCMC algorithm for the time-series process parameters. Values have been stored for every 50th iteration out of a total of 50,000 iterations. The results are consistent with observations drawn independently from the posterior distribution, or in some cases with low autocorrelation (that is, there is no evidence in the plots that we are still within the burn-in period). For each sub-plot, the range of values on the y-axis has been chosen so that the 1000 points plotted fill the available space vertically. The plots reveal that many of the process parameters (for example, ψ_{S2}) have highly skewed posterior distributions.

- Figure 5 shows corresponding plots for selected latent age, period and cohort

effects. In most cases, there does not seem to be significant persistence in values in the simulated Markov chain. The only plots that show a noticeable pattern are for the $S_2(t)$. However, we can see from these plots that the run of 50,000 iterations is more than long enough for us to have sampled from the full posterior distribution.

- Figure 6 provides fan charts 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 $q_i(t, x)$ in each year.[23] In contrast with 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 $q_i(t, x)$. Importantly, we can see that the credibility intervals for the CMI data (grey fans) up to 2005 are significantly wider than for the EW data (red fans), reflecting the smaller size of the dataset.

  For age 85, the EW fan widens out 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 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 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 7, 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 (green 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 are significantly wider.

  In the right-hand plots, the underlying red and green fans are the same as on the left. These now have superimposed on them grey fans produced using the single-population models. For the EW data, the grey fan is reasonably smooth and the results are consistent with the EW forecasts. However, the

---

[23]The 90% credibility interval is made up of the 5% and 95% quantiles of the marginal posterior distribution for each $q_i(t, x)$.

red fan 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 (grey and green) have fans of similar width.

The CMI plot (bottom right) differs in three important ways. First, the single-population grey fan is much less smooth. For age 75, the first 15 years of projections include values for the cohort effect that have been derived 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. The red and green fans, by contrast (bottom left), are smoother and much more plausible. Second, for the CMI data there is a greater difference between the central trends in the red and grey fans. 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. Third, consider the width of the parameter certain fans (green and grey). For EW, these were about the same width, while, for the CMI data, the green fan is narrower. In Figure 2, 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 2 which results in narrower fans and hence more confident predictions.

- In Figure 8, 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:

$$\sum_c R_3(c) = 0$$

$$\sum_c R_3(c)(c - \bar{c}) = 0$$

$$\sum_c S_3(c) = 0$$

$$\sum_t R_2(t) = 0$$

$$\sum_t S_2(t) = 0.$$

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 μ_{R2}, μ_{R3} and δ_{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 and 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 grey fan on top provides credibility intervals for $R_3(c)$ (a repeat from the bottom left plot). The wide blue fan in the background provides credibility intervals for the trend, $\mu_{R3} + \delta_{R3}(c - \bar{c})$. Clearly, 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[24] 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.

---

[24]The fans in Figure 6, for example, would, otherwise, widen out rapidly.

- In Figure 9, we focus on the correlation between future mortality improvement factors for the two populations 10 years ahead in 2015. We look at 1000 simulations of the 10-year-ahead improvement factors (IF) for ages 65, 75 and 85 plotted on a logarithmic scale

$$IF_i(2015, x) = \frac{q_i(2015, x)}{q_i(2005, x)}$$

for $T = 2015$.

Correlations between the simulated improvement factors are relatively high at around 0.86 for ages 75 and 85, but slightly lower (0.79) at age 65 for 2015. The lower correlation for age 65 reflects the fact that this mortality rate relies on the cohort effect for 1950. Our historical dataset takes us up to 1945 and so we need to simulate forward to 1950. For ages 75 and 85 in 2015, we use previously fitted cohort effects for 1940 and 1930, respectively. The added uncertainty in the cohort effects for populations 1 and 2 reduces the correlation between the age-65 simulated improvement factors.

With a 10-year horizon, the main driver of correlation is the correlation between annual innovations in the period effect. For longer time horizons, the correlations increase as the mean-reversion in the spreads between the period and cohort effects become more significant. This is useful information when it comes to hedging.

- Correlation is investigated further in Figure 10, where we plot the simulated correlation between the simulated improvement factors at ages 65, 75 and 85 as a function of the time horizon. 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 10. 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.[25] After 15 years, the age 75 mortality also includes simulated cohort effects, so the age

---

[25]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.

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 $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$, 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 8). The correlation between estimates of the cohort effect for these years is quite small and has the immediate effect of dragging down the age 65 correlation plot (Figure 10, left) relative to its PC counterpart (Figure 10, right). In the longer run, this uncertainty in estimates of historical latent effects is replaced by uncertainty in the process parameters such as μ_{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 $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$ on the random-walk drift μ_{R2}.

