DISCUSSION PAPER PI-1301

A General Procedure for Constructing
Mortality Models

Andrew Hunt and David Blake

February 2013

A General Procedure for Constructing Mortality Models

Andrew Hunt and David Blake*

Cass Business School, City University London

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

February 2013

Abstract

Recently, a large number of new mortality models have been proposed to analyse historic mortality rates and project them into the future. Many of these suffer from being over-parametrised or have terms added in an ad hoc manner which cannot be justified in terms of demographic significance. In addition, poor specification of a model can lead to period effects in the data being wrongly attributed to cohort effects which results in the model making implausible projections. We present a general procedure for constructing mortality models using a combination of a toolkit of functions and expert judgement. By following the general procedure, it is possible to identify sequentially every significant demographic feature in the data and give it a parametric structural form. We demonstrate using UK mortality data that the general procedure produces a relatively parsimonious model that nevertheless has a good fit to the data.

Keywords: Mortality modelling, age-period-cohort models, age period effects, cohort effects

1 Introduction

In recent years, there has been an explosion in the number of new mortality models that have been proposed. This has been triggered, in part, by the greater focus placed on longevity risk by demographers, actuaries and governments. It has also been prompted by the failure of existing models to identify adequately the full extent of the complexities involved in the evolution of mortality rates over time.

Yet these new models often involve ad hoc extensions to existing models that have questionable demographic significance. They also have difficulties providing realistic forecasts of specific mortality rates. Despite having more terms than the older models, they still fail to capture a lot of the information present in the data. Lacking a formal procedure for interrogating the data in order to establish what structure remains to be explained, modellers too often add new terms based on theoretical models or assumptions regarding the shape of the mortality curve rather than evidence. This is especially dangerous in models with cohort parameters intended to capture generational effects. The result of any mis-specification in these extra age/period terms can result in structure being wrongly attributed to the cohort effect. This is then projected incorrectly, moving up the age range with the passage of time, with the result that implausible forecasts are generated at higher ages.

In view of this, we felt that the time has come to take a fresh look at mortality model construction. But, rather than propose yet another new model, what we do in this paper is outline and implement a "general procedure" (GP) for building a mortality model from scratch, driven by a forensic examination of the data. Through an iterative process, the GP identifies every significant demographic feature in the data in a sequence, beginning with the most important. For each demographic feature, we need to apply expert judgement to choose a particular parametric form to represent it. To do this, we need a "toolkit" of suitable functions.

By following the GP, it is possible to construct mortality models with sufficient terms to capture accurately all the significant information present in the age, period and cohort dimensions of the data. In particular, the GP prevents structure in the data which is genuinely associated with an age/period effect being wrongly allocated to a cohort effect. The procedure is general in

the sense that it can be applied to any dataset to give a fully specified model tailored to the features of the population under consideration. Most significantly, the GP provides evidence for the addition of each term to an existing model: it allows each new term to be associated with a specific demographic and biological process driving the evolution of mortality rates.

Section 2 presents a summary of the structure of the class of mortality models we are considering and sets out the desirable properties that we believe a good mortality model should possess. The general procedure is discussed in Section 3. In Section 4, we apply the GP to data for men in the UK and describe how the steps in Section 3 operate in practice. In Section 5, we assess the goodness of fit of this model and check whether there is any remaining structure present in the fitted residuals. Section 6 compares the GP with the Lee-Carter model and with a procedure based on principal component analysis as an alternative method of constructing mortality models with multiple age/period terms. Finally, Section 7 concludes with an assessment of how the final model found measures up against our set of desirable properties from Section 2 as well as its advantages and disadvantages.

2   The structural form of mortality models

The majority of existing mortality models proposed in the actuarial literature fall into an age/period/cohort framework. This transforms the observed mortality rates and then fits a series of terms to account for the interactions between the age, the year of observation and the year of birth for the population within each cell of data. Mathematically, this can be written as:[1]

$$\eta \left(E \left(\frac{D_{x,t}}{E_{x,t}} \right) \right) = \alpha_x + \sum_{i=1}^{N} f^{(i)}(x;\theta_i)\kappa_t^{(i)} + \gamma_{t-x} \tag{1}$$

This equation has the following components:

- a link function η to transform the observed data into a form suitable for modelling. The raw data usually consists of death counts $D_{x,t}$ and exposures to risk $E_{x,t}$ at ages x and for years t;

---

[1]This structural form and the issues associated with identifying the terms in it are discussed in depth in Hunt and Blake (2013)

- a static life table α_x to capture the general shape of the mortality curve that does not change with time;

- N companion pairs of age/period terms $f^{(i)}(x;\theta_i)\kappa_t^{(i)}$, consisting of period terms $\kappa_t^{(i)}$ (or "trends") which give the evolution of mortality rates through time and age functions $f^{(i)}(x;\theta_i)$ which determine which segments of the age range these trends affect; and

- cohort parameters γ_{t-x} which determine the lifelong effects that are specific to different generations as discussed in Willets (2004), denoted by their year of birth;

Mortality models in this form include the Lee-Carter model proposed in Lee and Carter (1992) and extensions of this, such as those of Renshaw and Haberman (2003) and Yang et al. (2010); the age/period/cohort model of Hobcraft et al. (1982); the Cairns-Blake-Dowd family of mortality models (in Cairns et al. (2006a) and Cairns et al. (2009)); developments of this such as the models proposed by Plat (2009) and O'Hare and Li (2012); and various other mortality models not contained within these families such as the ones proposed in Wilmoth (1990) and Aro and Pennanen (2011). The models of the rate of mortality change proposed in Haberman and Renshaw (2012) and Mitchell et al. (2013) also fall within this structure for suitable choice of the link function $\eta_{x,t}$. These models and the relationships between them are discussed in greater depth in Hunt and Blake (2013). Models which fall outside this framework include those with a term which is constany across ages (such as the static Makeham model of mortality), the extension to the LC model proposed in Renshaw and Haberman (2006) (due to the presence of the $\beta_x^{(0)}$ term modifying the cohort parameters) and the P-splines models of Currie et al. (2004).

A good mortality model should satisfy the following "desirability criteria":

1. provide an adequate fit to the data, with sufficient terms to capture all the significant structure in the data;

2. be demographically significant in the sense that each age function can a) be identifiable with specific biological and socio-economic processes occurring at the ages of interest, and b) be biologically reasonable (as discussed in Cairns et al. (2006b));

3. be parsimonious, with the smallest number of terms needed to capture this structure, and with each term using as few parameters as possible;

4. be robust, in that parameter uncertainty should be low and small changes in the data should not result in significant changes in the estimates of the parameters and in our interpretation of them;

5. span the full age range, with sufficient terms to model the complex shape of and dynamics observed in mortality rates at younger ages; and

6. include cohort effects if justified by the data and allow for these to be clearly distinguished from age/period effects to allow plausible projections of the model.

The GP has been designed with these criteria (and the trade-offs between them) in mind. Most specifically, the GP chooses parametric age functions $(f^{(i)}(x; \theta_i))$ which take a specific functional form and are parametrised by a small number of variables θ_i, over more general non-parametric age functions, $\beta_x^{(i)}$, due to their parsimony and because we can use our judgement to assign demographic significance to the term in question. However, a key feature of the GP is to use the information discovered from first using a non-parametric age function to provide guidance on the shape of that demographic feature. This will improve the goodness of fit for each term and avoid the need to make a priori assumptions regarding which age functions to use.

3 A general procedure for constructing mortality models

The general procedure consists of the following steps:

1. start with a static life table α_x to capture the time-independent shape of the mortality curve across ages in the data set under consideration;

2. add a companion pair of non-parametric age and period functions $\beta_x \kappa_t$ to find the most significant age/period effect not captured by the model so far, where the age term β_x is free to take the shape that maximises the fit to the data;

3. observe the shape of the estimated age term β_x across ages and how κ_t has evolved through time;

4. check that the addition of the new pair of terms improves the overall goodness of fit to the data;

5. use judgement to select a specific smooth functional form $f(x; \theta)$ to replace the non-parametric age term β_x where the function is defined by a small number of free parameters θ;

6. check whether the fitted model with this specific functional form a) produces a similar evolution over time as the non-parametric term by comparing the fitted κ_t's for the two cases and b) achieves comparable improvements in the goodness of fit as the non-parametric term;

7. check whether the addition of the new companion pair of terms has significantly changed the shape of previously selected terms, in which case we might need to change and re-estimate the earlier terms;

8. repeat steps 2 to 7 until we are satisfied that the model captures all significant age and period structure in the data;

9. add a cohort term γ_{t-x} to capture any year-of-birth effects;

10. test the final model for goodness of fit and robustness, and the residuals for the properties of normality and independence, thereby confirming that there is no significant unexplained demographic structure remaining in the data;

11. compare the final model to alternative models estimated using the same data set.

Figure 1 shows a flow chart of the GP summarising these steps.

The GP is a data-driven procedure, with terms being selected based on their ability to capture features of the observed mortality rates. At high level, it is a specific-to-general model building procedure (as defined in Campos et al. (2005)) as it begins with a simple model and sequentially adds terms in order to build a model that fully reflects the features contained in the dataset under investigation. This approach is unavoidable, as to begin with a fully general mortality model, as required by the general-to-specific

Figure 1: Flow chart of the general procedure
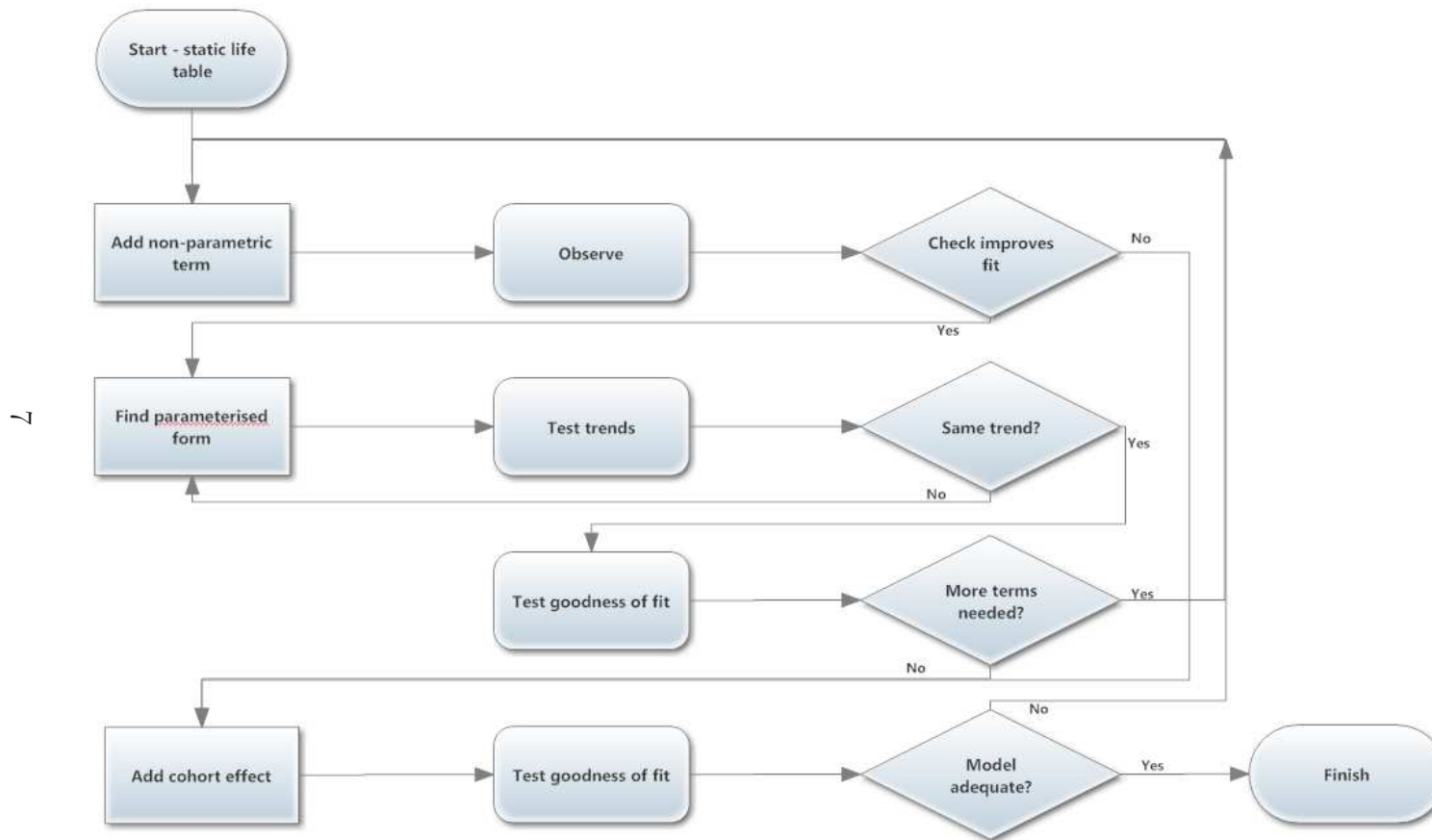
methodology, would contain such a large number of terms that it would be impossible to fit it to data and difficult to simplify. However, at the "micro" level, each age/period companion pair is added in a general-to-specific fashion - the most general form of the function is added to the model and then simplified into a specific, parametric form, whilst seeking to retain its explanatory power. Thus, we believe that the GP benefits from both model-building frameworks.