A significant difference between the left- and right-hand plots in Figure 10 is that on the left the correlation between $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$ 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 corelations 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 findings of Coughlan et al. (2009b).
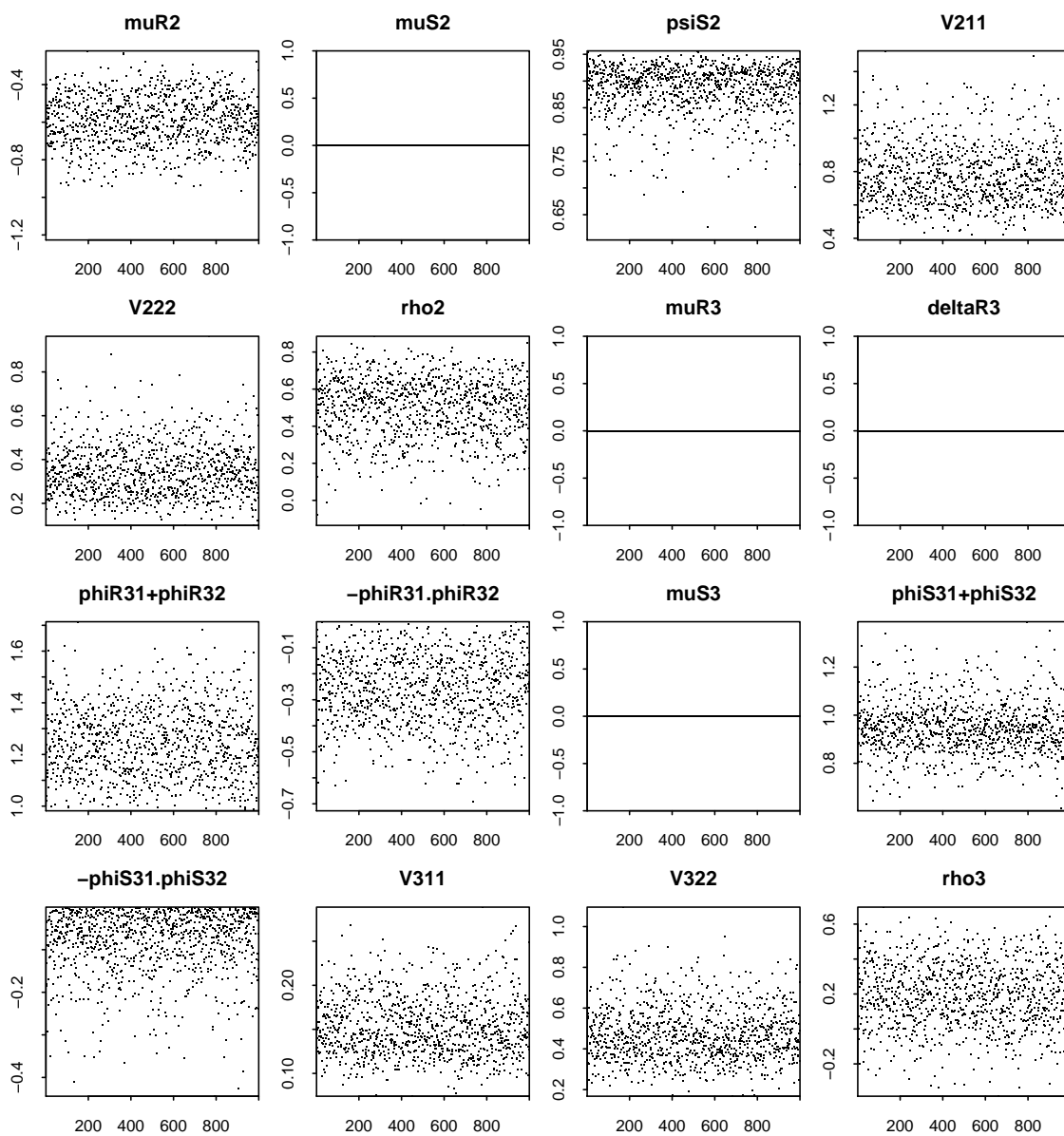
Figure 4: APC model, EW/CMI males. Metropolis-Hastings algorithm output: process parameters. 1000 values have been stored out of a Markov chain of length 50,000.
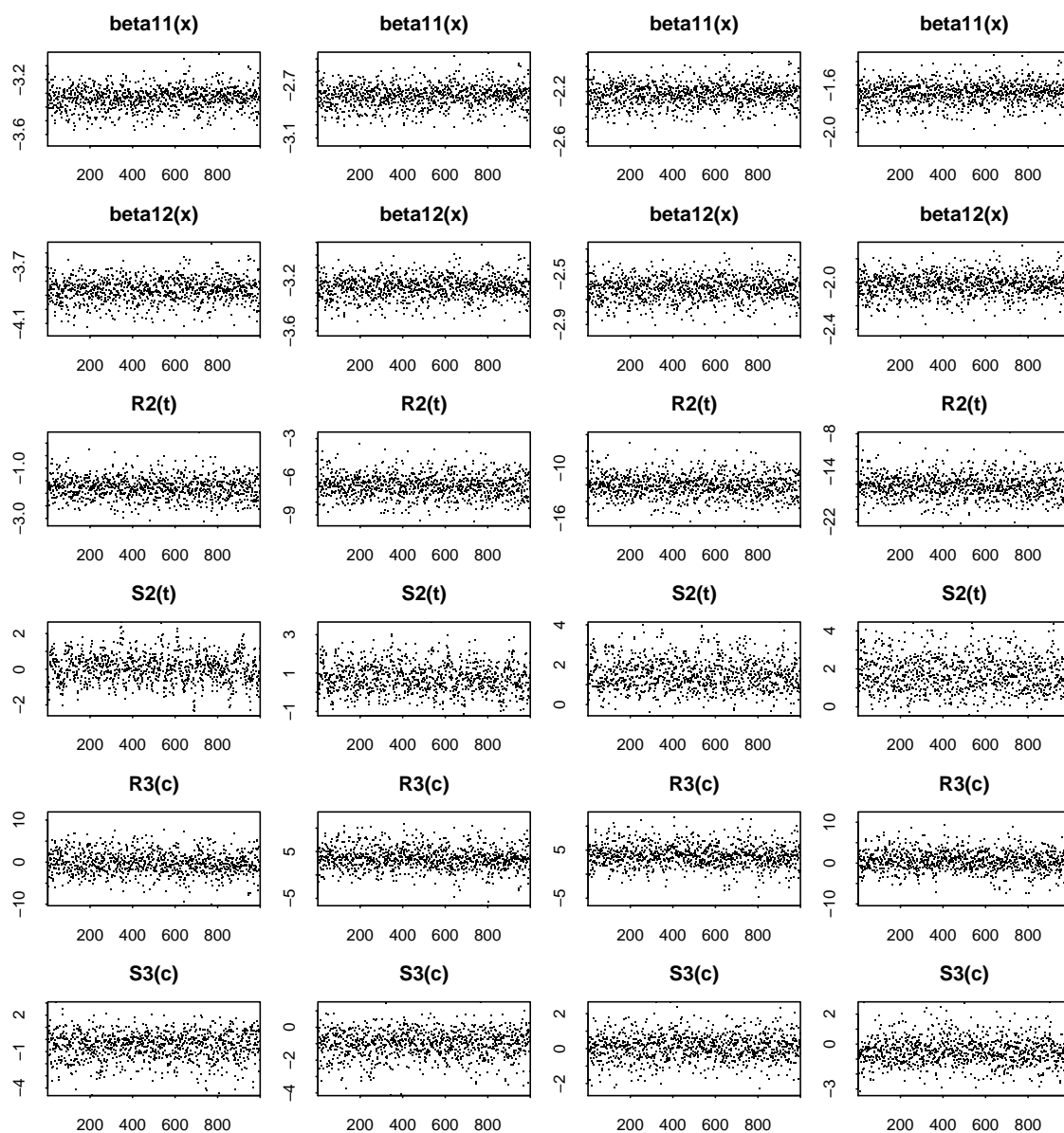
Figure 5: APC model, EW/CMI males. Metropolis-Hastings algorithm output: age, period and cohort parameters. 1000 values have been stored out of a Markov chain of length 50,000. Age effects are plotted for ages $x =$ 65, 71, 77, 83 (left to right). Period effects are plotted for years $t =$ 1969, 1978, 1987, 1996. Cohort effects are plotted for years of birth $c =$ 1885, 1900, 1915, 1930.
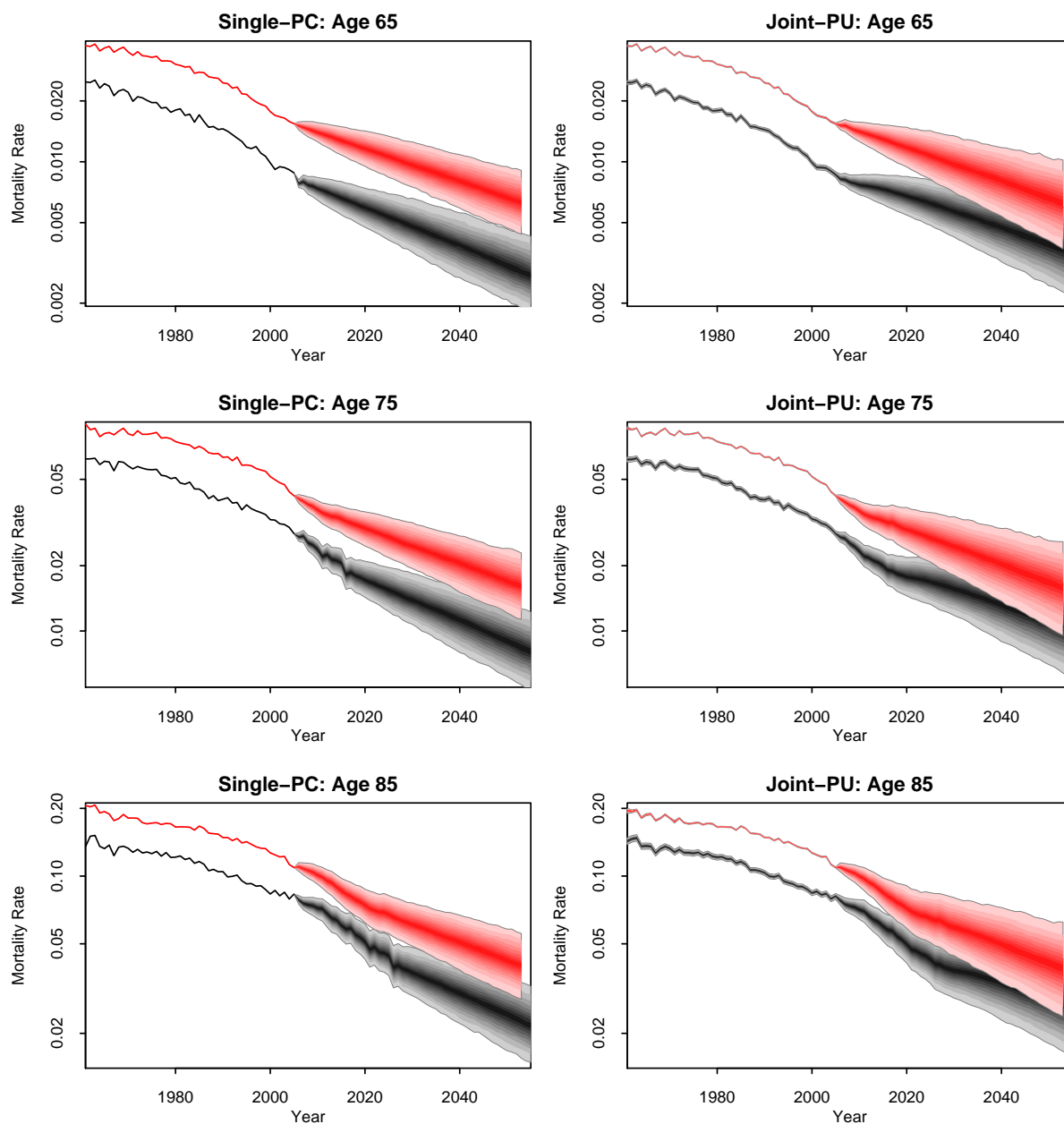
Figure 6: APC model. Mortality fan charts for ages 65, 75 and 85 for EW males (red fans) and CMI males (grey fans). Left-hand plots: fan charts constructed using the single-population models with no parameter uncertainty (PC). Right-hand plots: fan charts constructed using the joint-population model (MCMC) with parameter uncertainty (PU).
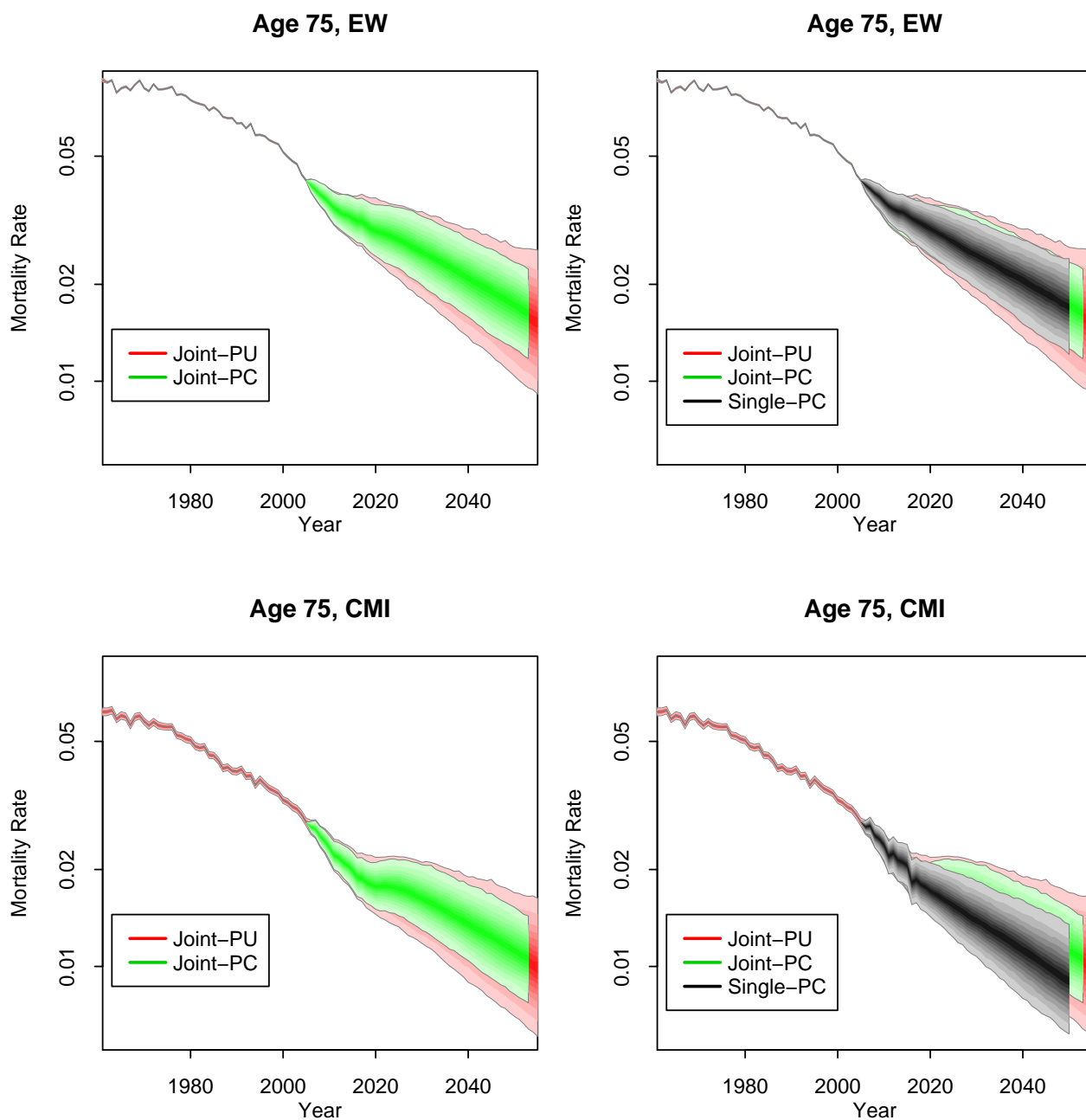
Figure 7: APC model, EW/CMI males. Comparison of fan charts based on (a) the new MCMC algorithm (red fans) using AR(2) cohort effects (left-hand panel) and (b) the original two-stage method using ARIMA(1,1,0) models for the cohort effect (right-hand panel).
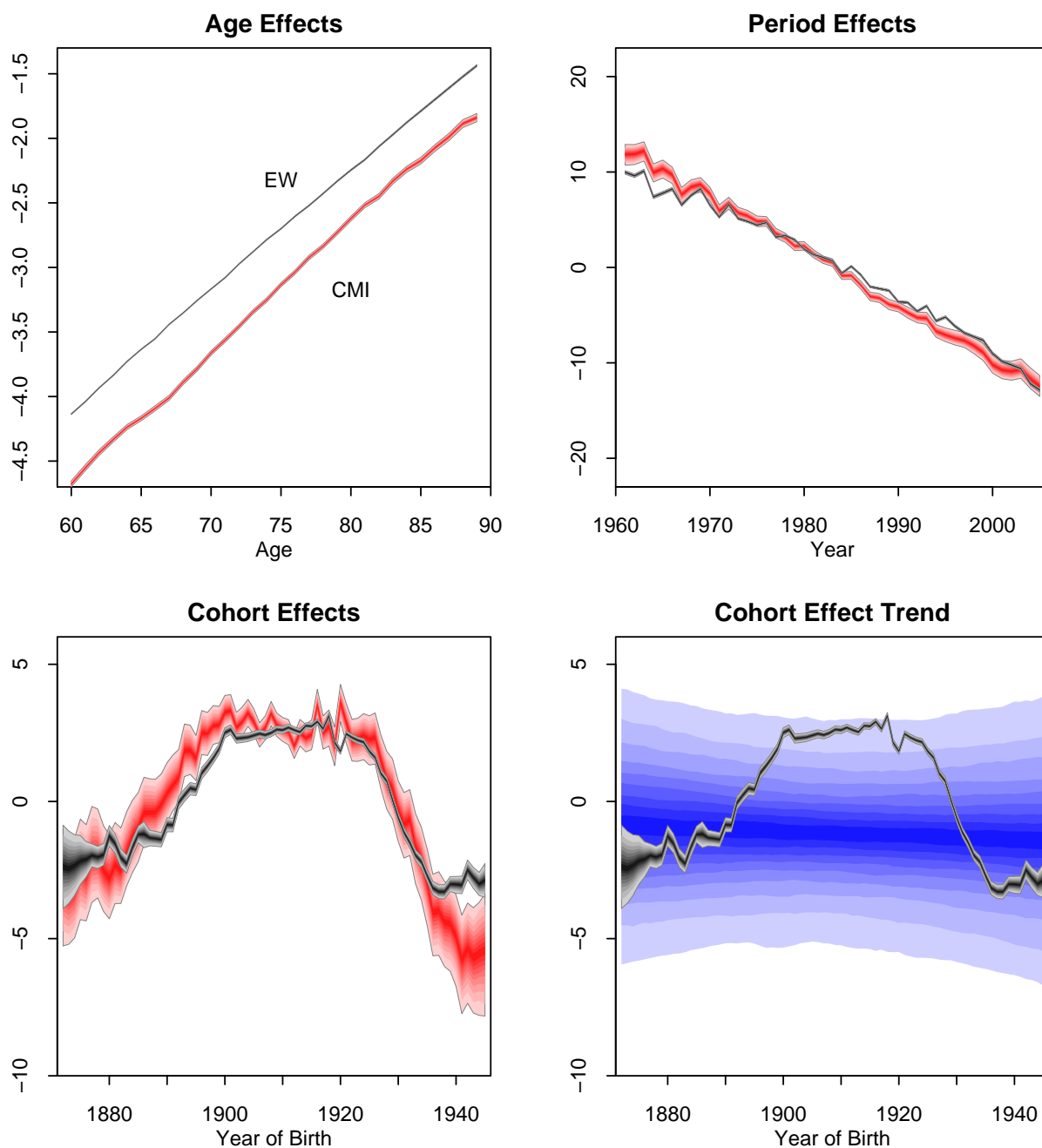
Figure 8: APC model. Age, period and cohort effects for EW males (grey fans) and CMI males (red fans). Bottom right: EW cohort effect (grey) and its underlying linear trend (blue fan). Age, period and cohort effects have been adjusted to satisfy identifiability constraints: $\sum_c R_3(c) = 0$, $\sum_c R_3(c)(c - \bar{c}) = 0$, $\sum_c S_3(c) = 0$, $\sum_t R_2(t) = 0$, $\sum_t S_2(t) = 0$. $R_3(c)$ is modelled as an AR(2) process around the linear trend $\mu_{R3} + \delta_{R3}(c - \bar{c})$. Bottom right shows credibility intervals for $\mu_{R3} + \delta_{R3}(c - \bar{c})$ (blue fans).
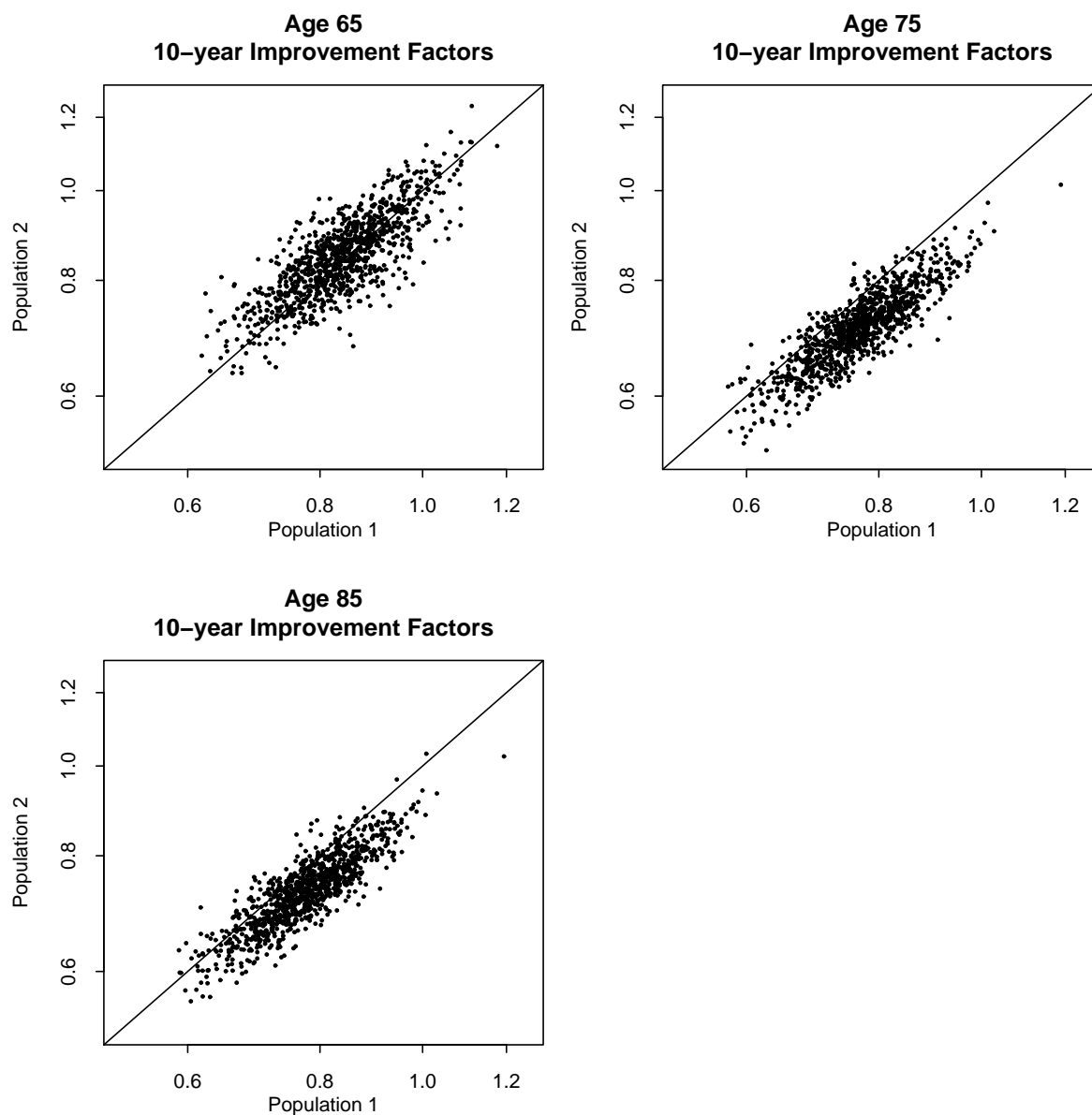
Figure 9: APC model. Correlation between EW males and CMI males projections. Simulated improvement factors over the period 2005 to 2015 (i.e. $q(2015, x)/q(2005, x)$). Correlations between improvement factors are 0.79, 0.86 and 0.86 for ages 65, 75 and 85 respectively.
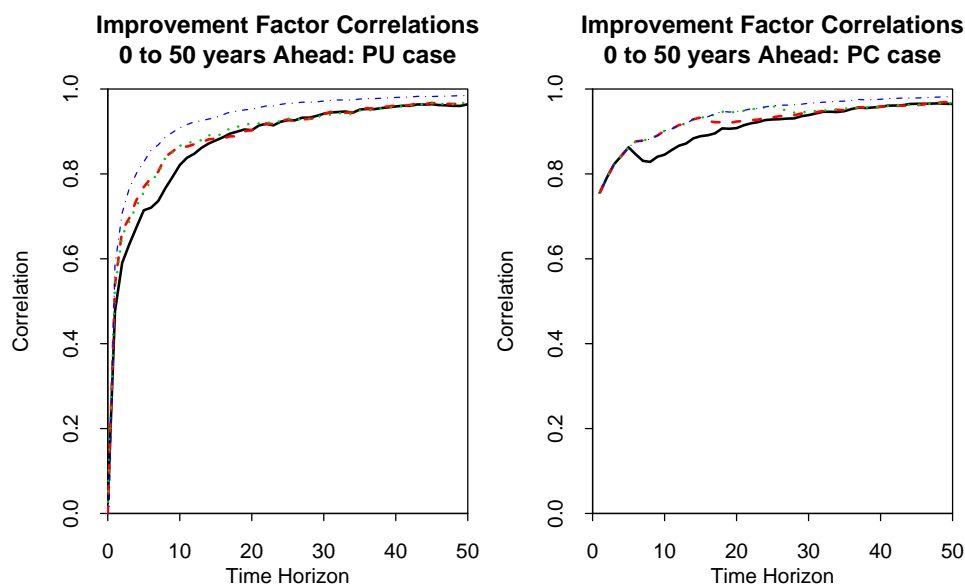
Figure 10: APC model. Correlation between simulated improvement factors in EW males and CMI males mortality rates as a function of the time horizon beyond 2005 – Age 65 (solid black line), 75 (dashed red line) and 85 (dotted green line) – and correlation between period effects $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$ as a function of the time horizon (blue dot-dashed line). Left: simulations incorporating parameter uncertainty (PU). Right: simulations with no parameter uncertainty (PC).
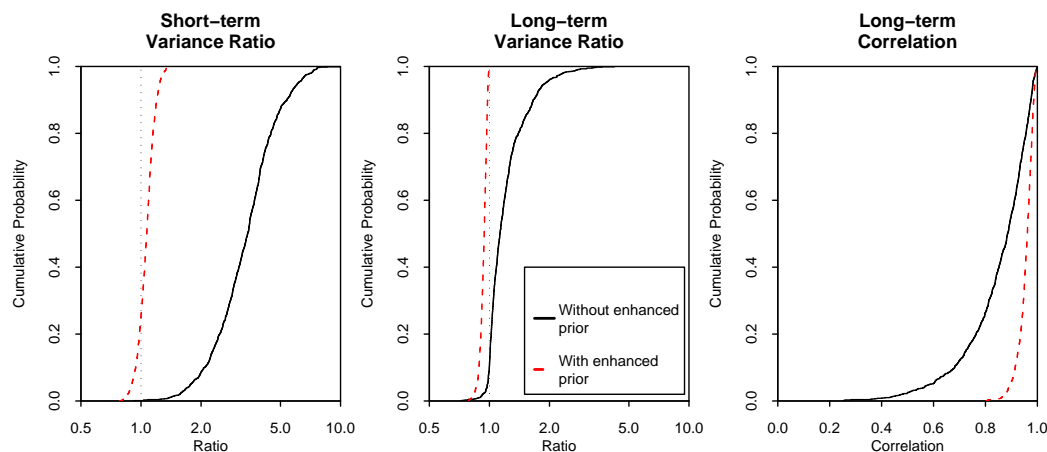
Figure 11: APC model, EW/CMI males. Left: Cumulative posterior distribution function (CPDF) of the ratio of the short-term volatility of $\gamma_c^{(32)}$ to the short-term volatility of $\gamma_c^{(31)}$. Middle: CPDF of the ratio of the long-term (stationary) variance of $\gamma_c^{(32)}$ to the long-term variance of $\gamma_c^{(31)}$. Hypothesis: both ratios should be close to 1. Right: Long-term correlation between $\gamma_c^{(32)}$ and $\gamma_c^{(31)}$. Hypothesis: long-term correlation should be reasonably strong. Solid black lines: without the use of the enhanced prior distributions (analysis 1). Dashed red lines: with the use of the enhanced prior distributions (analysis 2).

7 Analysis 2: EW/CMI Males, Ages 60-89, Years 1961-2005, Enhanced Priors

In Section 4.2, we discussed biologically reasonable criteria for the relationship between the short- and long-term variances of the two populations' cohort effects and their correlation. In Figure 11, we plot the posterior distribution for the relevant quantities.

In the left-hand panel, we plot the CDF corresponding to Analysis 1 (Section 6) (solid black line) of the ratio of the short-term (one-step-ahead conditional) variances of the population 1 and 2 cohort effects.[26] In the middle panel, we plot the CDF of the ratio of the long-term (unconditional) variances between the two cohort effects.[27] The right-hand panel shows the corresponding long-term correlation,[28] under Analysis 1 (solid black line).

We infer from these plots that a typical draw from the posterior distribution results in:

(a) $\gamma_c^{(32)}$ being very much more volatile in the short term compared with $\gamma_c^{(31)}$ (left-hand panel in Figure 11);

(b) $\gamma_c^{(32)}$ being significantly more variable in the long term compared with $\gamma_c^{(31)}$ (middle panel);

(c) reasonably strong correlation between $\gamma_c^{(32)}$ and $\gamma_c^{(31)}$ in the long term (right-hand panel).

Observations (a) and (b) are contrary to what we considered to be biologically reasonable in Section 4.2.

Consequently, we repeated our analysis using enhanced priors 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 variance of $\gamma_c^{(32)}$ to the conditional 1-step-ahead variance of $\gamma_c^{(31)}$.

---

[26]The short-term ratio is defined as $Var(\gamma_c^{(32)}|\mathcal{G}_{c-1}, V^{(3)})/Var(\gamma_c^{(31)}|\mathcal{G}_{c-1}, V^{(3)})$, 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 θ. Hence, $Var(\gamma_c^{(32)}|\mathcal{G}_{c-1}, V^{(3)}) = V_{11}^{(3)} - 2V_{21}^{(3)} + V_{22}^{(3)}$, and $Var(\gamma_c^{(31)}|\mathcal{G}_{c-1}, V^{(3)}) = V_{11}^{(3)}$.

[27]The long-term variance ratio is $Var(\gamma_c^{(32)})/Var(\gamma_c^{(31)})$.

[28]Defined as $cor(\gamma_c^{(32)}, \gamma_c^{(31)})$.

- 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, as we discuss below, their impact is limited primarily to precisely those factors that we seek to make more biologically reasonable. 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 restricts 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.

The dashed red lines in Figure 11 show the posterior distributions of the variance ratios and the long-term correlation when we incorporate these enhanced priors. We can see that the two populations now have similar levels of variability in the cohort effect in both the short and long term and that the long-term correlation is higher than before. Further analysis reveals that balancing the two variances is achieved partly by reducing the volatility of $\gamma_c^{(32)}$ and partly by $\gamma_c^{(31)} = R_3(c)$ becoming more volatile.

In Figure 12, we compare mortality fan charts using the enhanced priors (red) with the equivalent results without the use of enhanced priors. We can see that the introduction of these priors has little impact on the headline emergence of risk over time.

Figure 13 shows the correlation plots based on the enhanced priors with and without parameter uncertainty in the joint-population model. Comparing this with Figure 10, we see that correlations are slightly lower when we use enhanced priors. For age 65 in the PC case, the dip after 5 years when the simulated cohort effects kick in is more prominent. This is because the central cohort effect, $R_3(c)$, is more volatile.

Figure 14 focuses on the impact of using an enhanced prior on the fitted CMI cohort effect, $\gamma_c^{(32)}$. Again we can see that the impact of the enhanced prior is relatively modest. In the right-hand plot (enhanced priors), close inspection of the details does reveal some damping down of volatility compared with the left-hand plot. So, although we have obtained more consistent volatility estimates, this has not been at the expense of allowing the data to speak for themselves.

Enhanced priors were not felt to be necessary for modelling the period effects. Specifically, the 1-year-ahead variances of $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$ resulting from Analysis 1 were found already to be reasonably close and, therefore, biologically reasonable. However, variants of the framework might include an enhanced prior along the same lines as for the cohort effect.
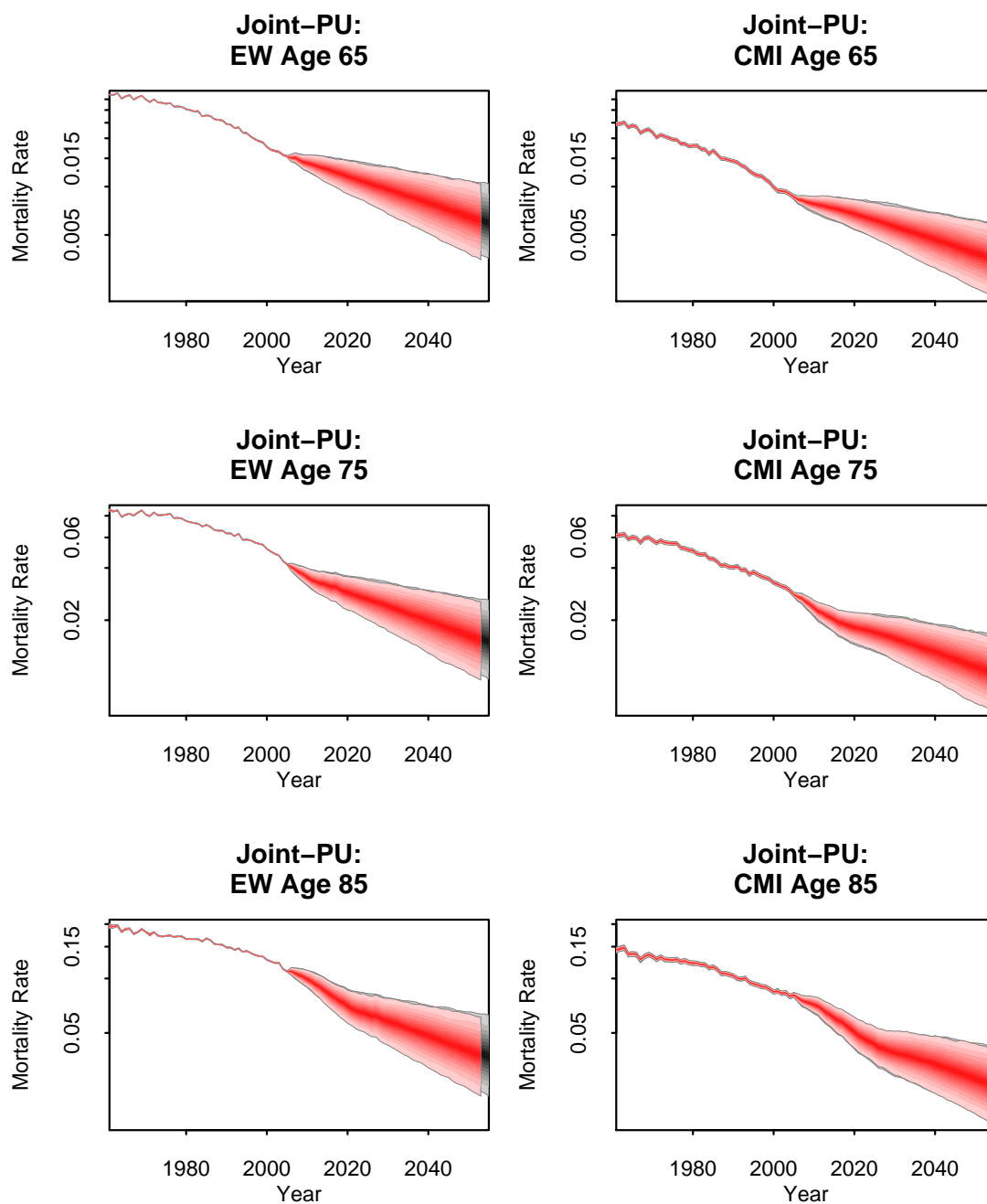
Figure 12: APC model. Mortality fan charts for ages 65, 75 and 85 with (red fans) and without (grey fans) the enhanced priors. Left-hand plots: EW males. Right-hand plots: CMI males.
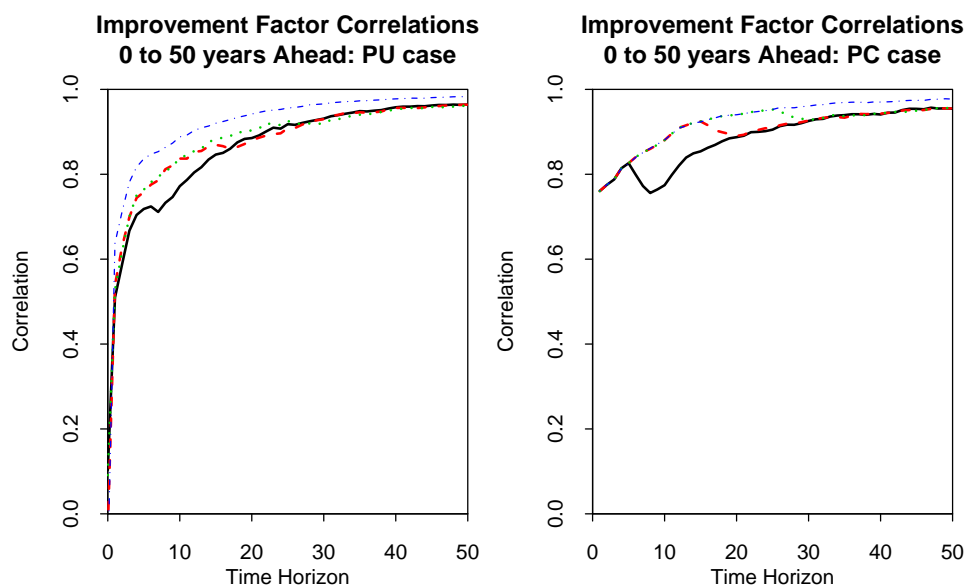
Figure 13: APC model with enhanced priors. Correlation between simulated improvement factors as a function of the time horizon – Age 65 (solid black line), 75 (dashed red line) and 85 (dotted green line) – and correlation between period effects $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$ as a function of the time horizon (blue dot-dashed line). Left: simulations incorporating parameter uncertainty (PU). Right: simulations with no parameter uncertainty (PC).
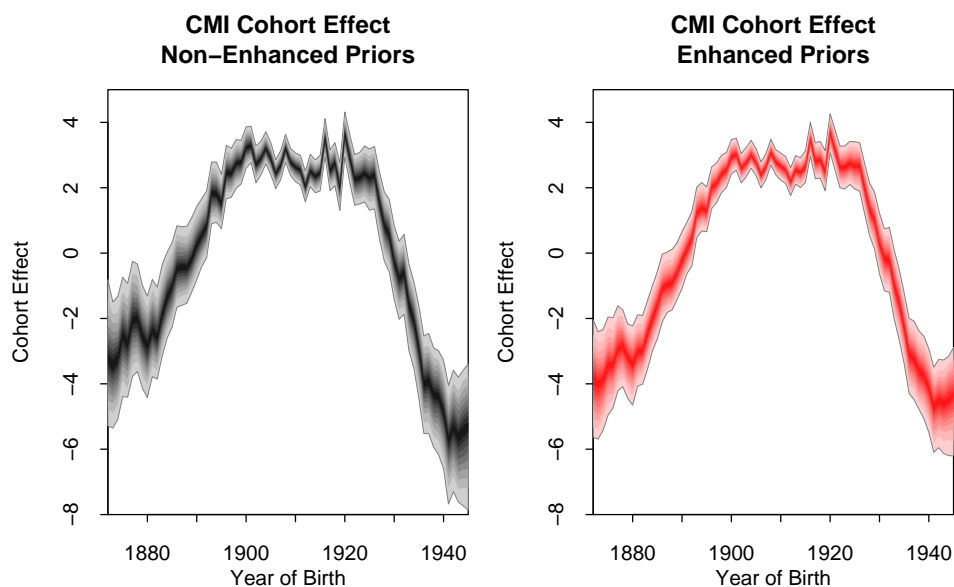
Figure 14: APC model. Cohort effect for CMI males without enhanced priors (left-hand plot), and with enhanced priors (right-hand plot).