The GP selects the functional form of the age/period terms in two stages. First, it allows each age/period term within the data to be identified by a non-parametric age function without requiring any a priori assumptions to be made by the modeller. Second, it allows the shape of these non-parametric age functions to guide the choice of parametric function that is selected from the toolkit to match as closely as possible the explanatory power of the former, whilst benefiting from parsimony in terms of the number of parameters to be estimated. However, judgement is required in the selection of the parametric function, although that the GP provides evidence to justify the decision made.

Appendix A gives details of the "toolkit" of parametric age functions needed to implement the GP; it also gives a general algorithm for estimating the free parameters in them. However, a toolkit is never complete and so we do not offer this as an exhaustive list of functions - only as those we have considered so far. Two highly desirable features for a function to be included in the toolkit are a small number of free parameters (in our experience, more than two free parameters leads to unstable estimates) and the ability to adjust the location of the function in the age range.

At each stage of the GP, we need to assess whether the resulting model is in accordance with our desirability criteria. First, we will need to test whether an additional age function improves the fit of the model to data. It is well known that a measure such as the log-likelihood will always show an improvement in the fit of a series of nested models to the data due to the increased number of free parameters. In order to achieve our desire for a parsimonious model, it is therefore necessary to penalise the number of free parameters used by considering a measure such as the Bayes Informa-

tion Criterion (BIC).[2] The log-likelihood is still useful, however, when adding an additional non-parametric term as the change in this measure represents the maximum possible improvement in the fit from the addition of a single new term. We can therefore use this maximum possible improvement as the benchmark for measuring the success of the specific parametric form being trialled: a parametric age function which produces 80-90% of the same improvement in log-likelihood can be regarded as highly desirable.

Second, we need to compare whether the structure identified by a non-parametric age function is the same as that found when a specific parametric function is introduced. Plots of the two are useful for revealing the general pattern of mortality change and identifying features such as trend changes and outliers that the two series have in common.

Finally, we will need to test the residuals from the data. As discussed in Pitacco et al. (2009), under a Poisson model for deaths (such as the one we use), the standardised deviance residuals $r_{x,t}$ are given by

$$r_{x,t} = sign(d_{x,t} - \hat{d}_{x,t}) \sqrt{\frac{2W_{x,t}}{\phi} \left(d_{x,t} ln \left(\frac{d_{x,t}}{\hat{d}_{x,t}} \right) - (d_{x,t} - \hat{d}_{x,t}) \right)}$$

with actual death count $d_{x,t}$, fitted death count $\hat{d}_{x,t} = E^c_{x,t} \mu_{x,t}$, and ϕ the scale parameter given by the total fitted deviance divided by the number of degrees of freedom[3] of the model. This assumes that the residuals have constant variance across age and time. For large expected death counts, these should be approximately standard normal variables, so we can test the residuals for normality using the Jarque-Bera test of the skewness and kurtosis to check this. The residuals should also be independent and show no obvious structure across ages, periods and cohorts. To look for structure within the residuals, we plot heat maps and visually inspect for obvious vertical, horizontal or diagonal banding patterns. This would indicate the presence of further age, period or cohort effects. We also calculate the correlations of the residuals with their neighbours in the age and period directions, and test these correlations against the assumption of independence.

---

[2]Defined as $max(L) - 0.5 \times$ No. free parameters $\times ln$(No. data points).
[3]Number of data points less number of free parameters.

9

To exit the cycle of adding new age/period terms, we need a stopping rule in the GP to determine when there are no further demographically significant age/period terms left unidentified in the data. Such a stopping rule will inevitably be subjective. This means that the GP is not a "black-box" algorithm - it requires the active engagement and exercise of judgement by the modeller at each stage of the model building process.

Finally, we add the cohort parameters as the last step in the GP. The reason for this reflects a preference for a model where the majority of the temporal dependence in the data is allocated to the age/period terms. In our experience, the pattern of fitted cohort parameters produced by some models does not seem to have any demographic significance and may be caused by the model trying to compensate for inadequate age/period terms. We therefore seek to avoid this in the GP.

4   Application of procedure to male UK data

To illustrate the GP, we apply it to data for men in the UK from 1950 to 2009 covering ages 0 to 100 (ungrouped) downloaded from the Human Mortality Database (Human Mortality Database (2012)). We restrict the data to the period since the Second World War as it is free from major conflicts and abrupt social upheaval. As the Human Mortality Database provides central exposures to risk for each age and year, we assume that the death counts are Poisson random variables and therefore use a log-link function for $\eta_{x,t}$ as it is the canonical link function for the Poisson distribution. We fit the model at each stage using Poisson maximum likelihood estimation using the algorithms described in Appendix A.

4.1   Stage 0 - Static life table

The static life table produced by fitting $ln(\mu_{x,t}) = \alpha_x$ constitutes the first step in the GP. The fitted values of α_x (not shown) show the usual pattern of mortality across the full age range: with high mortality rates at age zero due to infant mortality, the log-linear pattern of mortality increases at high ages (from 50 to 90) and the increased rates of mortality due to the accident hump between ages 15 and 25. Whilst the age function is refitted at each

stage of the GP, this shape does not change significantly throughout the different stages of the model building process.
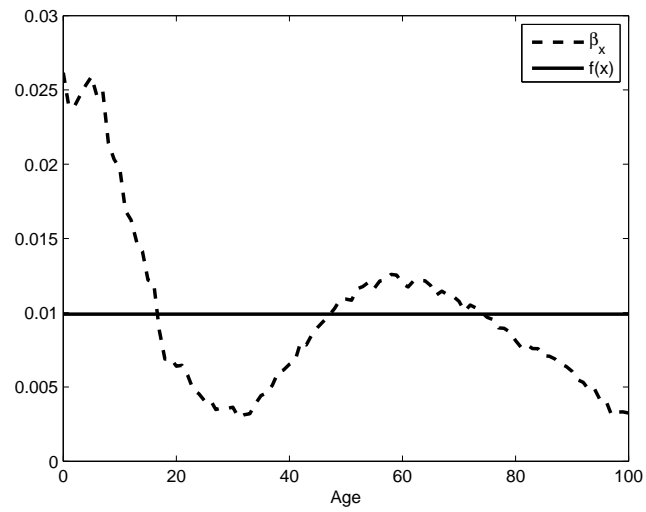
4.2 Stage 1 - First age/period term

The next step is to add the first non-parametric age/period term to the static model to arrive at $ln(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t^{(1)}$, which has the form of the Lee and Carter (1992) model. This gives the familiar β_x and $\kappa_t^{(1)}$ terms shown in Figure 2.

In the interests of parsimony and demographic significance, we believe that it is highly desirable to find a simpler parametric form than the age function of the Lee-Carter model to capture the impact of the dominant trend within the data - ideally the simplest age function that will capture the same trend. This parametric form should be continuous to avoid any issues with the smoothness of projected mortality rates. As the fitted β_x age function is positive across the whole age range, it might be felt to represent a general improvement in mortality rates across all ages. Appealing to this demographic significance, we therefore try the simplest possible age function - a constant. As Figure 2 shows, this simple age function effectively captures the same trend as the non-parametric β_x function with 100 fewer parameter, and achieves approximately 92% of the same improvement in log-likelihood. We are therefore satisfied that there is no need to use a more complex and less parsimonious age function, although we would expect that much of the age structure present in the fitted β_x will need to be captured by subsequent age/period terms.

Figures 2a and 2b shows the age and period functions generated by Stage 1 of the GP. We can see that the population has experienced sustained improvements in mortality which have accelerated slightly in recent years. The model also detects the increased mortality in 1951 owing to the influenza epidemic in that year which affected much of England.

So far, so good, but a plot of the residuals - not shown here - indicates that additional terms are necessary to fully capture all the structure within the data.

11

Figure 2: Age and period functions for Stage 1 of the general procedure
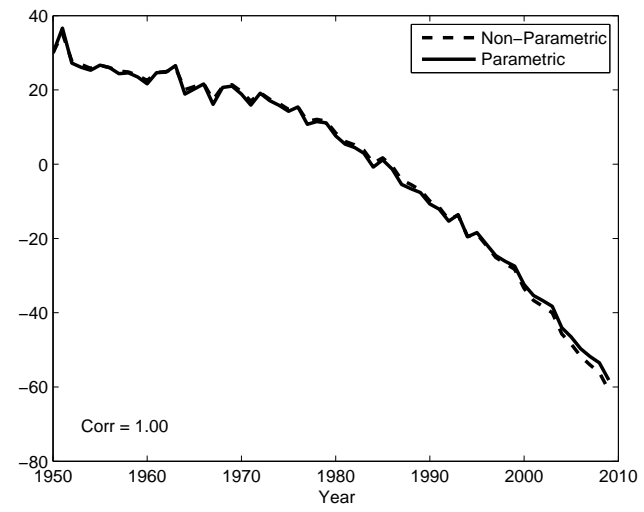
(a) Age functions

(b) Period functions

4.3   Stage 2 - Second age/period term

In order to find the next most significant age/period effect within the data, we now add another non-parametric structure to the model to arrive at

$$ln(\mu_{x,t}) = \alpha_x + f^{(1)}(x)\kappa_t^{(1)} + \beta_x\kappa_t^{(2)} \tag{2}$$

The fitted model gives the values of β_x and $\kappa_t^{(2)}$ shown in Figure 3. It is not a trivial task to select an appropriate parametric age function from the shape of β_x and this is where judgement becomes important. By inspection, the non-parametric age function appears to have two components - an upward-sloping linear trend across the entire age range and a large "hump" superimposed on the age range 10 to 50. Since we can assign different demographic significance to each of these features, it is appropriate that we separate them into two different age/period terms in the fully specified model. However, these trends will probably be highly correlated which is why the non-parametric function has combined them.