8   Analysis 3: EW/CMI Males, Ages 60-89, years 1961-2005, Enhanced Priors, ARIMA(1,1,0) Model for the Main Cohort Effect

In previous work (Cairns et al. 2008b), we modelled the cohort effect using an ARIMA(1,1,0) process. In Analysis 3, we replace the AR(2) model for $R_3(c)$ with an ARIMA(1,1,0) model (in programming terms, this is achieved by fixing $\phi_{R31} = 0.999$ in the AR(2) version of the model).

Results are very similar to the analysis 2 results for the AR(2) cohort model with enhanced priors. Although the correlation plot (Figure 15) differs slightly from the benchmark (Figure 10), we see it is quite close to the analysis 2 correlation plot (Figure 13) which also uses an enhanced prior.
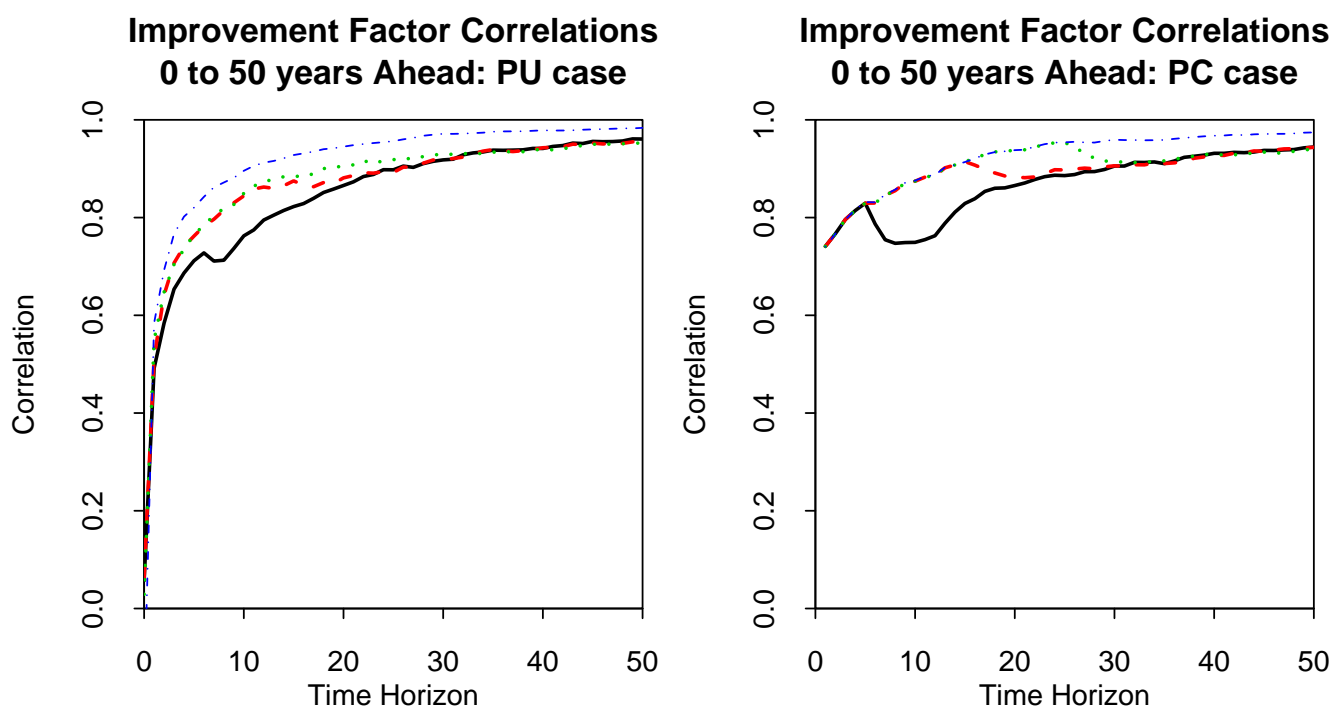
Figure 15: APC model, enhanced priors, ARIMA(1,1,0) model for $R_3(c)$. Correlation between simulated improvement factors as a function of the time horizon. Age 65 (solid black line), 75 (dashed red line) and 85 (dotted green line).

9 Analysis 4: EW/CMI Females, Ages 60-89, Years 1983-2003, Enhanced Priors

In our next analysis, we move on to consider EW versus CMI females, applying the MCMC approach with enhanced priors.

In contrast with the males data, data for CMI females are only available for the period 1983 to 2003. Additionally, the size of the females' CMI population is much smaller than that of CMI males. The small size of the data mainfests itself in two ways and both are linked to parameter estimation error. For the MCMC approach, we see in Figure 16 that credibility intervals for the age, period and cohort effects are much wider for the CMI females (and much wider than for CMI males, which has a larger population). We have also plotted parameter estimates (dots) using the single-population, two-stage approach. Here, the small CMI population results in very noisy estimates, particularly for the cohort effect.

Figure 17 shows fan charts for mortality at age 75, for the single-population PC model, the two-population PC model and the two-population PU model. As with EW males, the PC fans for EW females (upper right-hand plot, grey and green fans) are quite similar, while the PU fan (red) is rather wider. For the CMI data (lower plots), there is a rather bigger difference between the single-population model and the two-population model, the latter's fan being much narrower. Again this tells us that the MCMC approach is very effective at dampening the impact of small population noise.

The two right-hand plots in Figure 17 also reveal a further benefit of using the two-population model. When we compare the single-population projections for the EW and CMI populations (grey fans in both plots), we see that the EW and CMI forecasts cross over around 2050. Historically, CMI females' mortality has been lower than EW females reflecting the wealthier and healthier status of the insured lives. We might reasonably expect this to continue in the future, but this is not reflected in the single-population forecasts. This might be mitigated by altering the length of the lookback window or, perhaps, by making a subjective adjustment to the CMI drift parameters. The two-population model takes a more objective approach. The key assumption of non-divergence over time of the two populations, along with the relatively very large size of the EW population and the 0/1 weights in the spreads model, ensures that the EW projection dominates and the CMI forecasts move roughly in parallel in the long term. As a consequence, the two-population forecasts for CMI females (green and red fans) are significantly lower than the single-population forecast (grey fan).

Correlations are plotted in Figure 18. The PC correlations are a bit lower than the corresponding correlations for males reflecting different values for the process parameters governing the dynamics of $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$. There is also a bigger difference

between the PC and PU cases, which is the result of the wider credibility intervals for the CMI females' age, period and cohort effects (Figure 16).
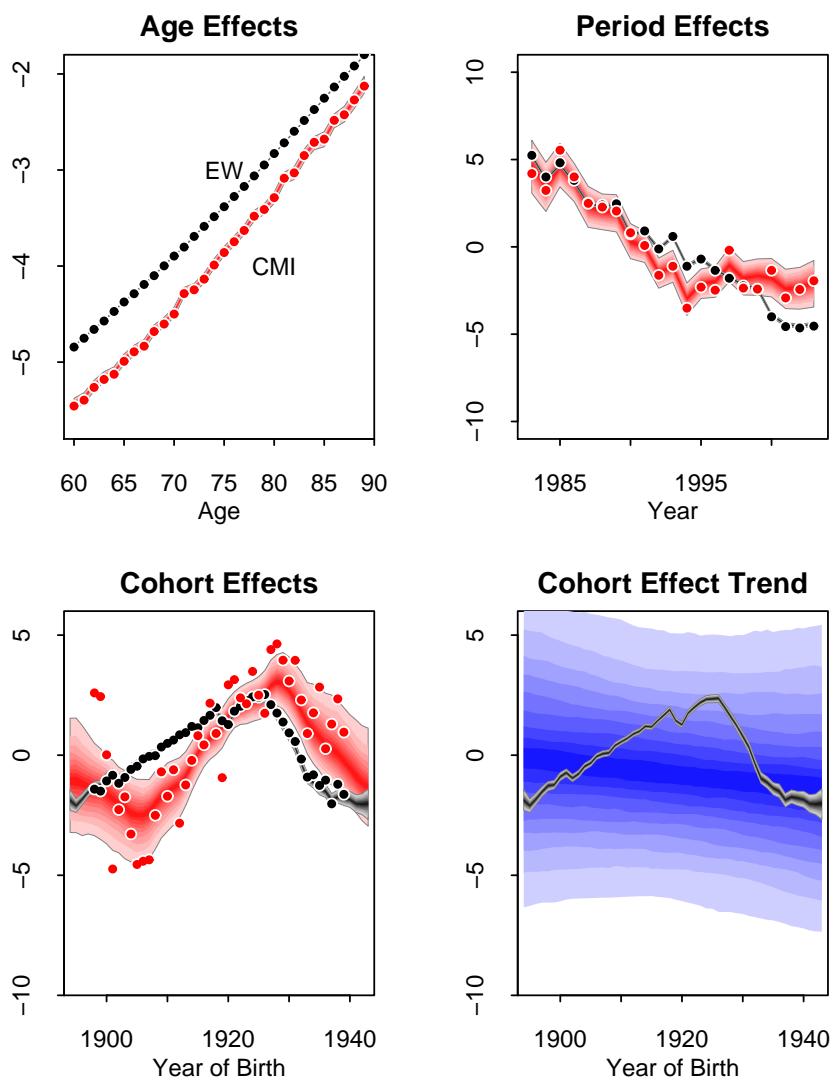

Figure 16: APC model, enhanced priors, 1983-2003. Age, period and cohort effects for EW females (grey fans) and CMI females (red fans). Bottom right: EW cohort effect (grey) and its underlying linear trend (blue fan). Age, period and cohort effects have been adjusted to satisfy identifiability constraints: $\sum_c R_3(c) = 0$, $\sum_c R_3(c)(c - \bar{c}) = 0$, $\sum_c S_3(c) = 0$, $\sum_t R_2(t) = 0$, $\sum_t S_2(t) = 0$. $R_3(c)$ is modelled as an AR(2) process around the linear trend $\mu_{R3} + \delta_{R3}(c - \bar{c})$. Bottom right shows credibility intervals for $\mu_{R3} + \delta_{R3}(c - \bar{c})$ (blue fans). The black (EW) and red (CMI) dots show corresponding point estimates for $\kappa_t^{(i2)}$ and $\gamma_c^{(i3)}$ effects using the two-stage single-population model.
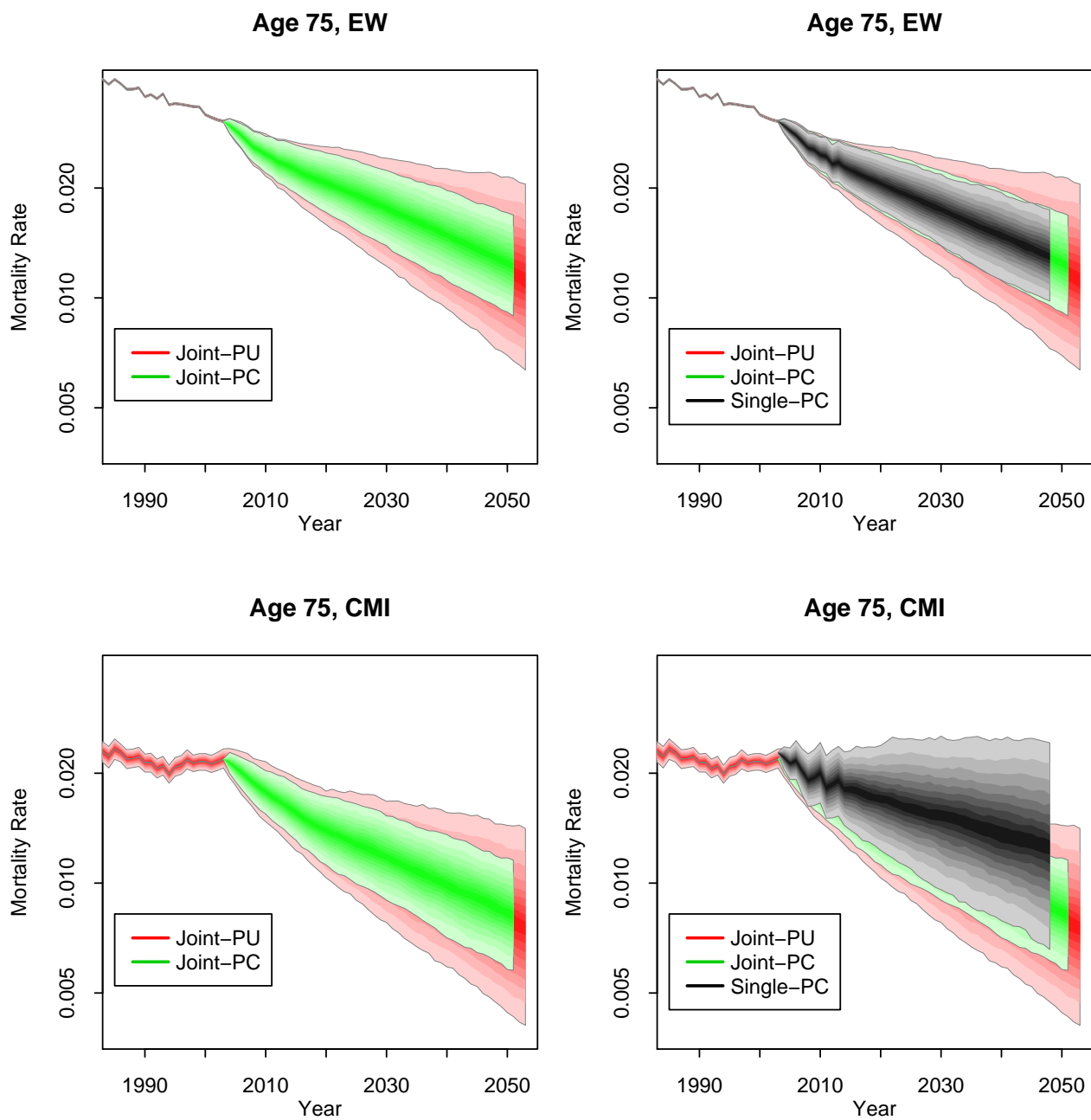
Figure 17: APC model, enhanced priors, 1983-2003. Comparison of fan charts for EW and CMI females based on (a) the new MCMC algorithm (red fans) using AR(2) cohort effects and (b) the original two-stage method using ARIMA(1,1,0) models for the cohort effect.
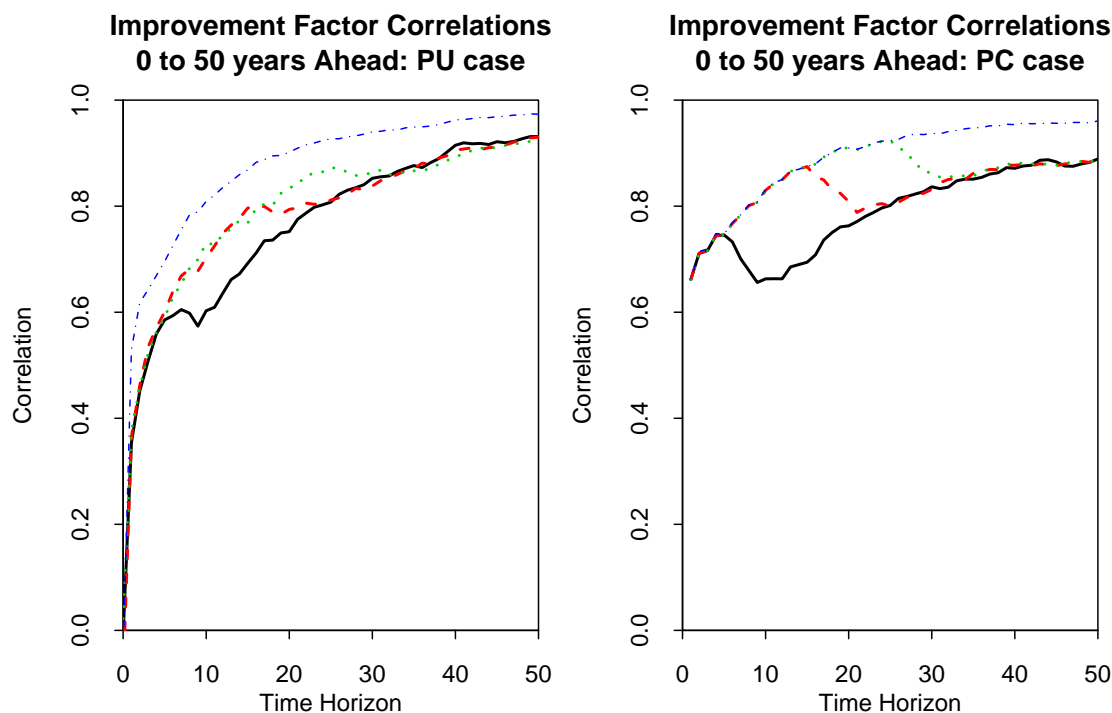
Figure 18: APC model, enhanced priors, EW/CMI females, 1983-2003. Correlation between simulated improvement factors as a function of the time horizon – age 65 (solid black line), 75 (dashed red line) and 85 (dotted green line) – and correlation between period effects $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$ as a function of the time horizon (blue dot-dashed line). Left: simulations incorporating parameter uncertainty (PU). Right: simulations with no parameter uncertainty (PC).

10 Analysis 5: EW/CMI Females, Ages 60-89, Years 1983-2003 versus 1961-2003 and 1961-2007

CMI females data are currently only available for years 1983-2003. In contrast, EW females data are currently available from 1961 to 2007. Here, we consider augmenting the basic 1983-2003 dataset with this additional EW data. A key advantage of the MCMC approach is that it allows us to include missing data: in this case, CMI data from 1961-1982 and 2004-2007. Missing data are dealt with by assigning a weight of zero to the relevant cells: just as we did for certain parts of the EW males dataset (Section 3). We do this in two steps: first, we add the extra data from 1961-1982 and, second, we add the extra data from 2004-2007.

We just consider one plot here, Figure 19. This plot allows us to compare forecasts for ages 65, 75 and 85 based on the basic and then the augmented datasets. The grey fans show projections based on the combined 1983-2003 dataset. The red fans are based on data from 1961-2003. The green fans are based on data from 1961-2007. We can comment as follows:

- The main impact of including data from 1961-1982 is to tilt slightly the central trajectory of both the EW and CMI projections. The average mortality improvement rate over 1961-2003 was lower than the 1983-2003 average and this results in a lower forecast improvement rate. The change in the width of the grey and red fan charts reflects a more complex combination of factors. Amongst these, the red fan will be narrower because there is less parameter uncertainty in the μ_{R2} drift term.

- When we include data from 2004-2007, we see a significant impact on the EW fans. The fan now starts to widen out only after 2007, and, for a given future year, the prediction interval is narrower, reflecting the shorter out-of-sample forecasting period. In general, the fans have been shifted down slightly.

- The impact of the inclusion of the 2004-2007 EW data on the CMI forecasts is more difficult to unravel. The most obvious impact is a shift down from the red to the green fan in the central projection mimicking the shift in the EW forecasts. In the longer term, the width of the fans is also narrowed slightly, in line with changes in the EW fans. The impact of the additional data in the shorter term is more difficult to see and this is because of the considerable uncertainty in estimates of the historical age, period and cohort effects as reflected in the width of the CMI mortality fans up to 2003.

In conclusion, the major impact on CMI projections comes from the basic coupling of this population to the EW females dataset (Figure 17). Using greater amounts of EW data (the most recent data, in particular) has a noticeable but less significant impact.
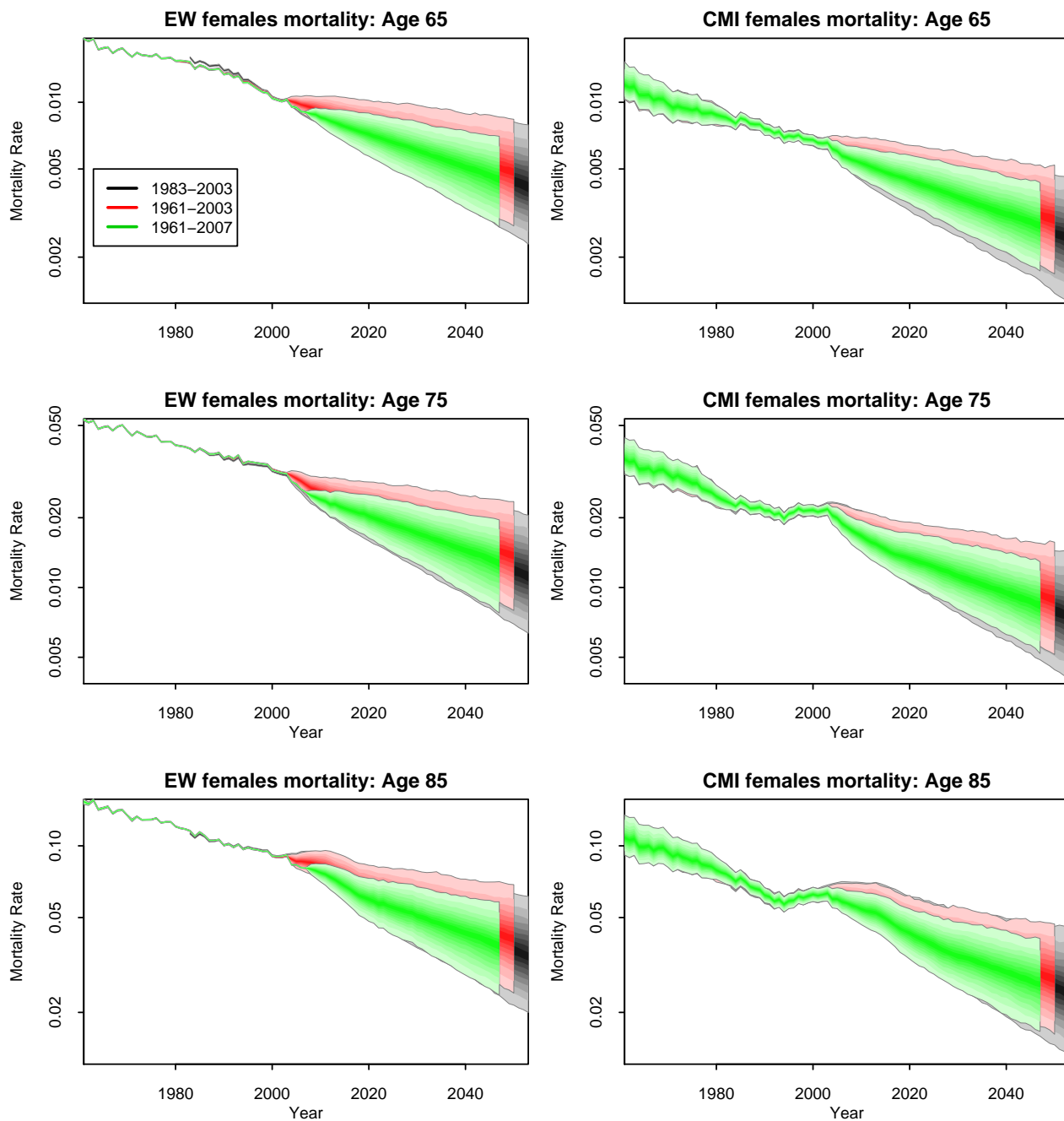
Figure 19: APC model, enhanced priors, EW/CMI females. Comparison of mortality fan charts for ages 65, 75 and 85 based on years 1961-2003 (grey), 1961-2005 (red), and 1961-2007 (green) and age range 60-89.

11 Analysis 6: EW/CMI Males, Ages 60-89, Years 1961-2005, Lee-Carter Model

In this section, we use the same dataset as Analysis 1 (Section 6), but now we use the LC model instead of the APC model.

The basic model here is

$$\log m_i(t, x) = \beta_x^{(1i)} + \beta_x^{(2i)} \kappa_t^{(2i)}$$

with the constraint that, for all x, $\beta_x^{(21)} = \beta_x^{(22)}$. The time-series model for $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$ is the same as before for the two-population version of the APC model. Priors are also as before for those parameters relating to $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$. The structure is represented in Figure 20.

With APC, the pseudo-Gibbs samplers for $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$ took advantage of the fact that the implicit $\beta_x^{(2i)} = 1/n_a$ were all constant. Now that the $\beta_x^{(21)}$ are age dependent, the ensuing non-linearities made the updating of the $\kappa_t^{(2i)}$ more difficult and we found that mixing was a bit slower than before, but still acceptable: see Figures 22 and 21.

We can make the following observations on the results:

- Credibility intervals for the age and period effects are plotted in Figure 23. As before, the intervals are significantly wider for the CMI population. The common $\beta_x^{(21)} = \beta_x^{(22)}$ function clearly declines with age, reflecting the slower rate of improvement in mortality at higher ages over the last 45 years.

- In Figure 24, we compare forecasts for the two-population models for LC and APC. We can see substantial differences between the two: in particular LC has much narrower fans at high ages, as has been remarked elsewhere (see, for example, Cairns et al. 2008b). Further, the LC fans (especially EW) have almost linear trends because of the lack of a cohort effect, whereas the impact of the cohort effect in APC is quite prominent. Thus at age 85, we see a steeper downwards trend after 2005 compared with LC as the 1930 "golden" cohorts move through.

- Figure 25 compares the one- and two-population PC and the two-population PU forecasts. Results are similar to those discussed earlier for APC.