We choose to fit a straight line as our choice of $f^{(2)}(x)$ as it is a simpler potential function than one with a hump shape; indeed it is the simplest possible function after a constant. In our experience, a straight line is often the second choice of age function that arises naturally when applying the GP, especially for data restricted to higher ages. This lends support for the use of the Cairns-Blake-Dowd class of models. A straight line can be interpreted as determining changes in the slope parameter in a Gompertz model of mortality for models with a logarithmic link function. This is related to the "rectangularisation" of the mortality curve, as a greater proportion of deaths at high age occur around the median age of death. We also note that $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are negatively correlated, consistent with the Strehler-Mildvan law of mortality discussed in Finkelstein (2012), for instance.
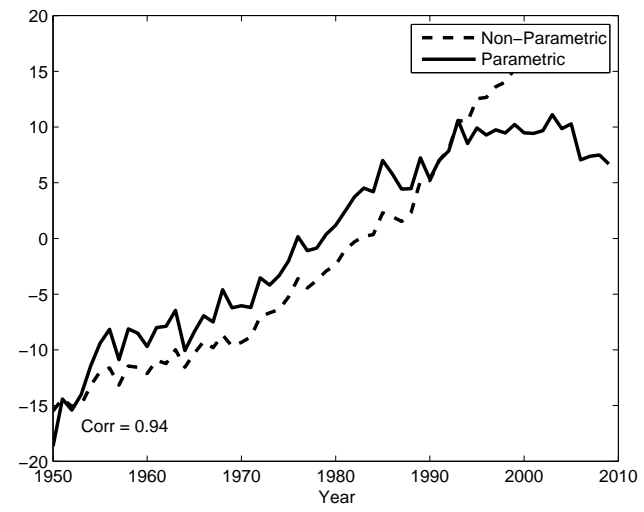
4.4   Stage 3 - Third age/period term

Our discussion of the choice of an appropriate age function at Stage 2 should give us a strong idea as to the appropriate shape of the age function for Stage 3. The GP gives us the evidence to support or reject our conjecture by first

Figure 3: Age and period functions for Stage 2 of the general procedure
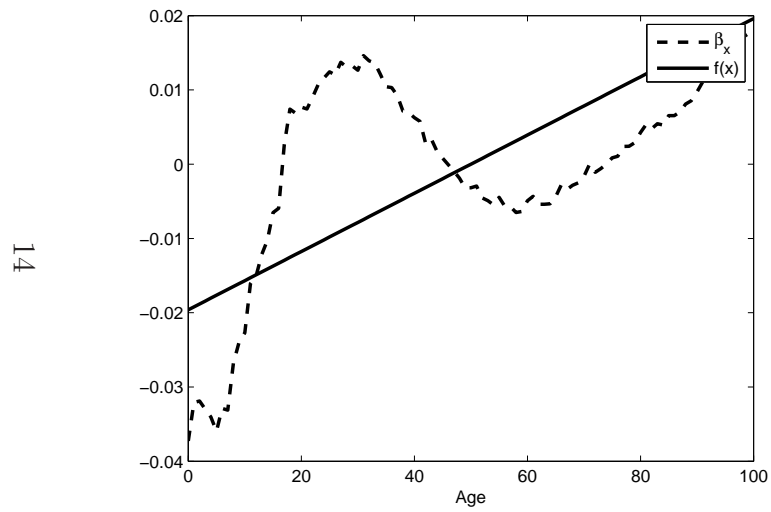
(a) Age functions

(b) Period functions

extending the model with a new non-parametric age/period term

$$ln(\mu_{x,t}) = \alpha_x + \sum_{i=1}^{2} f^{(i)}(x)\kappa_t^{(i)} + \beta_x \kappa_t^{(3)} \tag{3}$$

The fitted non-parametric model gives the values of β_x and $\kappa_t^{(3)}$ shown in Figure 4. This confirms that a suitable choice for $f^{(3)}(x)$ could indeed be some form of hump-shaped function centred around age 25 and so we experiment with

$$f^{(3)}(x) = \frac{1}{\sigma}exp\left(-\frac{(x-\hat{x})^2}{\sigma^2}\right) \tag{4}$$

This function has two free parameters, $\hat{x}$ and σ which, by analogy with the normal distribution, govern the location of the hump and its width. These are estimated using Poisson maximum likelihood estimation. We choose the starting values for these parameters by observing the pattern of the β_x function, before applying our optimisation algorithm. The final, fitted values should not be overly sensitive to the initial choice. If they are, this indicates that the choice of age function may be inappropriate and will cause problems with the model when additional terms are added.

The final fitted $f^{(3)}(x)$ and $\kappa_t^{(3)}$ functions are shown in Figure 4. When adding a new term to the model, we need to check that it does not significantly alter the demographic interpretation of the previous terms. Plots of the first two terms - not presented here - indicate that they have not changed significantly due to the presence of the third term.

Visual inspection of the heat map of residuals in Figure 5 shows us that a) there appear to be additional age/period effects in the data, most obviously centred on age 0 and age 18 and b) there is a clear need for a cohort effect in the model as shown by the prominent diagonal lines on the heat map indicating features which follow individual years of birth as they age. The evidence gleaned from the heat map plot is useful when deciding on subsequent terms, especially when trying to determine if the shape shown by an exploratory $\beta_x\kappa_t$ function is trying to approximate for a cohort effect - something we believe is essential to avoid.

15

Figure 4: Age and period functions for Stage 3 of the general procedure
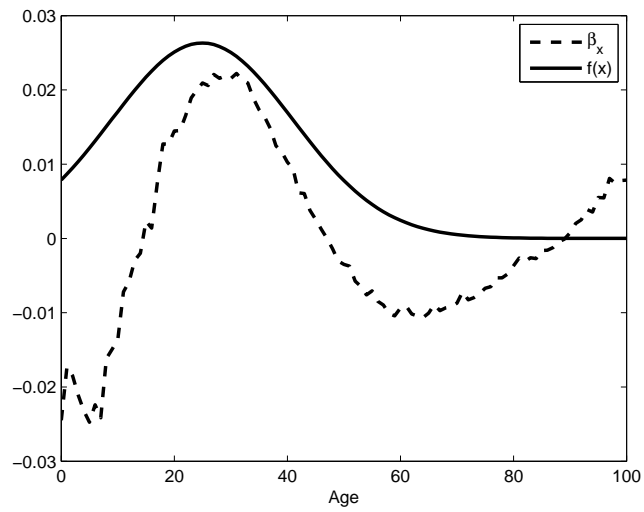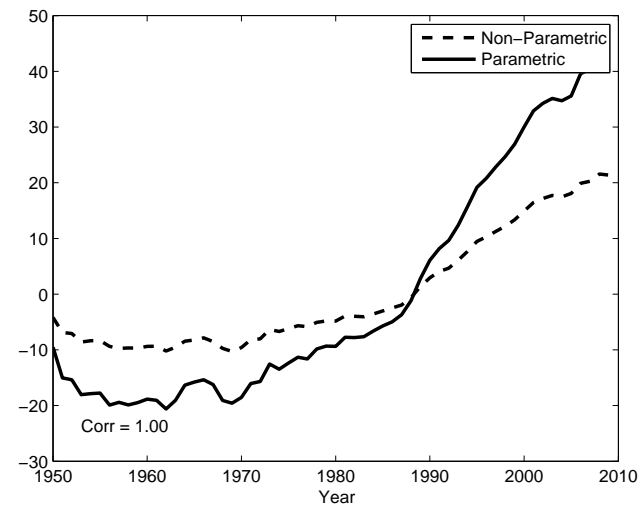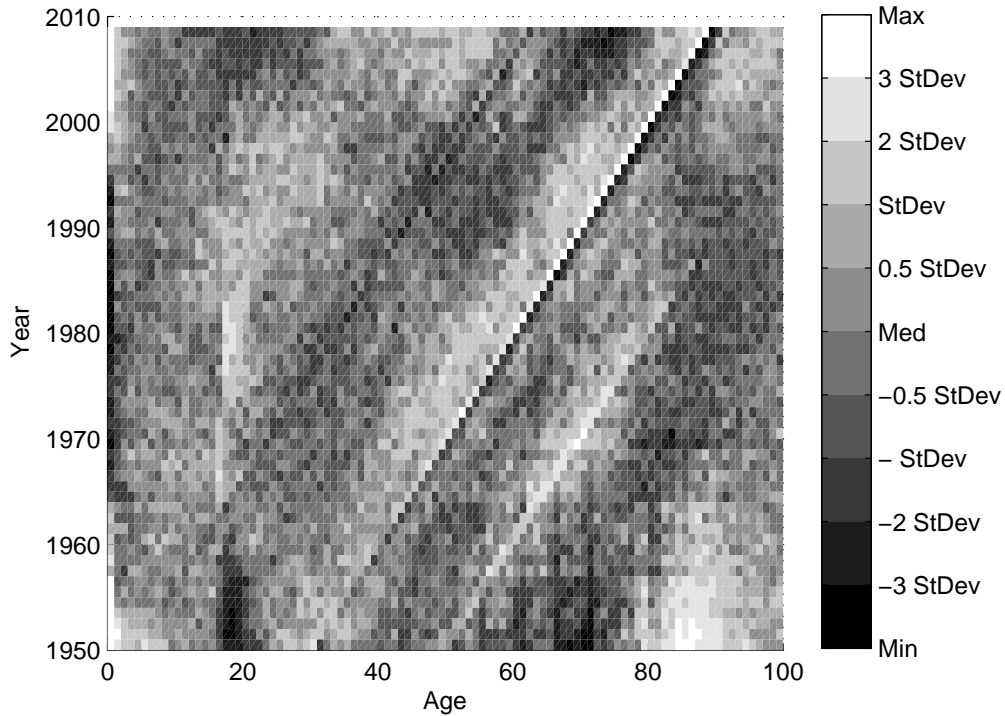

(a) Age functions

(b) Period functions

Figure 5: Heat map of residuals from Stage 3

4.5   Stage 4 onwards - Additional age/period terms

The format of the GP from Stage 4 onwards follows the same pattern as for Stages 1, 2 and 3: choose an appropriate functional form for the age term in order to capture the main effect revealed by the non-parametric $\beta_x \kappa_t$ term.

We have already dipped into our toolkit of age functions, most notably by using the two-parameter Gaussian function at Stage 3. Stage 4 and onwards require us to have a far greater range of functions available in the toolkit that we can potentially use. Appendix A contains a list of the parametric functions considered in this analysis.

Figure 6 shows plots of the final fitted age functions $f^{(i)}(x)$ and trends $\kappa_t^{(i)}$ for $i = 4, 5, 6, 7$. It is useful to note that the order of discovery of these functional forms provides a natural order of importance for the age terms.

The age functions we have fitted are:

- Stage 4: a broken linear function similar to the payoff of a put option, which we can associate with childhood mortality rates;[4]

- Stage 5: a Rayleigh function, which we associate with the postponement of deaths from late middle age to old age that results from medical improvements over the past 60 years;

- Stage 6: a log-normal function centred on ages 18-19 which we associate with the peak age of the accident hump; and

- Stage 7: a normal function centred on ages 55 to 65 which may be associated with the major causes of death in late middle age, such as lung cancer and coronary heart disease and the efforts made to tackle them.

The residual heat map for Stage 7 (Figure 7) is dominated by the diagonal lines representing the cohort effects which have been excluded from the model so far. This might lead us to conclude that we have extracted all of the important age/period effects from the data. This is confirmed by adding a further exploratory non-parametric term to the model. Whilst the resulting BIC for the model does increase, there is little structure to the β_x fitted (shown in Figure 8a) except for the periodic pattern at high ages which is clearly trying to capture a series of cohort effects.[5] We therefore conclude that, for UK male data over the sample period, there are seven distinct age/period effects in the data.

---

[4]This function can be thought of as a very simple linear spline with a single knot, similar to those used as basis functions in Aro and Pennanen (2011). More complex splines could also be considered as part of the toolkit of age functions.

[5]We have tested whether the use of an indicator function at age 18 or a narrow, triangular "spike" function similar to those used in Aro and Pennanen (2011) and centred on this age would improve the goodness of fit. However, when using the BIC which penalises for excessive parametrisation, the use of these functions did not improve the fit of the model. The use of an indicator function also leads to mortality rates at age 18 being fit perfectly which does not accord with our desire for parsimony and may lead to discontinuous mortality rates and this is not biologically reasonable.

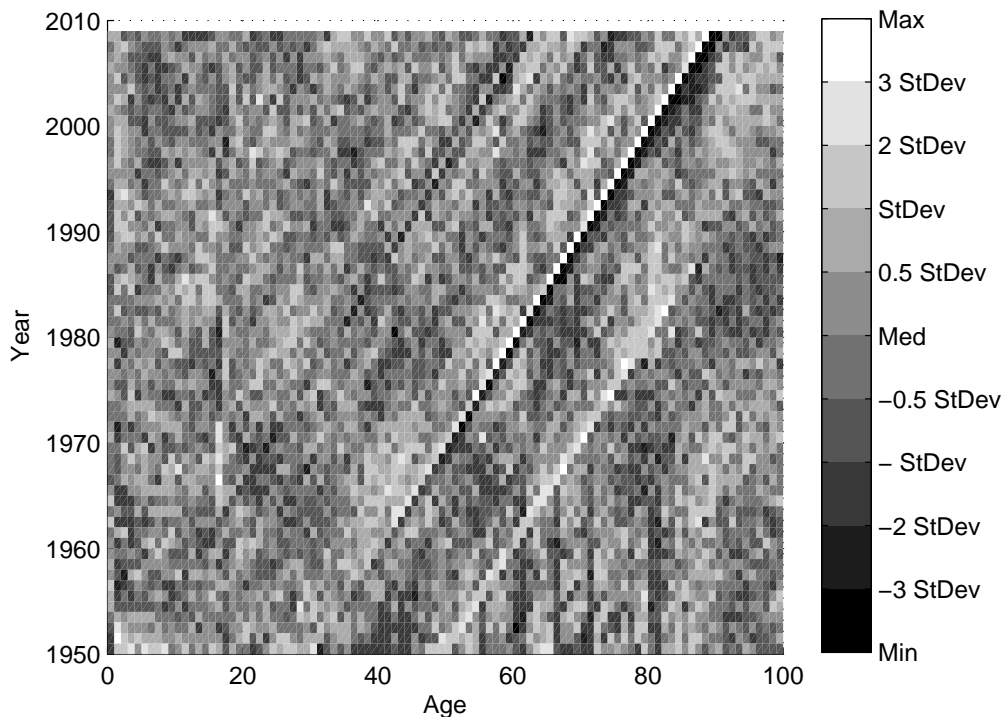Figure 6: Age and period functions for Stages 4 to 7 of the general procedure
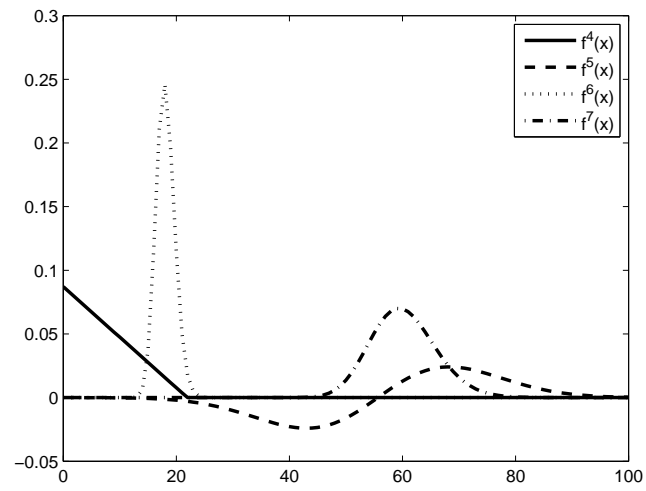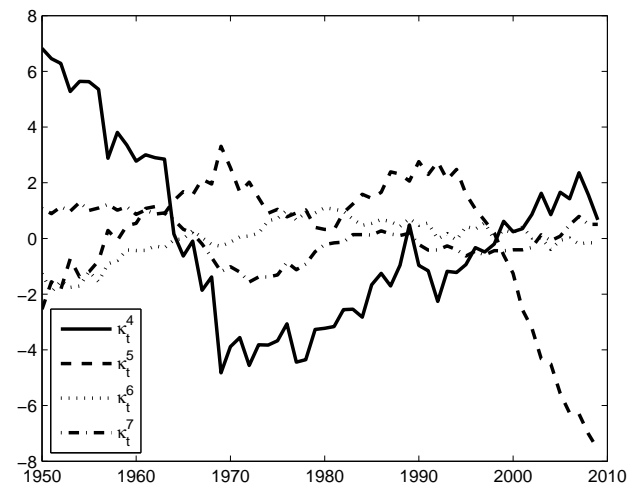
(a) Age functions

(b) Period functions

Figure 7: Heat map of residuals from Stage 7

4.6 Stage 8 - Cohort term

The final stage is to add the cohort parameters γ_{t-x} to yield the final model

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

Due to the limited number of observations on very early and late cohorts, we do not estimate cohort parameters in the first and last ten years of birth. Instead, we linearly interpolate these to zero for smoothness. The final model gives the cohort parameters shown in Figure 9. Adding a cohort term to the model also creates additional issues with the identifiability of the parameters, which are solved by applying extra identifiability constraints. The full set of identifiability constraints required by the final model produced by the GP is given in Appendix A.

20

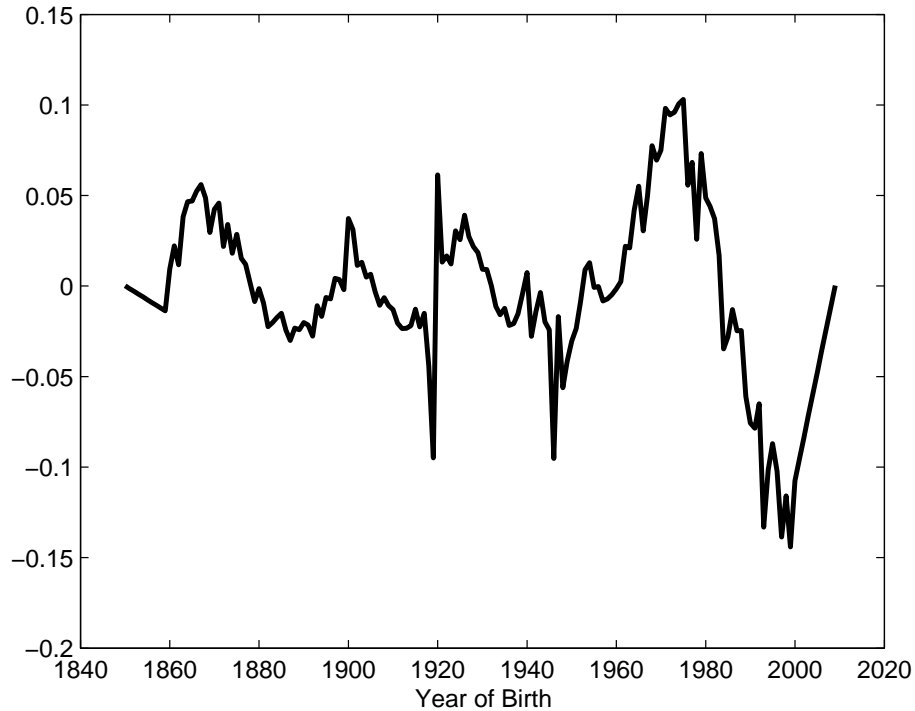Figure 8: Non-parametric age and period functions at the end of Stage 7 of the general procedure
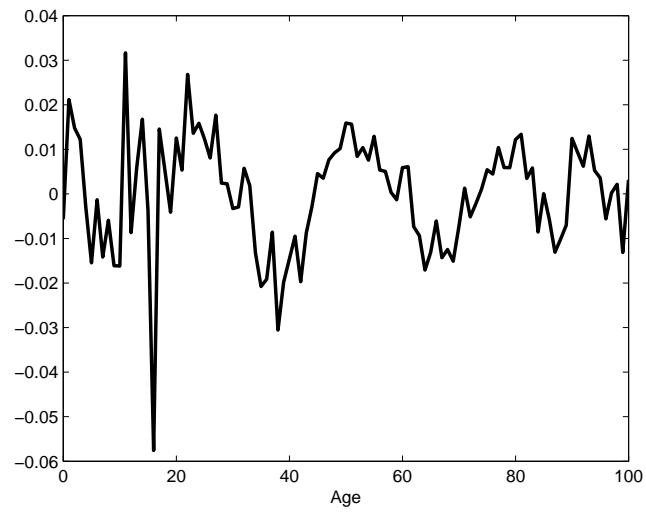
(a) Age function
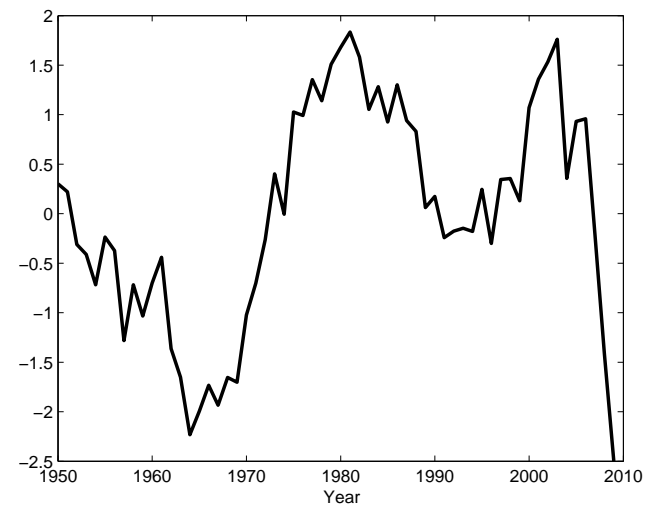
(b) Period function

Figure 9: γ_{t-x} cohort effects from Stage 8 of the general procedure

From this, we can identify the major features of interest and can try to
relate them to the life histories of the affected cohorts. Most obviously, there
is a clear discontinuity between years of birth 1918 and 1919. This may relate
to the impact of the influenza epidemic that year. Alternatively, it could be a
data artefact caused by a flood of births after the First World War distorting
the assumptions used to construct exposures to risk (for a discussion, see
Richards (2008)). Following this is the decline in cohort mortality observed in
Willets (1999, 2004) and discussed in Murphy (2009) relating to the "golden
cohort" of individuals born in the late 1920's and early 1930's. We also
observe a further (although smaller) discontinuity between 1945 and 1946
relating to the end of the Second World War, strengthening the data artefact
argument presented in Richards (2008). We are unsure what demographic
significance the excess cohort mortality observed for years of birth between
1960 and 1980 has. These are individuals currently aged between 30 and

22

50 and therefore we have limited mortality experience data for them and so any attempt at assigning demographic significance is somewhat speculative. However, this feature is robust when adjusting the range of the data for the model and when additional age/period terms are added. This feature will be also be significant for projecting mortality rates if this excess mortality is continued later into life. Finally, we observe a distinct cohort effect for individuals born around the year 1900 (which again is robust to the model and data specification). This may be due to the formative impact of experience during the First World War as young men and the lifetime health effects this may have induced.

5 Testing the final model

Our final model consists of the seven age period terms described in Table 1 plus terms for the static life table α_x and the cohort parameters γ_{t-x}.

Term	Description	$f^{(i)}(x) \propto$	Demographic Significance
1	Constant	1	General level of mortality
2	Linear	$x - \bar{x}$	"Gompertz slope", rectangularisation
3	Normal	$exp(-\frac{(x-\hat{x})^2}{\sigma^2})$	Young adult mortality
4	"Put option"	$(x_c - x)^+$	Childhood mortality
5	Rayleigh	$(x - \hat{x})exp(-\rho^2(x - \hat{x})^2)$	Postponement of old age mortality
6	Log-normal	$\frac{1}{x}exp(-\frac{(ln(x)-\hat{x})^2}{\sigma^2})$	Peak of accident hump
7	Normal	$exp(-\frac{(x-\hat{x})^2}{\sigma^2})$	Late middle / old age mortality

Table 1: Age/period terms in the final model

Figure 10 shows (on a logarithmic plot) the contribution each of these

terms makes to improving the goodness of fit (measured by the BIC) of the model. It can be seen that the majority of the improvement in goodness of fit comes from the first three age/period terms. However, the other terms (as well as being statistically and demographically significant) are still important in describing genuine structure in the data. Without them, the cohort term - as the final catch-all term added to the model - would attempt to capture this structure, leading to it being wrongly specified and generating inaccurate and implausible forecasts of mortality rates when projected.

Figure 10: Improvement in goodness of fit at different stages of the general procedure
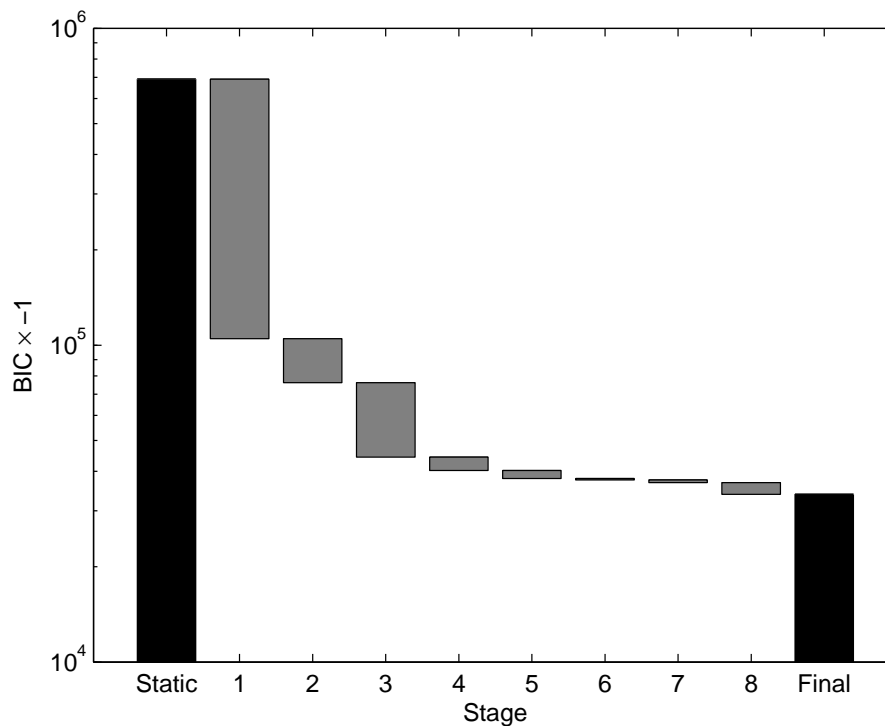

Our final model should, ideally, satisfy the desirable properties relating to the adequacy and goodness of fit of the model discussed in Section 2. Specifically

1. it should provide a good and parsimonious fit to the data (which should

have been achieved through the model fitting procedure);

2. it should extract all of the significant structure from the data, leaving residuals which are independent and identically distributed; and

3. it should give parameter estimates which are robust to small changes in the data.