- Correlations over time are plotted in Figure 26. The lack of a cohort effect means that there is very little difference between the correlations at different ages. The correlations are slightly higher (more so for age 65) than under APC, with a smaller difference between the PC and PU cases, suggesting that uncertainty in in estimates of historical cohort effects in APC is quite significant.
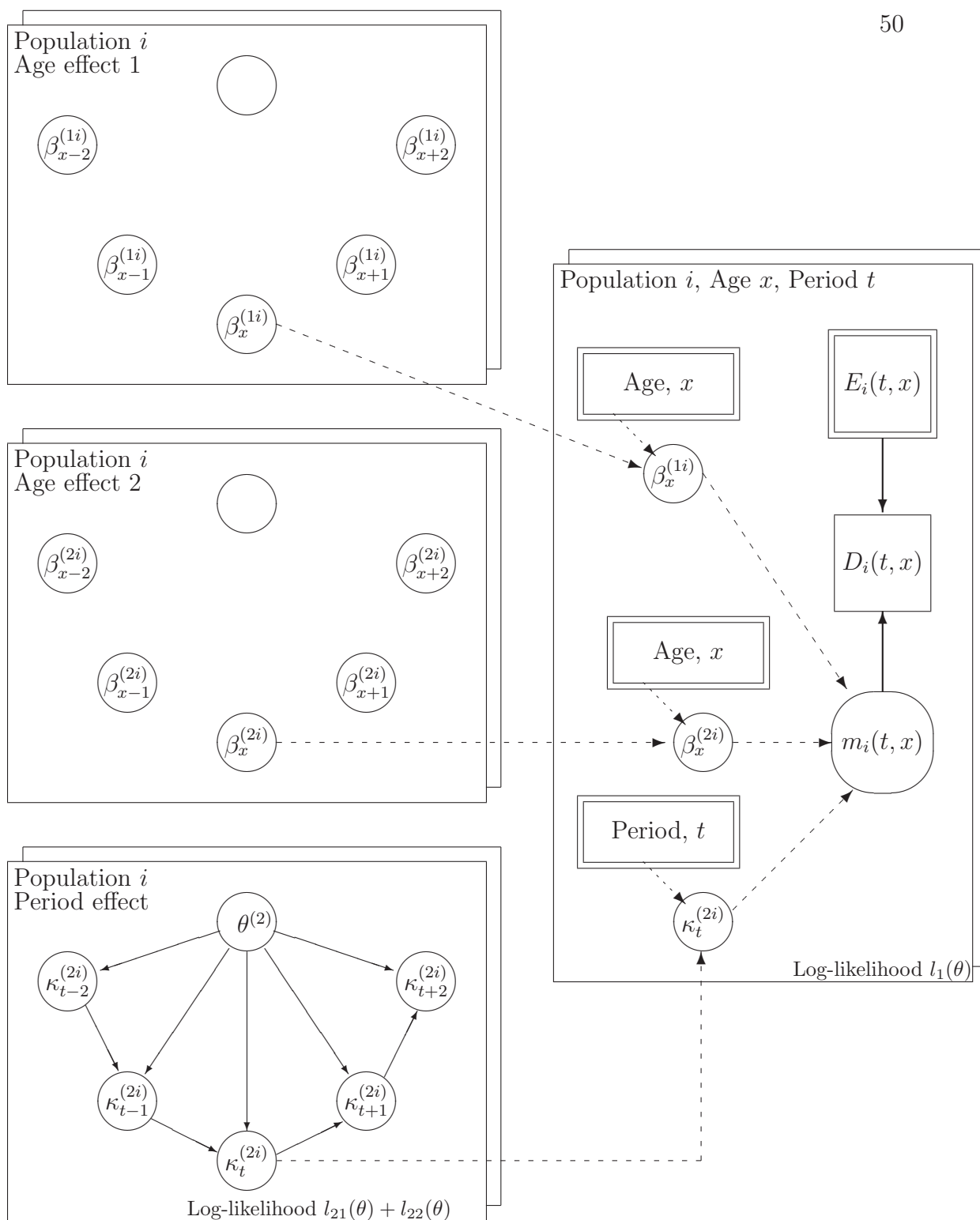
Figure 20: Directed acyclic graph for the two-population version of the Lee-Carter model. Details as in Figure 3.
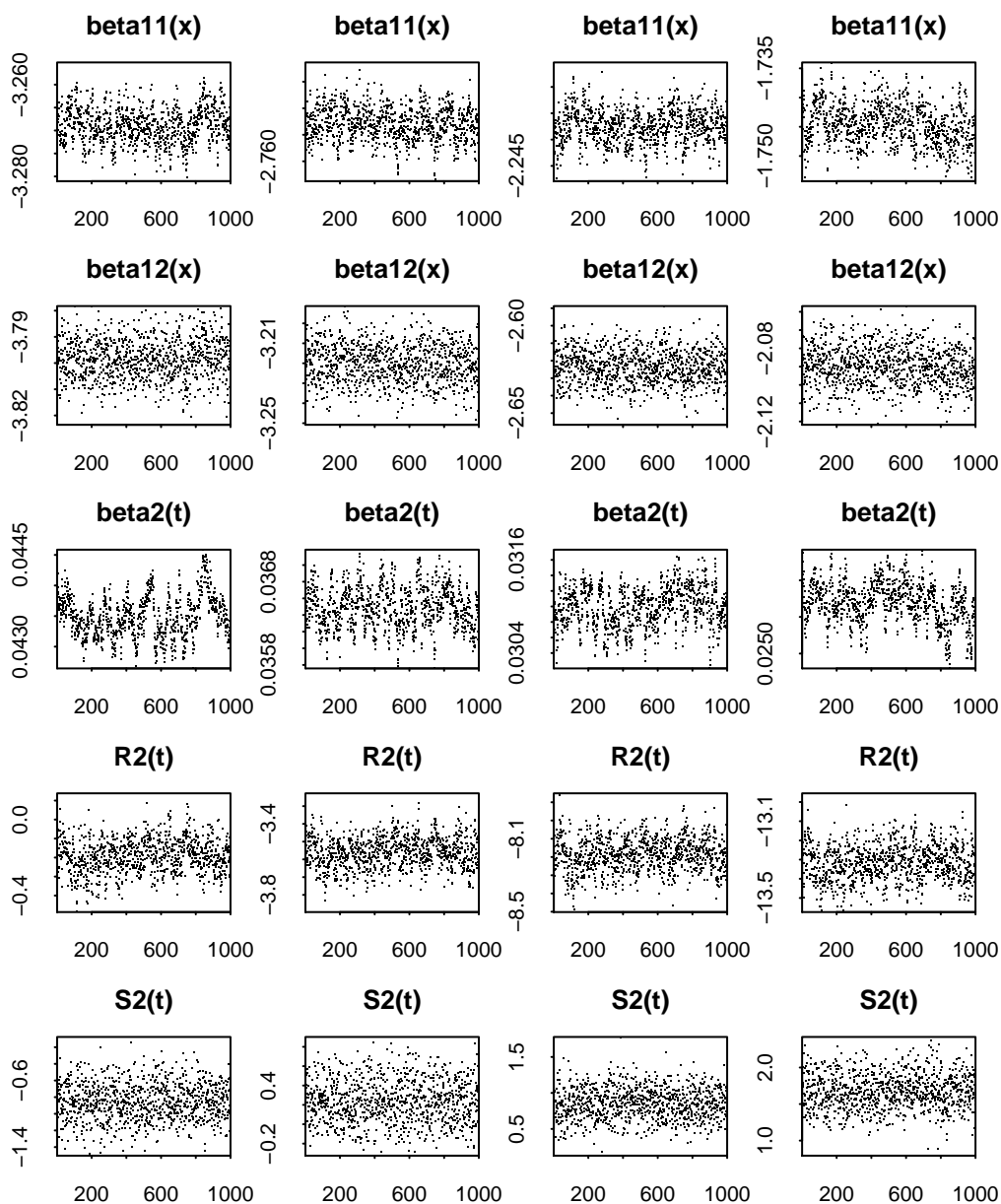
Figure 21: LC model, EW/CMI males. Metropolis-Hastings algorithm output: age, period and cohort parameters. 1000 values have been stored out of a Markov chain of length 50,000. Age effects are plotted for ages $x =$65, 71, 77, 83. Period effects are plotted for years $t =$1969, 1978, 1987, 1996.
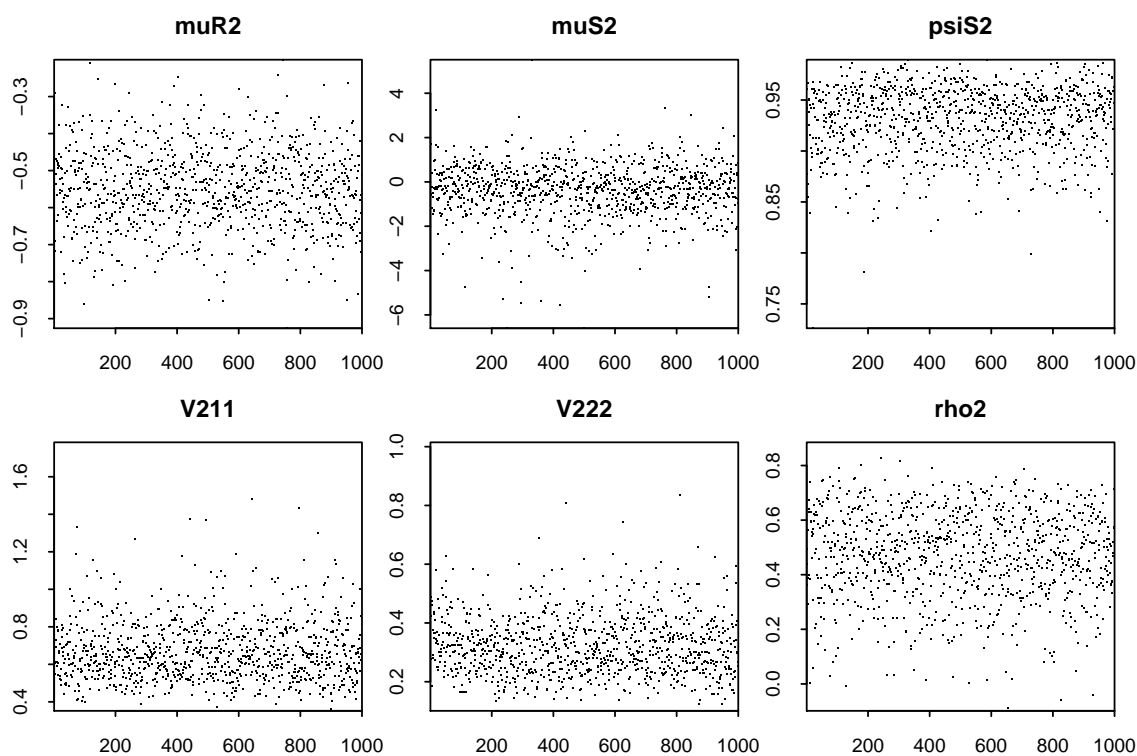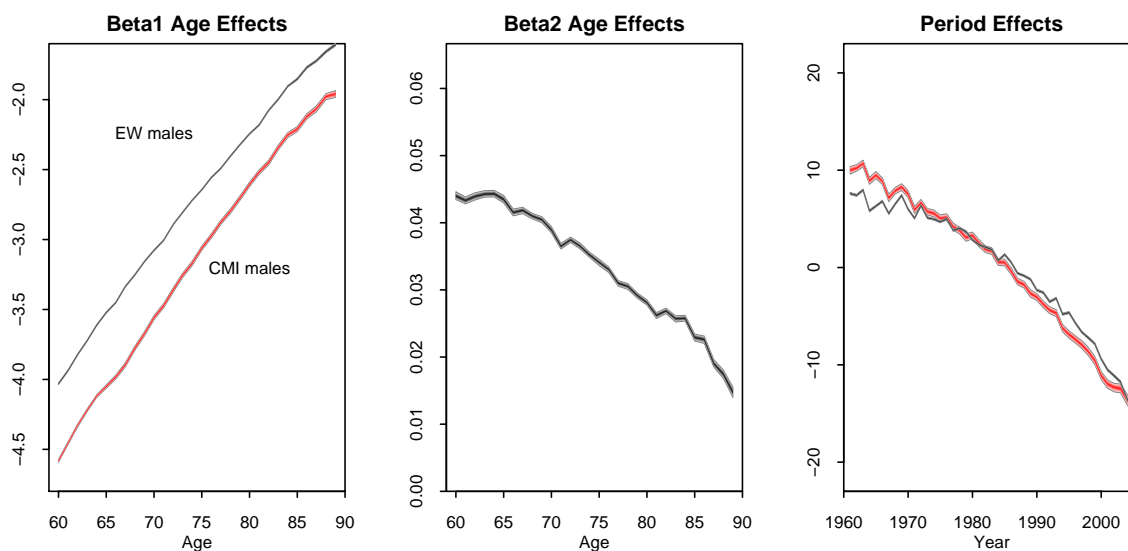
Figure 22: LC model, EW/CMI males. Metropolis-Hastings algorithm output. Process parameters. 1000 values have been stored out of a Markov chain of length 50,000.

Figure 23: LC model. Age, period and cohort effects for EW males (grey fans) and CMI males (red fans). Age, period and cohort effects have been adjusted to satisfy identifiability constraints: $\sum_t R_2(t) = 0$, $\sum_t S_2(t) = 0$.
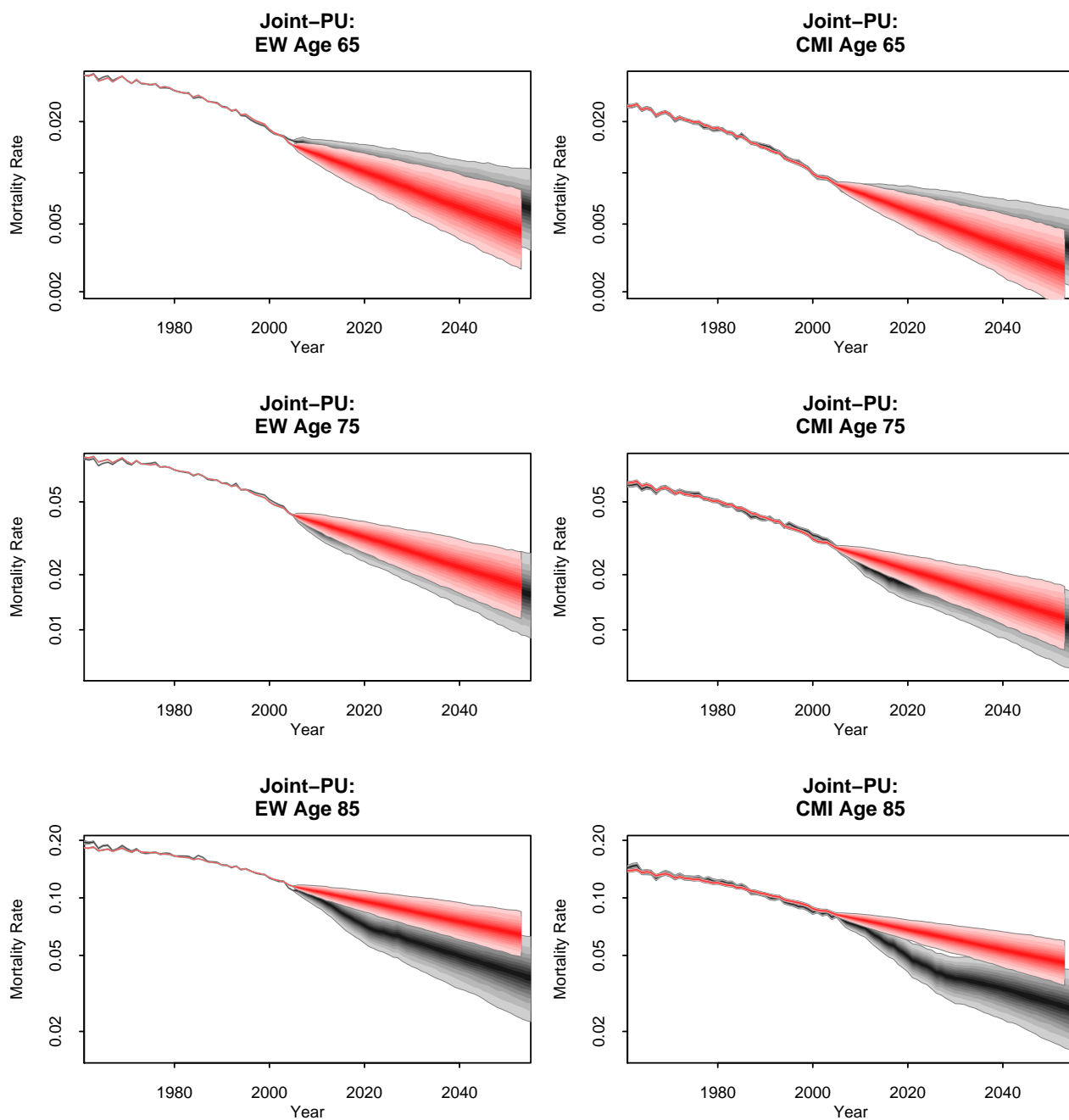
Figure 24: LC model, EW/CMI males. Comparison of fan charts based on (a) the MCMC algorithm for the two-population LC model (red fans) and (b) the MCMC algorithm for the APC model without enhanced priors (grey fans).
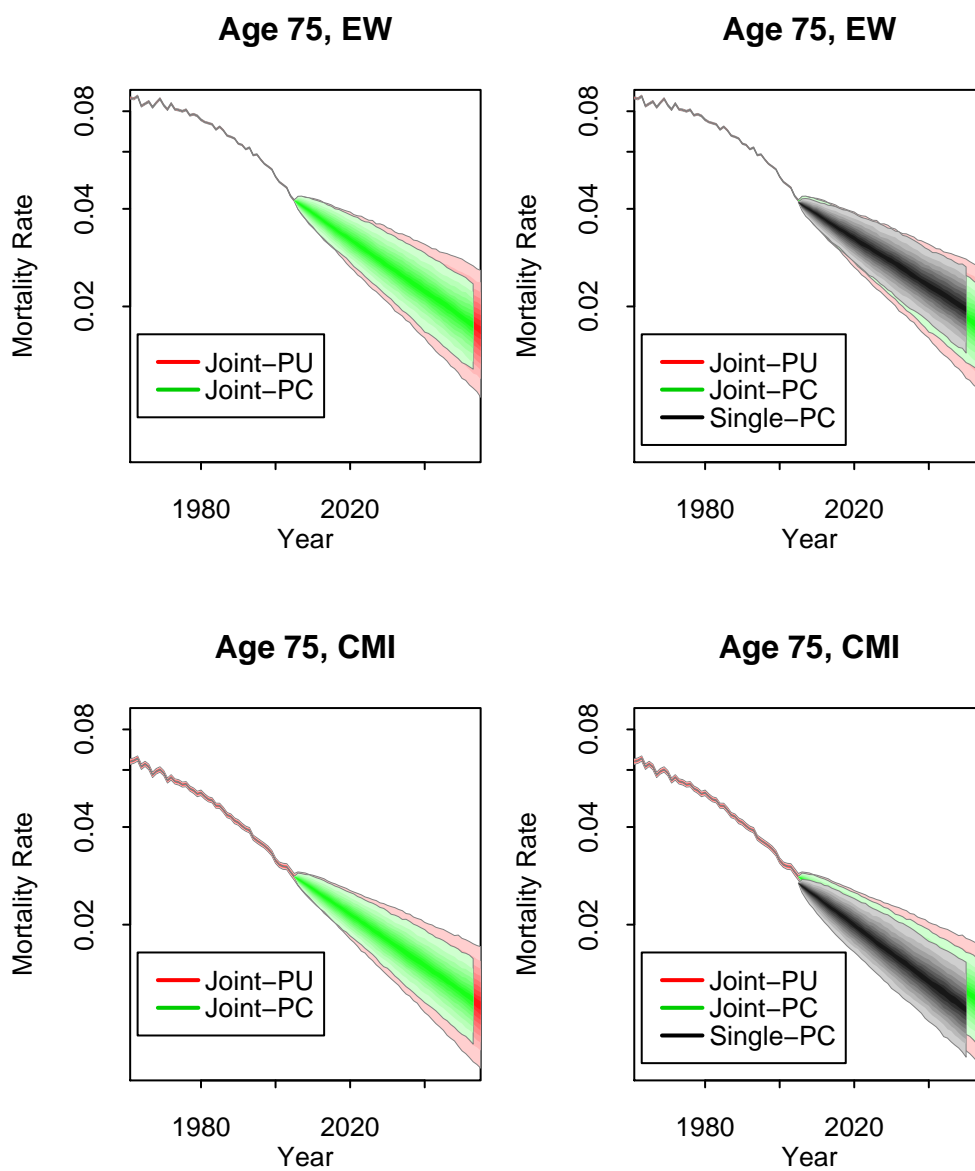
Figure 25: LC model, EW/CMI males. Comparison of fan charts based on (a) the new MCMC algorithm (red fans) and (b) the original two-stage method (grey fans).
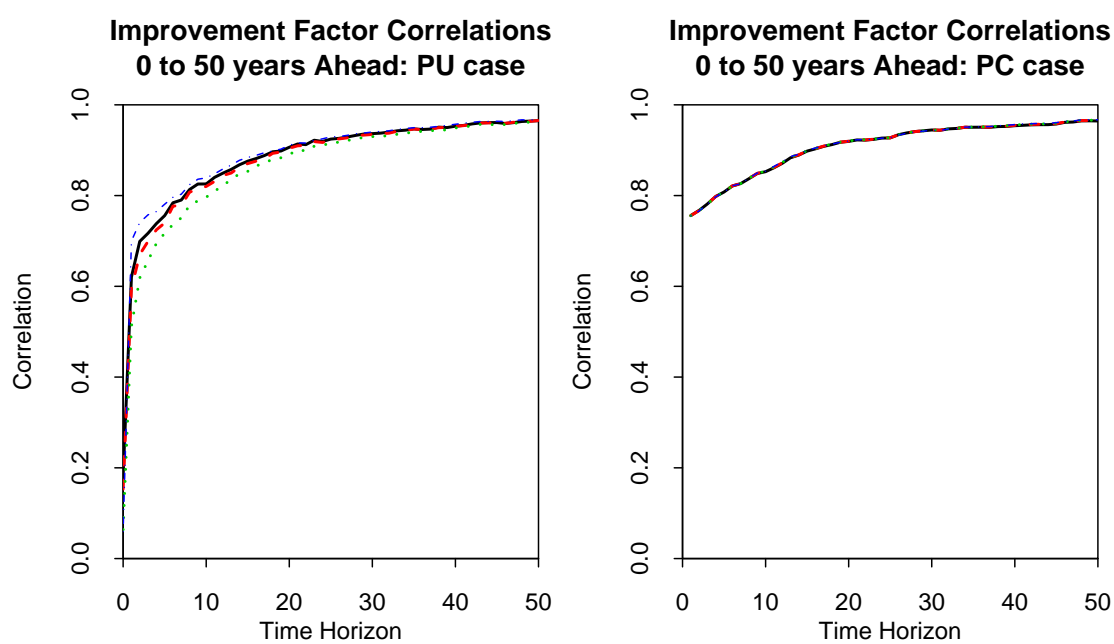
Figure 26: LC model, EW/CMI males. Correlation between simulated improvement factors as a function of the time horizon – age 65 (solid black line), 75 (dashed red line), and 85 (dotted green line) – and correlation between period effects $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$ as a function of the time horizon (blue dot-dashed line). Left: simulations incorporating parameter uncertainty (PU). Right: simulations with no parameter uncertainty (PC).