Figure 11: Heat map of residuals from Stage 8
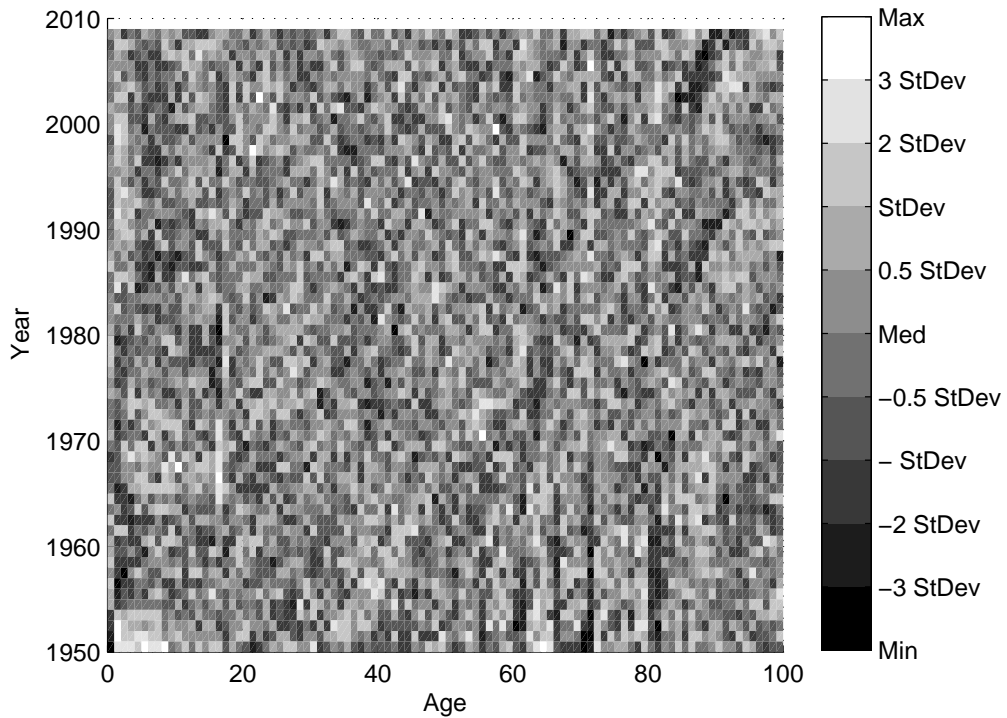
To test for structure within the standardised deviance residuals, we extend the procedures in Dowd et al. (2010b) We first plot the heat map shown in Figure 11. This shows an apparent lack of any major age/period or cohort features and there are very few "hot" and "cold" regions or clusters in the plot. We then calculate the sample moments of the residuals which are shown in Table 2. With large exposures and death counts and assuming the residuals have constant variance, we can use an approximation to assume that they

are $N(0,1)$ variables under the null hypothesis and so use the Jarque-Bera statistic to test for this.

	Residual mean	Standard deviation	Residual skewness	Residual kurtosis	Jarque-Bera statistic
General procedure	-0.01	0.94	-0.03	3.38	37.70
Lee-Carter	-0.02	0.98	0.47	9.75	11,700
PCA	0.00	0.94	0.06	3.26	21.25

Table 2: Properties of the residuals from Stage 8 of the general procedure and the Lee-Carter and PCA models

The critical statistic for the Jarque-Bera test at 95% is 5.99, whilst at 99% it is 9.21. This means that we decisively reject the assumption of normality for the standardised deviance residuals. Next, we consider the correlations of the residuals with those adjacent in the age and period directions, i.e.

$$\rho_x^X = corr(\epsilon_{x-1,.}, \epsilon_{x,.})$$
$$\rho_t^T = corr(\epsilon_{.,t-1}, \epsilon_{.,t})$$

Figure 12 shows the plot of these correlations against age and year and the relevant statistics if we test against the null hypothesis of independence (a two-tailed test at 95% significance) for the final model from the general procedure. Clearly, the hypothesis of independence is not supported overall. Testing these jointly (i.e., as a series of independent binomial trials where the probability of failure is 5% under the null) confirms the lack of independence in both the age and period directions at the 99% level.

This lack of normality and independence should be investigated further. In practice, this may be due to isolated outliers (often caused by data errors) or due to structural changes within the data. This would cause the variance of the residuals to change with age or time. Plots of the residuals from the model against age, period and cohort (not shown) indicate that there are no extreme outliers that would need to be investigated and that the variance of the residuals is roughly constant. Therefore, it is probable that there is unexplained structure remaining within the data which is not captured by the model However, comparing these results to those from the PCA model and

other models such as the Lee-Carter model show that the GP gives results which are at least as good as those from alternative mortality models.[6]

Figure 12: Correlations and tests statistics for residuals from the general procedure
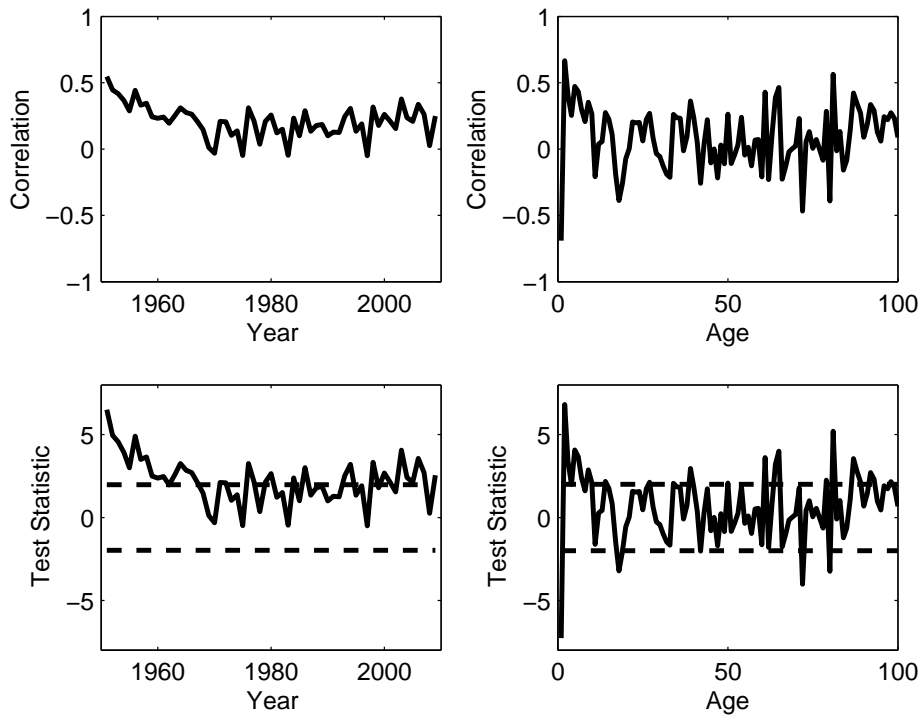
We also perform a number of tests of the robustness of the model to changes in the data. These include:

1. Fitting the model to different periods of data by increasing the start date sequentially from 1950 to 1980;

2. Bootstrapping the standard deviance residuals using a method based on the procedure of Koissi et al. (2006) to test the extent of parameter uncertainty; and

---

[6]We will compare the relative performance of alternative mortality models in Section 6.

3. Removing ages and years from the data by setting their weights to zero to test that none of the age/period functions are overly sensitive to specific ages and years.

The first of these tests is based on the procedure in Cairns et al. (2009). Graphs of the fitted parameters (not shown but available from the authors) indicate that the model fits similar patterns for the evolution of the different $\kappa_t^{(i)}$ period functions and slowly varying age functions as the age range of the data is changed.

The second robustness test we perform is to look at parameter uncertainty under residual bootstrapping. Standard bootstrapping techniques, such as that implemented by Koissi et al. (2006) were developed for use with the Lee-Carter model and assume that the residuals from the model are independent. However, this assumption is not valid.[7] Nevertheless, for simplicity, we implement an approach based on this method of residual bootstrapping in order to test our final model for parameter uncertainty. This method samples randomly from the fitted residuals and adds them to the fitted mortality surface to generate artificial death counts, to which the model is refitted to generate new parameter estimates. In this fashion, the degree of parameter uncertainty can be ascertained. The plots in Figure 13 depict fan charts (see Dowd et al. (2010a)) showing the 90% confidence interval for the period and cohort parameters produced by this bootstrapping procedure using 1,000 simulations. As can be seen, the underlying pattern of the parameters remains unchanged and there is no evidence to suggest that any terms are not significant when allowance is made for parameter uncertainty. The age functions are not shown, but these are considerably more robust to the effect of parameter uncertainty than the period and cohort effects.

As a final test of the model, we systematically remove ages and years from the data by setting their weights to zeros and then refitting the parameters. This tests if any of the fitted functions are overly sensitive to the specific rows or columns of the data grid, and the model's ability to interpolate sensibly for missing data. Figures 14 and 15 shows the impact of this analysis on

---

[7]More recently, stratified and block-bootstrapping procedures have been used (see D'Amato et al. (2011)) as have those based on geo-statistical techniques which look at the correlation structure across residuals (see Debón et al. (2008, 2010)).

the cohort parameters γ_{t-x} and on the age/period terms $f^{(6)}(x)$ and $\kappa_t^{(6)}$.[8] As can be observed, while removing specific ages and years can distort the cohort parameters at the end of the range of data, it does not substantially affect those estimated across more data points in the centre of the range. $\kappa_t^{(6)}$ is also robust under this analysis.[9]. We are therefore satisfied that our final model is robust under small changes to the data.

---

[8]This age/period term was chosen as the most specific age function fitted and therefore probably the most susceptible to uncertainty under this analysis.

[9]Corresponding graphs for the age functions and other period functions, not shown here, also show considerable robustness.

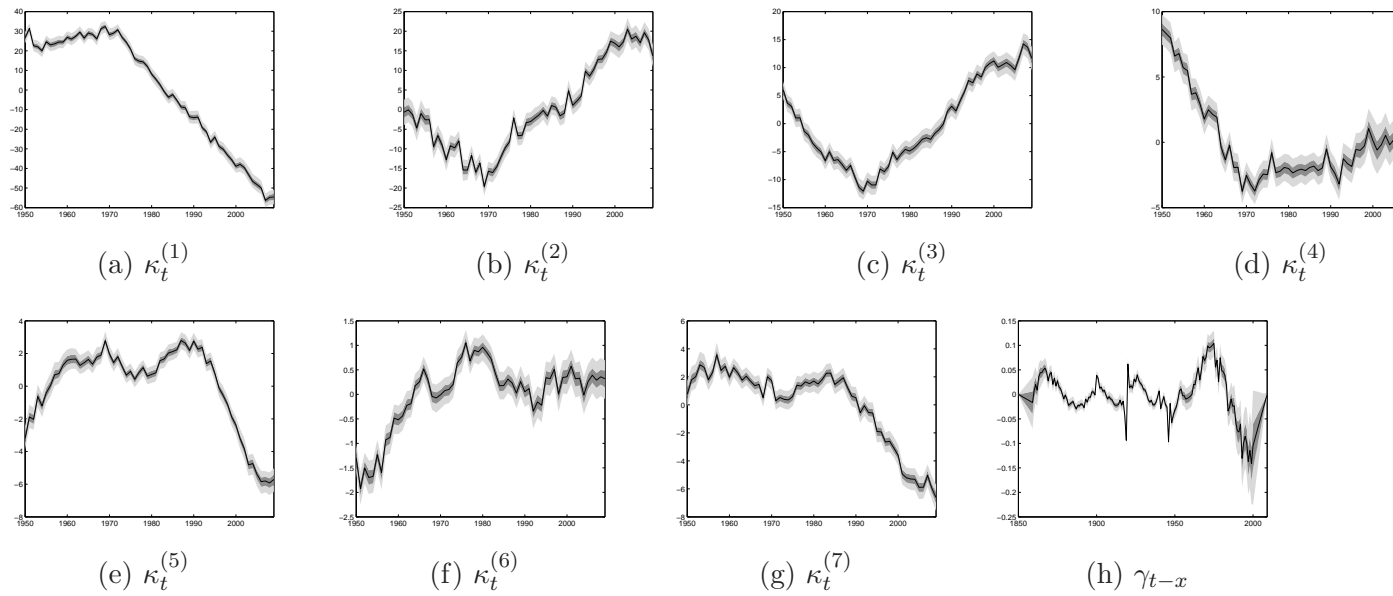Figure 13: Parameter uncertainty due to residual bootstrapping