12 Analysis 7: US/California Males, Ages 60-84, Years 1980-2003, Lee-Carter Model

Results for the application of the two-population MCMC LC model to the US/California males datasets are shown in Figures 28 to 32.

- Mortality is lower in California compared to the US (Figure 30), and divergence of rates can be observed (bottom left) at age 84 if the two populations are modelled in isolation.

- Compared with the equivalent EW analysis, we see that fan charts for US mortality (Figures 30 and 31) are rather narrower, reflecting lower volatility in the historical period effects in the US and California, $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$. A tentative explanation for this is the the greater geographical diversity of the US leading to a dampening of systematic annual fluctuations caused by environmental factors.

  There is a noticeable difference between the parameter certain forecasts for California using the one- and two-population models (Figure 31) which results largely from differences in the estimates of $\beta_x^{(22)}$ (recall that this parameter dictates both the central improvement rate at specific ages and also the level of uncertainty in the LC model).

  Figure 31 also suggests that parameter uncertainty has a more significant role to play in forecasts, particularly in the μ_{R2} parameter, and this might reflect the shorter dataset used compared to the EW/CMI dataset.

- Correlations between the simulated improvement factors are plotted in Figure 32. These are lower than the equivalent correlations for EW/CMI and merely reflect specific parameters estimated in this case ($V^{(2))}$, ψ_{S2}). The PC and PU cases are reasonably close suggesting that the large population sizes mean that Poisson noise is not a major factor.
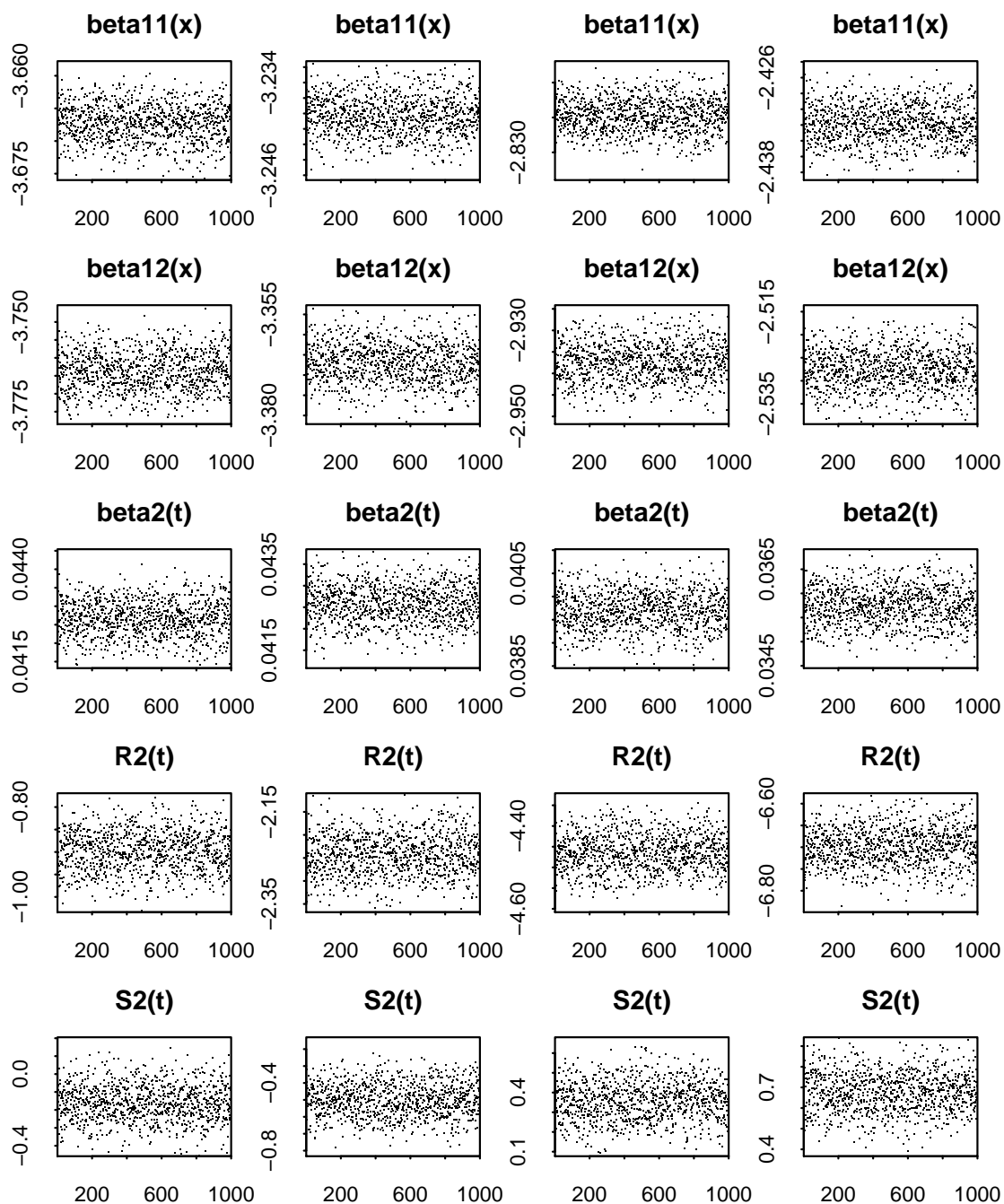
Figure 27: LC model, US/California males. Metropolis-Hastings algorithm output. Age, period and cohort parameters. 1000 values have been stored out of a Markov chain of length 50,000. Age effects are plotted for ages x =64, 69, 74, 79. Period effects are plotted for years t =1983, 1988, 1993, 1998.
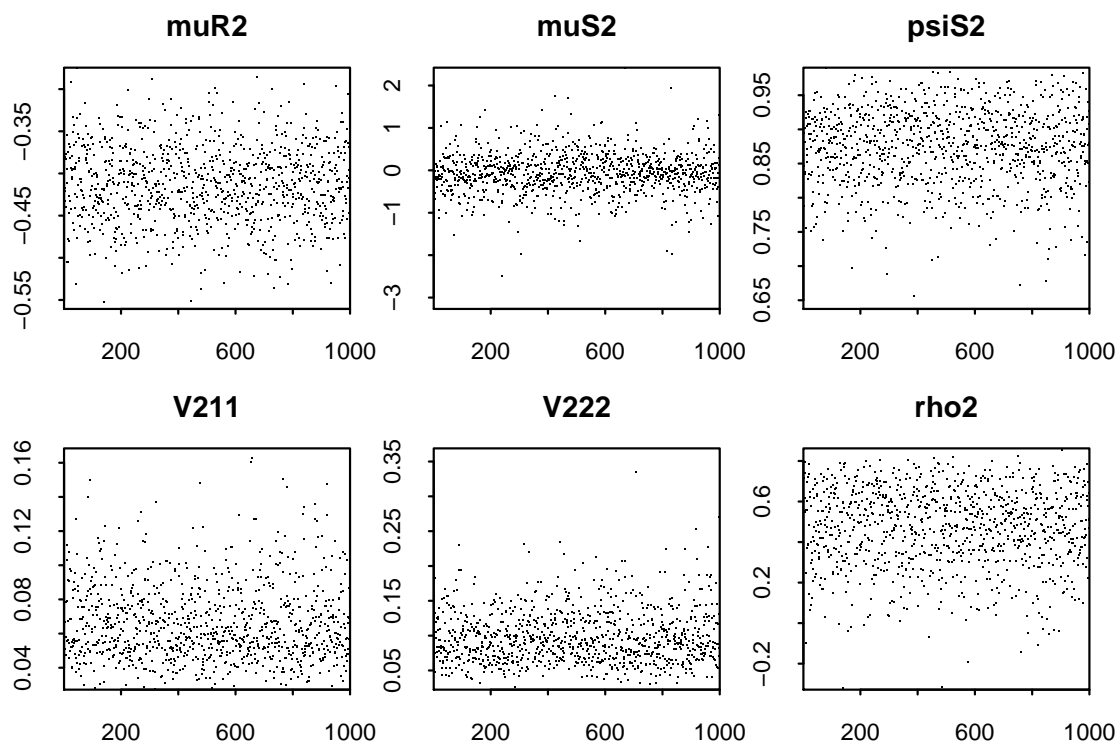
Figure 28: LC model, US/California males. Metropolis-Hastings algorithm output: process parameters. 1000 values have been stored out of a Markov chain of length 50,000.
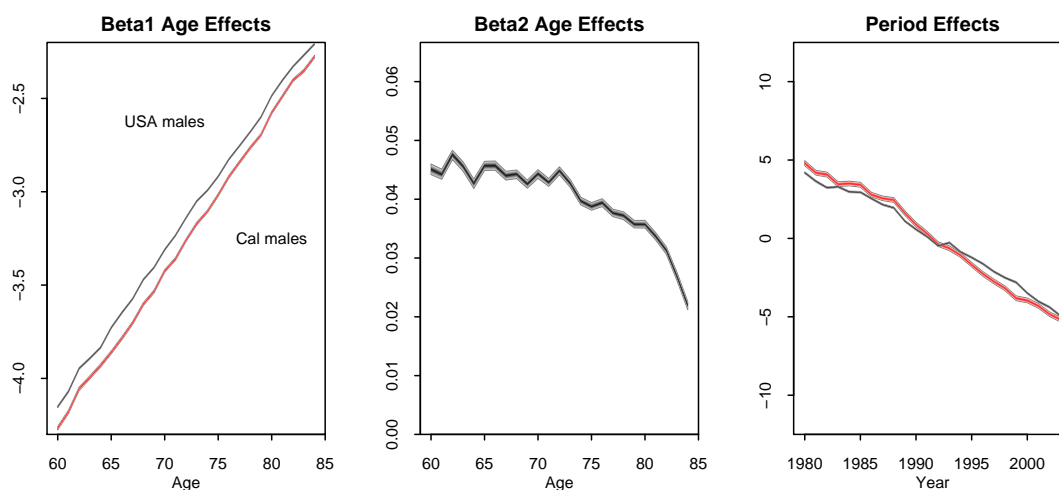