(a) $\kappa_t^{(1)}$

(b) $\kappa_t^{(2)}$

(c) $\kappa_t^{(3)}$

(d) $\kappa_t^{(4)}$

(e) $\kappa_t^{(5)}$

(f) $\kappa_t^{(6)}$

(g) $\kappa_t^{(7)}$

(h) γ_{t-x}

Figure 14: Parameter uncertainty due to removal of one age of data

(a) $\kappa_t^{(6)}$

(b) $f^{(6)}(x)$

(c) γ_{t-x}
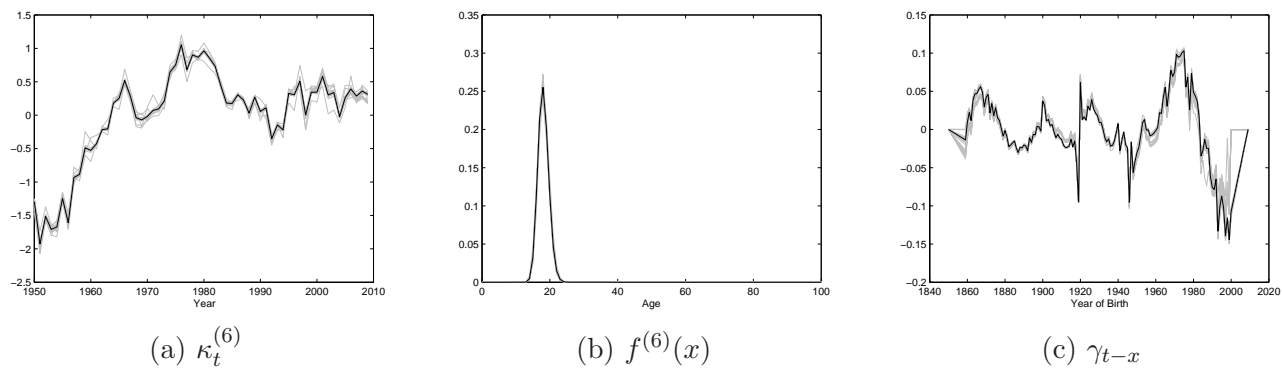
Figure 15: Parameter uncertainty due to removal of one year of data

(a) $\kappa_t^{(6)}$

(b) $f^{(6)}(x)$

(c) γ_{t-x}
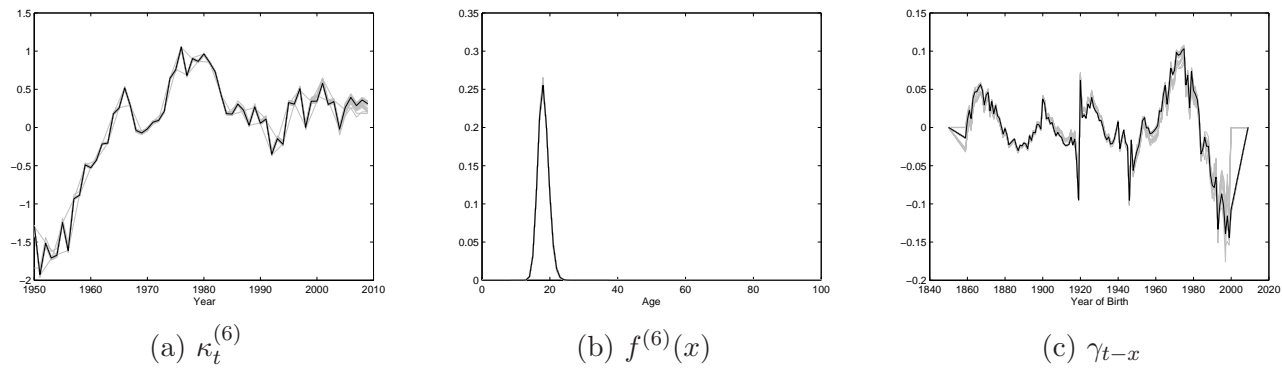
6 Comparison with alternative models

The model produced by the GP in Section 5 had some unexplained structure according to our analysis of the residuals. How serious a problem is this? Perhaps the best way to answer this question is to compare the model from the GP with some alternative mortality models: the Lee-Carter model (as the most widely used mortality model) and a method based on principal component analysis which extends the Lee-Carter approach with multiple age/period and cohort terms.

The Lee-Carter (LC) model, introduced in Lee and Carter (1992) has subsequently been much studied, developed and extended, most notably in the work of Lee (2000); Brouhns et al. (2002); Booth et al. (2002); Renshaw and Haberman (2003, 2006); Hyndman and Ullah (2007). It has rapidly become the benchmark mortality model against which others are compared (for instance in Cairns et al. (2009) or Plat (2009)) and so is a natural starting point for comparing the model produced by the GP against. However, it is a relatively simple model with only one age/period term and no cohort term, and so we would expect the GP to give significantly better fits to the data.

Principal component analysis (PCA) is a particular implementation of the singular value decomposition used in the work of Lee and Carter (1992) - see Huang et al. (2009) for more details. It is therefore the natural extension of the Lee-Carter methodology capable of giving multiple age/period terms. It finds age and period functions that explain the maximum amount of variance (across the period dimension) in the model. PCA and singular value decomposition have long been used in the study of mortality rates: for example Wilmoth (1990) uses it to detect higher order age/period functions, Booth et al. (2002) and Renshaw and Haberman (2003) both propose its use to extend the Lee-Carter model with additional age/period terms and the models of Hyndman and Ullah (2007) and Yang et al. (2010) use it directly to fit multiple age/period effects. However, it cannot directly find cohort effects. Therefore a direct comparison of PCA with our model is not appropriate.

In order to compare procedures, we use a method similar to that used in Wilmoth (1990). We first use PCA to find age/period functions for $ln(\mu_{x,t})$ in the absence of cohort effects. We then add a cohort effect to the under-

lying model and use the PCA age/period effects as the starting point when maximising the Poisson log-likelihood using the algorithms in Appendix A. This process is repeated for different numbers of age/period terms and the model with the highest BIC selected for comparison against our final model.

6.1 Results

Table 3 compares the three models and shows the goodness of fit to our dataset. The LC is a single factor model and so it is unsurprising that the other two models give considerably better fits to the data, although at the cost of a far greater number of parameters. The PCA method also requires substantially fewer age/period terms to achieve a very similar goodness of fit to the model produced by the GP. Because each of these age functions has approximately one hundred free parameters compared with a maximum of two using the GP, this does not result in fewer free parameters, however. Further, as we are primarily interested in the evolution of mortality rates over the period, we consider that it is desirable to have a high proportion of the parameters relating to the period and cohort effects of interest. This is not the case in the PCA model.

Model	No. A/P terms	No. free parameters	Log-likelihood	BIC
General procedure	7	679	-3.09×10^4	-3.38×10^4
Lee-Carter	1	259	-5.13×10^4	-5.25×10^4
PCA	3	735	-3.07×10^4	-3.39×10^4

Table 3: Goodness of fit for the different models

Figures 16 and 17 show the age and period functions for the GP and PCA procedure - the age and period functions for the LC model are the same as those shown in Figure 2 for the non-parametric terms. We find it difficult to assign demographic significance to the age functions in the LC and PCA models. The cohort parameters for the GP and PCA models are shown in Figure 18 - there is no corresponding plot for the LC model due to the absence of a cohort term. Here it is worth noting the similarities as well as the differences in the fitted parameters. Both approaches detect

the discontinuities after the First and Second World Wars and the increase in cohort mortality for years of birth around 1900 and between 1960 and 1980.

However, there are substantial differences in both the magnitude and the pattern of cohort parameters. Cohort effects for the GP are less pronounced than those from the PCA procedure. In addition, the PCA model fails to find a sustained decrease in cohort mortality for the "golden cohort" discussed previously. Most seriously, there appear to be large cohort effects at the beginning and end of the range of years of birth which are not explainable demographically. We believe that these effects are trying to compensate for the second and third age functions in the PCA model, which do not tend to zero at high ages (as shown in Figure 17a). This has very serious effects when these models are projected into the future. We therefore believe that the cohort parameters produced by the GP are more biologically reasonable and demographically significant than those fitted by the PCA procedure.

Table 2 above shows the moments and results of the Jarque-Bera tests on the residuals for the three approaches. We note that none of the three models tested give normally distributed standardised residuals, although the residuals from the GP and PCA models come considerably closer than those from the LC model.

We also compare plots of the residual heat maps in Figure 19 and test for correlation amongst the standardised deviance residuals in Figure 20 from the Lee-Carter and PCA models in Figure 20 - comparable plots for the GP are shown in Figures 11 and 12 respectively. The heat maps for the Lee-Carter and PCA models shows obvious clusters in the fitted residuals, indicating that there is still substantial structure remaining in the residuals of the PCA model. The LC residuals in particular show the clear need for a cohort term to capture the impact of the cohorts born after the First and Second World Wars. The PCA model yields residuals which are closer to normality than the GP, although they still do not pass the Jarque-Bera test. However, the correlations across residuals from the PCA procedure are higher than from the GP. Probably this is due to the smaller number of age/period terms. However, adding additional terms to the PCA model results in worse BICs and therefore will not improve the goodness of fit. This reinforces the conclusion that there is still structure in the data which is not adequately captured by the PCA model.

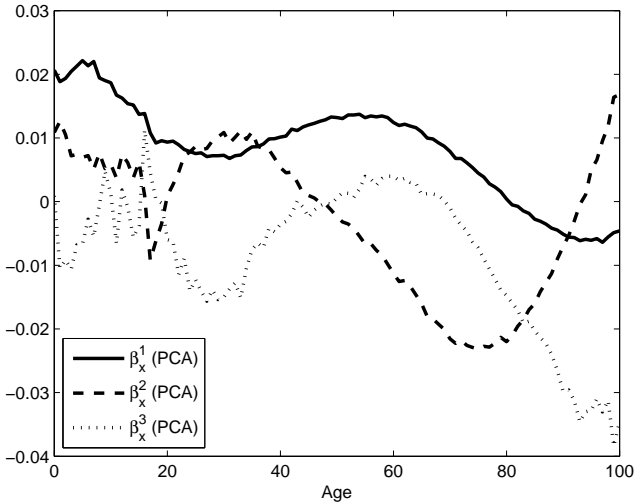Figure 16: Age and period functions for the general procedure
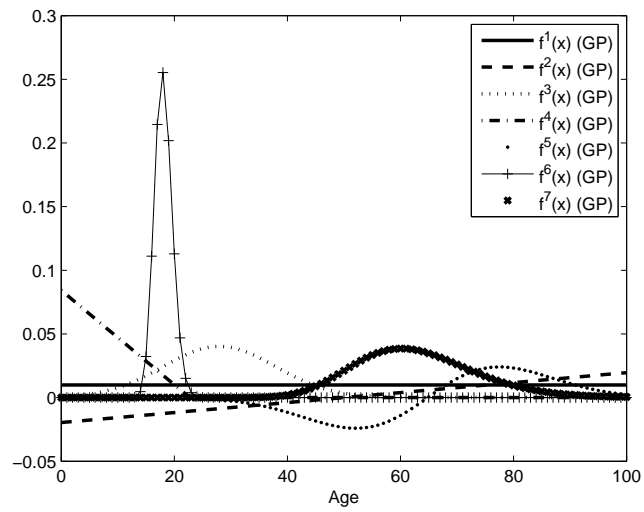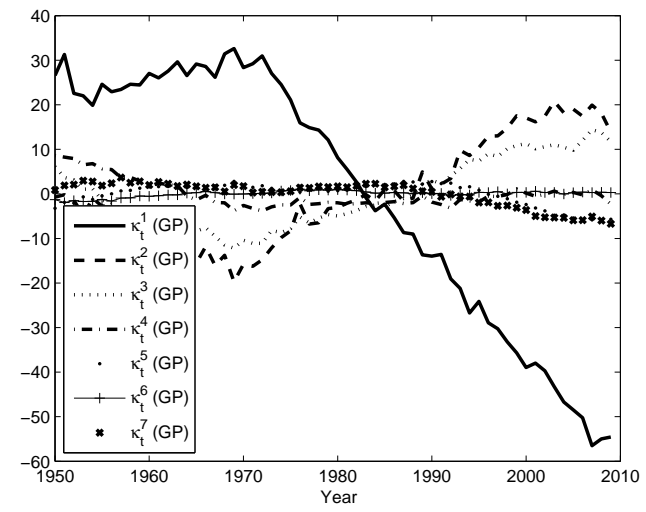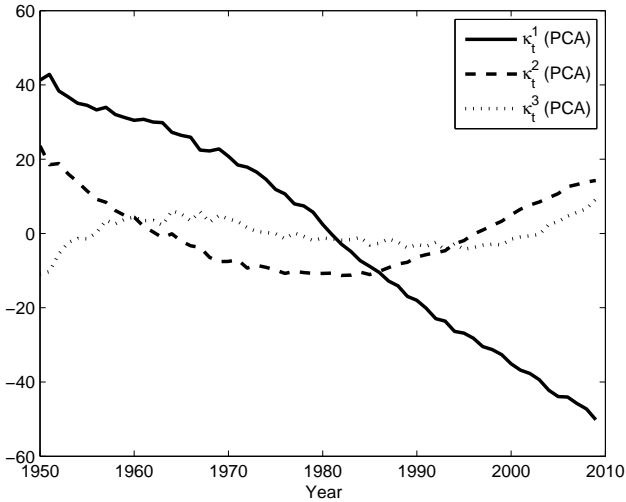
(a) Age functions

(b) Period functions

Figure 17: Age and period functions for the PCA model

(a) Age functions

(b) Period functions

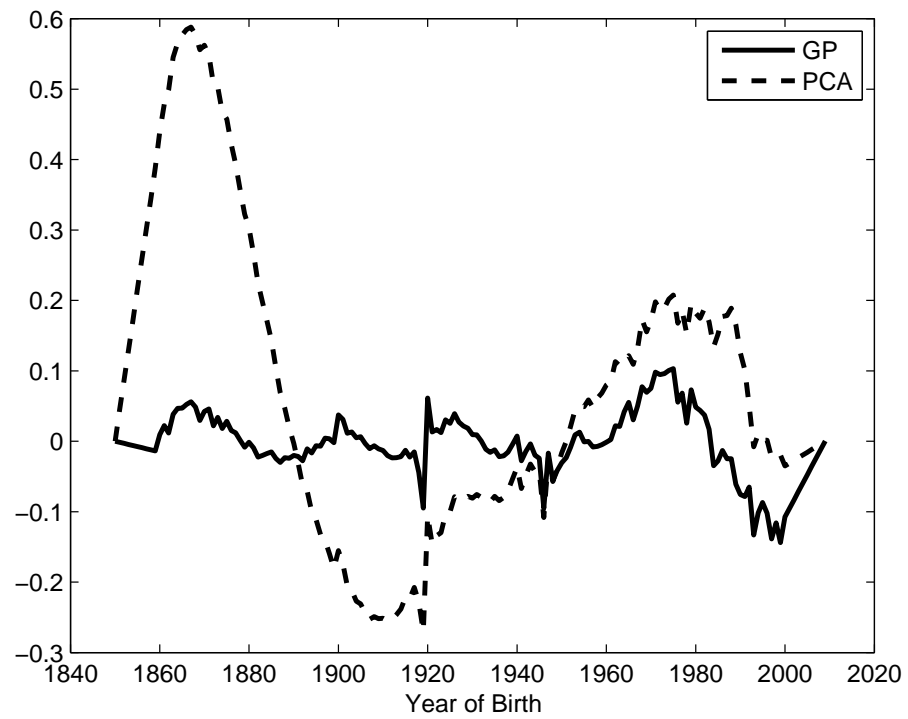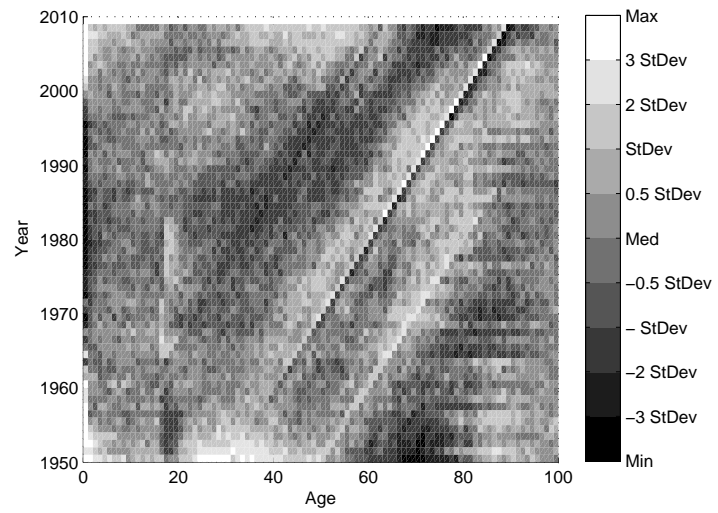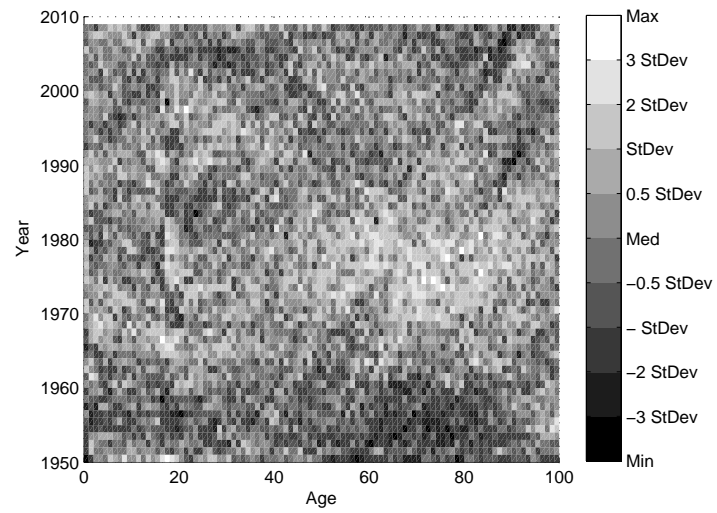Figure 18: Cohort parameters for the GP and PCA models
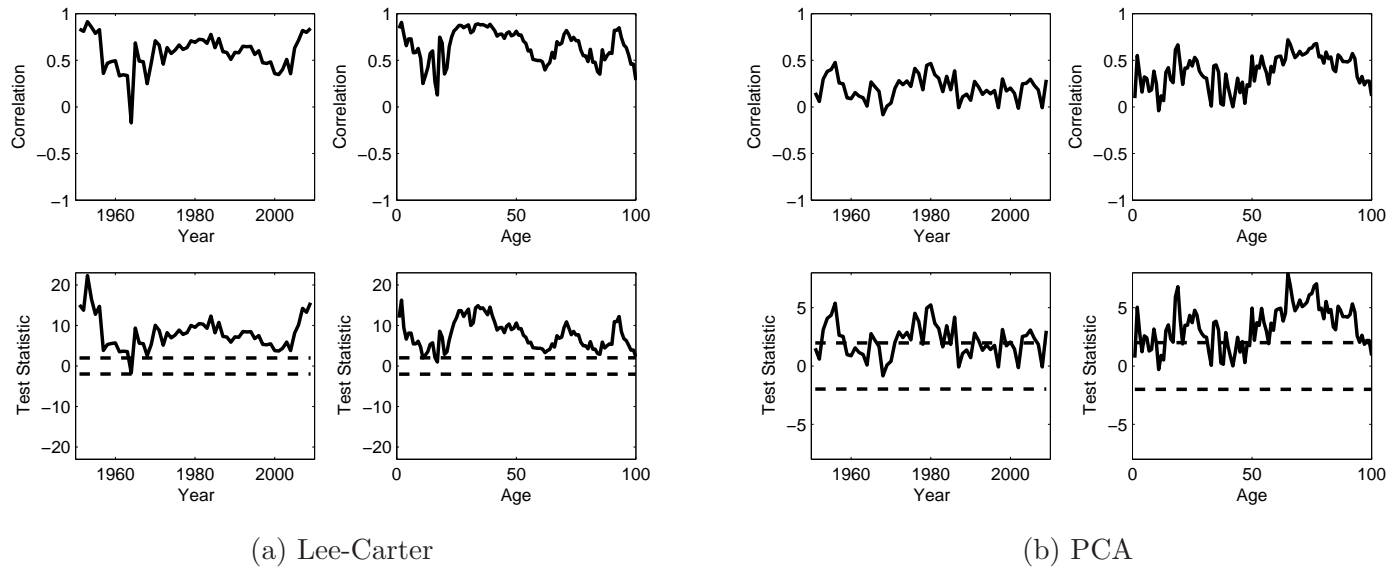
Figure 19: Residual heat maps for the Lee-Carter and PCA models

(a) Lee-Carter

(b) PCA

Figure 20: Residual correlations across age and period for the Lee-Carter and PCA models

(a) Lee-Carter

(b) PCA

7 Conclusions

As the level of interest in longevity risk increases, it becomes increasingly important to be able to construct more sophisticated mortality models reliably and robustly. These will need to capture most of the identifiable structure in mortality rates within the data - which calls for more terms - but to do so with the smallest number of free parameters - which calls for parsimony. Where cohort effects are believed to be real and important, they will need to be captured by the model. However, they must also be clearly distinguished from age/period effects in order that they can be projected correctly. This, in practice, means that all the significant age/period effects must be identified before any attempt is made to estimate the cohort effect. Finally, terms within the model should be capable of being associated with underlying biological or social processes. This requires judgement to be used to guide their projection and aid their communication with other, non-technical, stakeholders who are subject to longevity risk and wish to understand the implications.

In this paper, we have introduced a new, general procedure for constructing mortality models. The general procedure is driven by forensically examining the data to provide evidence for the selection of each and every term in the final model produced. We believe this improves the goodness of fit of the model parsimoniously and with demographic significance. We have applied the general procedure to a specific dataset, associated each term generated with an underlying demographic and/or socio-economic factor for the population being modelled, analysed the residuals to confirm that there is no identifiable structure remaining in the data which is not captured by the model, and compared the results with those from other methods of constructing mortality models.

The general procedure requires the modeller to engage intelligently with the data and make various, subjective decisions in its implementation. It is not a "black box" algorithm which can be deployed mechanically on various datasets, but rather requires a substantial investment of time to understand the underlying forces driving mortality within the population of interest and how these forces can be represented mathematically. But far from this being a disadvantage, we would argue that our approach accords perfectly with good model building practice, which seeks to move beyond a purely algorithmic approach in order to understand better the underlying structure of the

data.

In conclusion, we believe that the general procedure is capable of producing models which are in accordance with the desirability criteria of adequacy of fit to the data, demographic significance, parsimony, robustness and completeness (by including sufficient terms to cover all ages and cohorts).

However, we are aware that in order to be practically useful, a good fit to historic data needs to be accompanied by the ability to use the model to make reasonable forecasts of future mortality rates. Projecting models with multiple age/period and cohort terms consistently is a difficult problem as the historic time series are often highly correlated and display curvature, outliers or subtle trend changes which need to be accommodated (as have been described in Li and Chan (2005); Li et al. (2011); Coelho and Nunes (2011)). We therefore intend to address this issue in future work.

A Appendix: Algorithms and toolkit of function

In order to implement the general procedure, we need the ability to introduce new terms to existing models and to fit these to data. As we have central exposures to risk from the Human Mortality Database (Human Mortality Database (2012)), we adopt a Poisson likelihood maximisation approach which enables us to do this quickly and efficiently. This procedure is based on that implemented in Brouhns et al. (2002) and is described in Algorithm 1 at high level below.

The fitting algorithm used by the general procedure differs from the Brouhns et al. (2002) method in that the log-likelihood is maximised with respect to each set of parameters sequentially rather than simultaneously. It could be argued that this may lead the algorithm to find local rather than global maxima for the parameter values. In practice, we have not found this to be an issue and believe it can be largely resolved through finding the full set of invariant transformations of the parameters within the model (as discussed in Hunt and Blake (2013)). The maximisation of each set of parameters is done as per Algorithm 2 below.