Figure 29: LC model. Age, period and cohort effects for US (grey fans) and California males (red fans). Age, period and cohort effects have been adjusted to satisfy identifiability constraints: $\sum_t R_2(t) = 0$, $\sum_t S_2(t) = 0$.
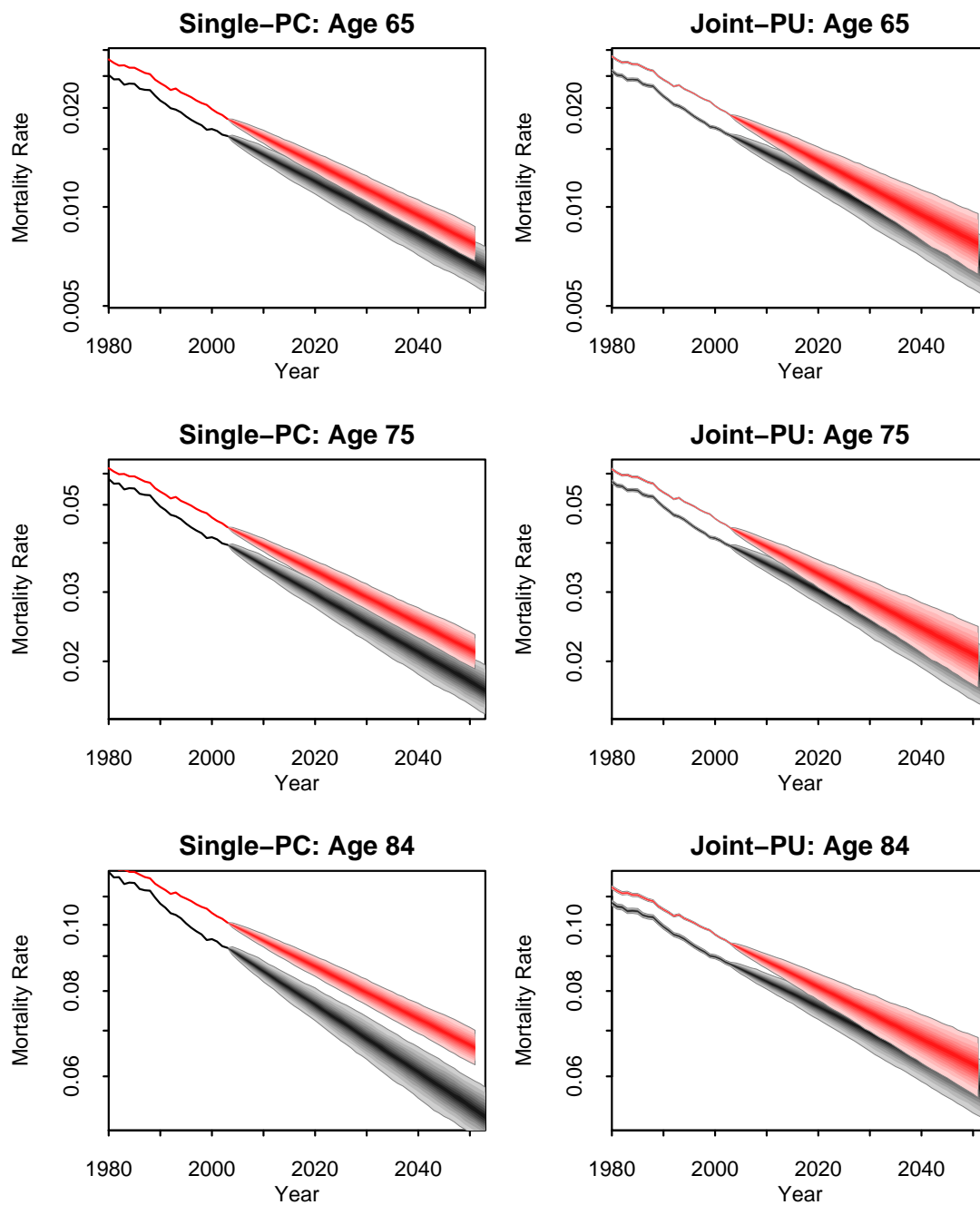
Figure 30: LC model. Mortality fan charts for ages 65, 75 and 85 for US males (red fans) and California males (grey fans). Left-hand plots: fan charts constructed using the single-population models with no parameter uncertainty (PC). Right-hand plots: fan charts constructed using the joint-population model (MCMC) with parameter uncertainty (PU).
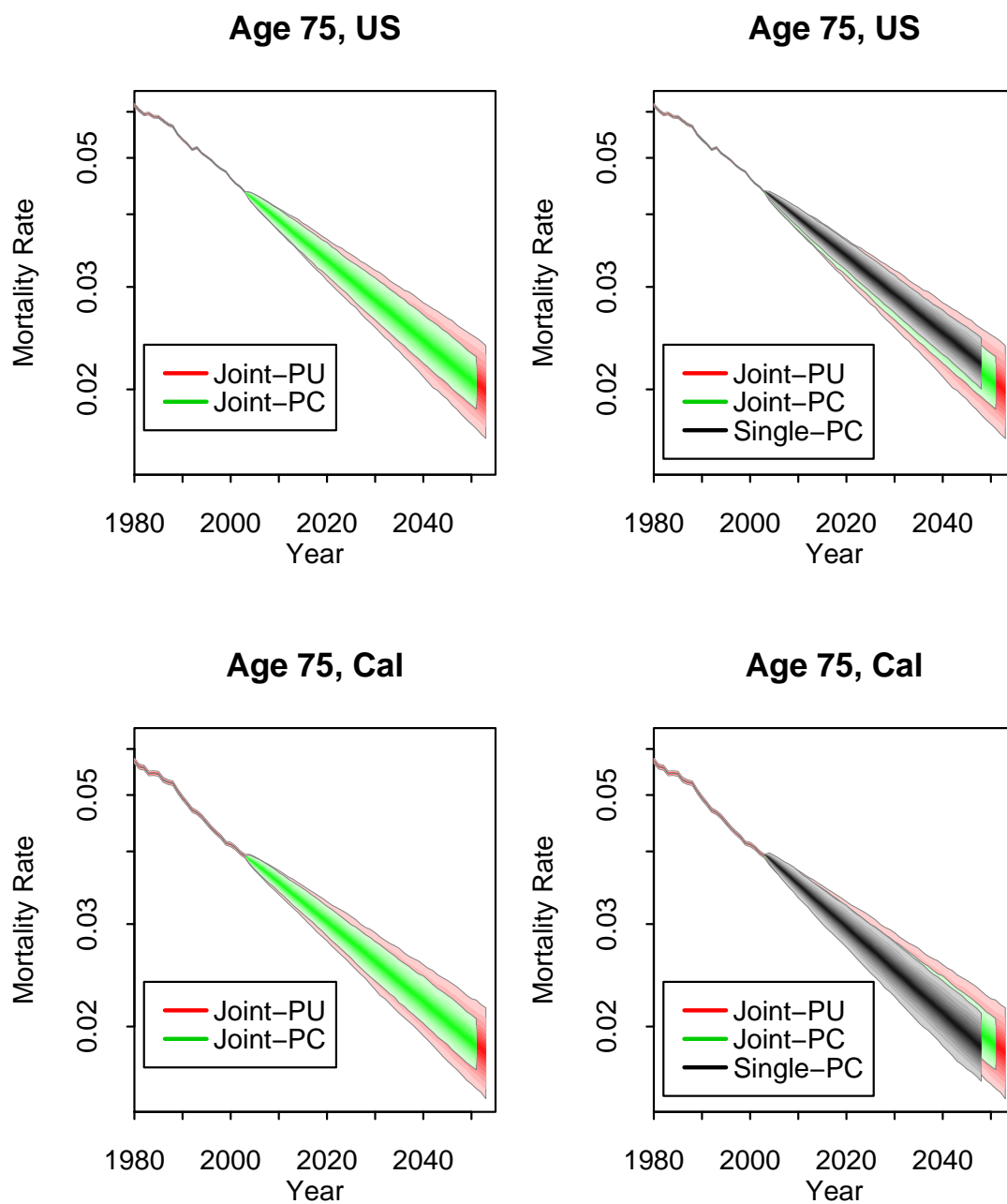
Figure 31: LC model, US/California males. Comparison of fan charts based on (a) the new MCMC algorithm (red fans) and (b) the original two-stage method (grey fans).
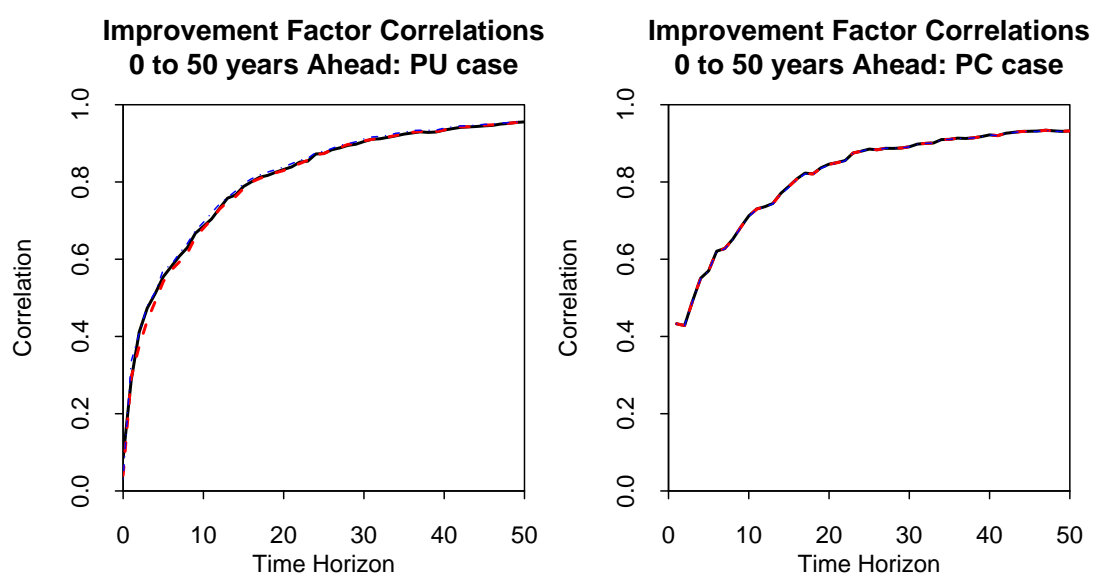
Figure 32: LC model, US/California males. Correlation between simulated improvement factors as a function of the time horizon – age 65 (solid black line) and 80 (dashed red line) – and correlation between period effects $\kappa_t^{(21)}$ and $\kappa_t^{(22)}$ as a function of the time horizon (blue dot-dashed line). Left: simulations incorporating parameter uncertainty (PU). Right: simulations with no parameter uncertainty (PC).

13 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 mortality rates (i.e., the basis) 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 mortality 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 rates 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 7, 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 (green) fans in the two-population case is actually less for both populations than the (grey) fans in the single-population case. Fourth, the correlations between the estimated mortality improvement factors for two populations are consistent with the historical data and can actually be higher when enhanced priors are used, although this is not a universal finding.

However, the approach remains sensitive to the underlying stochastic mortality models used and will compound any weaknesses in these models. For example, the single-factor Lee-Carter model produces fan chart projections of mortality rates that are "too narrow", given the volatility of historical mortality rates (as we found in a previous study, Cairns et al. (2008b)). The approach is also very sensitive to the model used to estimate the cohort effect. The results (e.g., correlations between mortality rates at different ages) are also sensitive to whether a cohort effect is present. 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 is some way. 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 mod-

elling related populations, particularly the basis and 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.

Acknowledgments

References

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., 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., 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., and Dowd, K. (2008a) 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., and Khalaf-Allah, M. (2008b)

Mortality density forecasts: An analysis of six stochastic mortality models. Pensions Institute Discussion Paper PI-0801.

Coughlan, G.D., Emery, S., and Kolb, J. (2004). HEAT (Hedge effectiveness analysis toolkit): A consistent framework for assessing hedge effectiveness. Journal of Derivatives Accounting, 1(2): 221-272.

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.

Coughlan, G., Epstein, D., Sinha, A., and Honig, P. (2007b) q-Forwards: Derivatives for transferring longevity and mortality risk. Available at www.lifemetrics.com.

Coughlan, G.D. (2009a). 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., Cairns, A.J.G., Blake, D. and Dowd, K., (2009b) Longevity hedging: A framework for basis risk and hedge effectiveness. Forthcoming working paper. Presented at the Fifth International Longevity Risk and Capital Markets Solutions Conference, New York, September 2009.

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. (2009) Mixed dynamic and static risk minimization with an application to survivor swaps. 19th International AFIR Colloquium, Munich, September 2009.

Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., Epstein, D. and Khalaf-Allah, M. (2008a). Evaluating the goodness of fit of stochastic mortality models. Pensions Institute Discussion Paper PI-0802.

Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., Epstein, D. and Khalaf-Allah, M. (2008b). Backtesting stochastic mortality models: An ex-post evaluation of multi-period-ahead density forecasts. Pensions Institute Discussion Paper PI-0803.

Gilks, W.R., Richardson, S., and Spiegelhalter, S.J. (1996) *Markov chain Monte Carlo in Practice.* Chapman and Hall, New York.

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. (2009) Modelling adult mortality in small populations: The SAINT model. Pensions Institute Discussion Paper PI-0902.

Kogure, A., Kurachi, Y., and Kitsukawa, K. (2009) A Bayesian evaluation of longevity risk: Model comparison, measuring and pricing. Working paper, Keio University.

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.

Loeys, J., Panigirtzoglou, N., Ribeiro, R.M. (2007) Longevity: A market in the making. Available at 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.

Oeppen, J., and Vaupel, J.W. (2002) Broken limits to life expectancy. *Science*, 296: 1029-1030.

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: 530550.

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.

Richards, S.J. (2008) Applying survival models to pensioner mortality data. To appear in *British Actuarial Journal*.

Tuljapurkar, S., Li, N., and Boe, C. (2000) A universal pattern of mortality change in G7 countries. *Nature* 405: 789-792.

A   Glossary of notation

β_x^{11}:   population 1 age effect

β_x^{12}:   population 2 age effect

$\kappa_t^{(21)}$:   population 1 period effect

$\kappa_t^{(22)}$:   population 2 period effect

$\gamma_c^{(31)}$:   population 1 cohort effect

$\gamma_c^{(32)}$:   population 2 cohort effect

$R_2(t)$:   central period effect

$S_2(t)$:   spread between population 1 and 2 period effects

$R_3(c)$:   central cohort effect

$S_3(c)$:   spread between population 1 and 2 cohort effects

μ_{R2}:   drift of $R_2(t)$

μ_{S2}:   mean reversion level for $S_2(t)$

ψ_{S2}:   AR(1) autoregressive parameter for $S_2(t)$

$V^{(2)}$:   1-step-ahead covariance matrix for $(R_2(t), S_2(t))$

(μ_{R3}, δ_{R3}):   level and trend of $R_3(c)$

μ_{S3}:   mean reversion level for $S_3(c)$

$\phi_{R31}, \phi_{R32}, \phi_{S31}, \phi_{S32}$:   AR(2) autoregressive parameters for $R_3(c)$ and $S_3(c)$