Algorithm 1 Algorithm for Poisson likelihood maximisation

1: Set initial starting values and calculate initial log-likelihood
2: **while** Increase in log-likelihood less than threshold value (e.g. 10^{-4}) **do**
3: Maximise log-likelihood with respect to α_x holding all other parameters constant
4: **for** Each age/period term i **do**
5:    Maximise log-likelihood with respect to $\kappa_t^{(i)}$ holding all other parameters constant
6:    Maximise log-likelihood with respect to free-parameters θ in age function $f^{(i)}(x; \theta_i)$ or with respect to β_x holding all other parameters constant
7: **end for**
8: Maximise log-likelihood with respect to γ_{t-x} holding all other parameters constant if model contains a cohort term
9: Impose identifiability constraints through use of invariant transformations
10: Calculate updated log-likelihood
11: **end while**
12: Calculate residuals and BIC

Algorithm 2 Algorithm for maximisation of individual parameters

---

1: Start with values for maximisation passed from parent algorithm
2: **while** Increase in log-likelihood less than threshold value (e.g. 10^{-2}) **do**
3:     Calculate first derivative of log-likelihood with respect to parameters $\frac{dL}{d\xi}$
4:     Calculate second derivative of log-likelihood with respect to parameters $\frac{d^2 L}{d\xi^2}$
5:     Update estimate of parameters $\hat{\xi} = \xi - \phi \frac{\frac{dL}{d\xi}}{\frac{d^2 L}{d\xi^2}}$
6:     Impose simple identifiability constraints, e.g. level of $\kappa_t^{(i)}$, using invariant transformations
7:     Update fitted surface $\mu_{x,t}$ and log-likelihood
8: **end while**
9: Return updated parameter estimates, fitted mortality rates and log-likelihood to parent algorithm

---

This is nothing more than the repeated application of the Newton-Raphson procedure. The parameter $\phi \in [0, 1)$ is a simple scaling which can be lowered to improve the stability of parameter estimates (albeit at the cost of increasing the run time of the algorithm).

Models produced by the GP will not be fully identified and so will require additional identifiability constraints to be robustly estimated. A discussion of the origin and nature of this lack of identifiability and the selection of appropriate identifiability constraints is given in Hunt and Blake (2013). In summary, we impose the following identifiability constraints upon the final

model from Stage 8

$$\sum_t \kappa_t^{(i)} = 0 \quad \forall i \tag{5}$$

$$\sum_x |f^{(i)}(x; \theta_i)| = 1 \quad \forall i \tag{6}$$

$$\sum_x \beta_x f^{(i)}(x; \theta_i) = 0 \quad \forall i \tag{7}$$

$$\sum_c n_c \gamma_c = 0 \tag{8}$$

$$\sum_c n_c \gamma_c (c - \bar{c}) = 0 \tag{9}$$

$$\sum_c n_c \gamma_c ((c - \bar{c})^2 - \sigma_c) = 0 \tag{10}$$

We refer to Equation 6 as the normalisation of the age function. In order to normalise age functions with free parameters θ_i, we must modify the form of the age function so that $\sum_x |f^{(i)}(x; \theta_i)|$ is not a function of θ. This is usually achieved by multiplying it by a "self-normalisation" function $N(\theta_i)$. Equation 7 is only applied in exploratory models with a non-parametric term in order to maximise the distinctness of the age/period terms. In Equations 8 to 10, n_c is the number of observations of each year of birth in the data, $\bar{c}$ is the mean year of birth and $\sigma_c = \frac{\sum_c n_c (c - \bar{c})^2}{\sum_c n_c}$. These constraints are imposed to ensure that the fitted cohort parameters have mean zero and no linear or quadratic trends.

The functions in the toolkit we have developed so far are given in Table 4 along with the free parameters they require and the self-normalisation functions $N(\theta)$. In this, the age range is assumed to run from age 0 to age x_n, giving n ages in total with $\bar{x} = \frac{1}{n} \sum_{x=0}^{x_n} x = 0.5(x_n - 1)$ and $\sigma_x = \frac{1}{n} \sum_{x=0}^{x_n} (x - \bar{x})^2 = \frac{1}{12} x_n (x_n - 1)(x_n + 1)$. Some of these normalisations are only approximate or are true up to a constant, so it is still necessary to rescale the age functions after applying Algorithm 2 to optimise the value of the free parameters. Illustrative plots of the age functions are given in Figure 21.

Name	Function $f(x) \propto$	Normalisation $N(\theta)$	Free Parameters
Constant	1	$\frac{1}{n}$	none
Linear	$x - \bar{x}$	$\frac{1}{\bar{x}(\bar{x}+1)}$	none
Quadratic	$(x - \bar{x})^2 - \sigma_x$	$\frac{12}{x_n(x_n+2)^2}$	none
"Put option"	$(x_c - x)^+$	$\frac{1}{x_c(x_c-1)}$	x_c - pivot
"Call option"	$(x - x_c)^+$	$\frac{1}{(x_n-x_c)(x_n-x_c-1)}$	x_c - pivot
Exponential	$exp(-\lambda x)$	$1 - exp(-\lambda)$	λ - width
Gumbel	$exp(exp(-\lambda x))$	λ	λ - width
Normal	$exp\left(-\frac{(x-\hat{x})^2}{\sigma^2}\right)$	$\frac{1}{\sigma}$	$\hat{x}$ - location σ - width
Log-Normal	$\frac{1}{x}exp\left(-\frac{(ln(x)-\hat{x})^2}{\sigma^2}\right)$	$\frac{1}{\sigma}$	$\hat{x}$ - location σ - width
Rayleigh	$(x - \hat{x})exp(-\rho^2(x-\hat{x})^2)$	$0.5\rho^2$	$\hat{x}$ - location ρ - width^{-1}
Ellipse	$\sqrt{1 - \frac{(x-\hat{x})^2}{a^2}}$	$\frac{2}{a\pi}$	$\hat{x}$ - location a - width

Table 4: Age functions in toolkit

References

Aro, H., Pennanen, T., 2011. A user-friendly approach to stochastic mortality modelling. European Actuarial Journal 1 (S2), 151–167.

Booth, H., Maindonald, J., Smith, L., 2002. Applying Lee-Carter under conditions of variable mortality decline. Population Studies 56 (3), 325–336.

Brouhns, N., Denuit, M., Vermunt, J., 2002. A Poisson log-bilinear regression approach to the construction of projected lifetables. Insurance: Mathematics and Economics 31 (3), 373–393.

Cairns, A., Blake, D., Dowd, K., 2006a. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. Journal of Risk and Insurance 73 (4), 687–718.

Figure 21: Age functions in toolkit
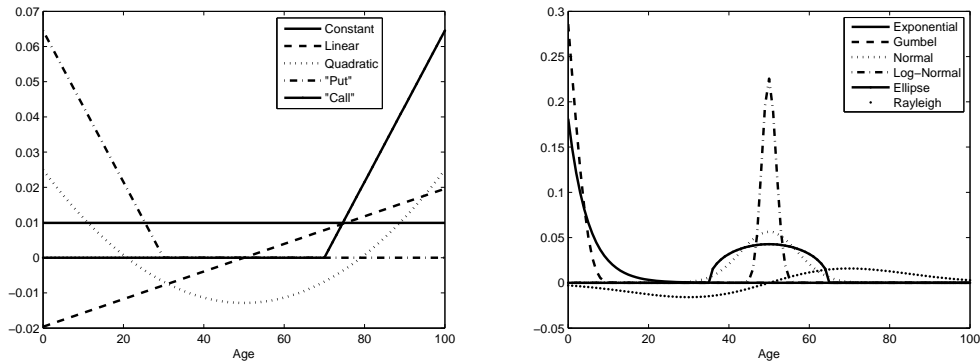
Cairns, A., Blake, D., Dowd, K., 2006b. Pricing death: Frameworks for the valuation and securitization of mortality risk. ASTIN Bulletin 36 (1), 79.

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

Campos, J., Ericsson, N. R., Hendry, D. F., 2005. General-to-specific modeling: An overview and selected bibliography. Federal Reserve Discussion Papers.

Coelho, E., Nunes, L. C., 2011. Forecasting mortality in the event of a structural change. Journal of the Royal Statistical Society: Series A (Statistics in Society) 174 (3), 713–736.

Currie, I., Durbán, M., Eilers, P., 2004. Smoothing and forecasing mortality rates. Statistical Modelling 4 (4), 279–298.

D'Amato, V., Lorenzo, E. D., Haberman, S., 2011. The Poisson log-bilinear Lee-Carter model: Applications of efficient bootstrap methods to annuity analyses. North American Actuarial Journal 15 (2), 315–333.

Debón, A., Martinez-Ruiz, F., Montes, F., 2010. A geostatistical approach for dynamic life tables: The effect of mortality on remaining lifetime and annuities. Insurance: Mathematics and Economics 47 (3), 327–336.

Debón, A., Montes, F., Mateu, J., Porcu, E., Bevilacqua, M., 2008. Modelling residuals dependence in dynamic life tables: A geostatistical approach. Computational Statistics & Data Analysis 52 (6), 3128–3147.

Dowd, K., Blake, D., Cairns, A., 2010a. Facing up to uncertain life expectancy: The longevity fan charts. Demography 47 (February), 67–78.

Dowd, K., Cairns, A., Blake, D., Coughlan, G., Epstein, D., Khalaf-Allah, M., 2010b. Evaluating the goodness of fit of stochastic mortality models. Insurance: Mathematics and Economics 47 (3), 255–265.

Finkelstein, M., 2012. Discussing the Strehler-Mildvan model of mortality. Demographic Research 26, 191–206.

Haberman, S., Renshaw, A., 2012. Parametric mortality improvement rate modelling and projecting. Insurance: Mathematics and Economics 50 (3), 309–333.

Hobcraft, J., Menken, J., Preston, S. H., 1982. Age, period and cohort effects in demography: A review. Population Index 48 (1), 4–43.

Huang, J. Z., Shen, H., Buja, A., 2009. The analysis of two-way functional data using two-way regularized singular value decompositions. Journal of the American Statistical Association 104 (448), 1609–1620.

Human Mortality Database, 2012. Human Mortality Database.

Hunt, A., Blake, D., 2013. On the structure and identifiability of mortality models. Work in Progress.

Hyndman, R., Ullah, M., 2007. Robust forecasting of mortality and fertility rates: a functional data approach. Computational Statistics & Data Analysis 51 (10), 4942–4956.

Koissi, M., Shapiro, A., Hognas, G., 2006. Evaluating and extending the Lee-Carter model for mortality forecasting: Bootstrap confidence interval. Insurance: Mathematics and Economics 38 (1), 1–20.

Lee, R. D., 2000. The Lee-Carter method for forecasting mortality, with various extensions and applications. North American Actuarial Journal 4 (1), 80–93.

Lee, R. D., Carter, L. R., 1992. Modeling and forecasting U.S. mortality. Journal of the American Statistical Association 87 (419), 659– 671.

Li, J., Chan, W., 2005. Outlier analysis and mortality forecasting: The United Kingdom and Scandinavian countries. Scandinavian Actuarial Journal 2005 (3), 187–211.

Li, J., Chan, W., Cheung, S., 2011. Structural changes in the Lee-Carter mortality indexes: Detection and implications. North American Actuarial Journal 15 (1), 13–31.

Mitchell, D., Brockett, P., Mendoza-Arriaga, R., Muthuraman, K., 2013. Modeling and forecasting mortality rates. Insurance: Mathematics and Economics 52 (2), 275–285.

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

O'Hare, C., Li, Y., 2012. Explaining young mortality. Insurance: Mathematics and Economics 50 (1), 12–25.

Pitacco, E., Denuit, M., Haberman, S., Olivieri, A., 2009. Modelling longevity dynamics for pensions and annuity business. Oxford University Press.

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

Renshaw, A., Haberman, S., 2003. Lee-Carter mortality forecasting with age-specific enhancement. Insurance: Mathematics and Economics 33 (2), 255–272.

Renshaw, A., Haberman, S., 2006. A cohort-based extension to the Lee-Carter model for mortality reduction factors. Insurance: Mathematics and Economics 38 (3), 556–570.

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

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

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

Wilmoth, J., 1990. Variation in vital rates by age, period and cohort. Sociological Methodology 20, 295–335.

Yang, S., Yue, J., Huang, H., 2010. Modelling longevity risks using a principal component approach: A comparison with existing stochastic mortality models. Insurance: Mathematics and Economics 46, 254–270.