SUBJECT SA0 OF THE INSTITUTE AND FACULTY OF ACTUARIES

Colin O'HARE (MMath. D.A.T. C.F.I.)

Essays in modelling mortality rates

Subject SA0 Advisors:

Dr Shane WHELAN (F.F.A., F.S.A., F.S.A.I.)
James Joyce Library Building
School of Mathematical Sciences
Statistics and Actuarial Science
University College Dublin

Dr Youwei LI (BSc., MSc., PhD(Finance), PhD(Math))
Queen's University Management School
Queen's University Belfast
Riddel Hall
Belfast

prepared at Queens University Management School

Declaration

The work of this thesis is my own and where material submitted by me for another degree or work undertaken by me as part of a research group has been incorporated into the thesis, the extent of the work thus incorporated has been clearly indicated

Signed ..... (Candidate)

Date .....

Dedication

This submission for SA0 of the Institute and Faculty of Actuaries is dedicated to my wife Diane and my daughter Lottie who have put up with my constant ramblings about research, mortality, statistics and data and never once asked me to explain myself.

I would also like to dedicate this work to the late Professor William Parry F.R.S. (1934 - 2006), Professor of Dynamical Systems and Ergodic Theory at Warwick University whom I knew well in the years between 1996 and 2004. During that very short period he instilled in me enough enthusiasm about the research process that 15 years after graduating from the University of Warwick School of Mathematics, and after a decade working and growing in the actuarial profession as a pension actuary I am able to return to my natural place in academia.

Colin O'Hare
September 2012

Acknowledgements

First and foremost I want to thank my supervisors Dr Youwei Li and Dr Shane Whelan. It has been an pleasure to work with them. It would not have been possible to write this submission for SA0 without their help and support. Their advice and friendship has been invaluable, and I am extremely grateful.

Secondly, I would like to also offer my appreciation to Dr Trevor Watkins and Georgina Warren who have answered my many SA0 related questions. Also thanks go to the Institute and Faculty of Actuaries, who through their forward looking education strategy created a route to fellowship via research and so enabled me to carry out the research activities that I enjoy whilst having those activities recognised by the profession of which I am proud to be a member.

Colin O'Hare
September 2012

Abstract

The field of mortality risk and longevity risk and in particular the accurate forecasting and financial management of such risks has become a topic of great interest to academics, actuaries and financial professionals. As individual life expectancies continue to improve and the era of low equity returns and low interest rates persists the current mechanisms for providing adequate coverage for individuals in their later years are coming under strain. In addition the development of financial hedging products has enabled many financial risks to be laid off and has exposed longevity risk as arguably the most significant un-hedged risk in the developed world. The new visibility of longevity risk and the constraint capacity of insurance firms and pension providers to accommodate it has led to a need and desire to create innovative ways to lay this risk off to new parties, namely the capital markets. This has resulted in the area of longevity becoming a key new growth area for the capital markets who are designing the products to be able to isolate, transfer and manage this risk.

Essential to the desire to create mechanisms to transfer longevity risk is the need to accurately forecast mortality rates. This will ensure that new products are priced appropriately and that a transparent market with willing sellers and buyers of the risk can emerge. Current research in mortality modelling does not prioritise the forecasting of mortality rates, instead it focuses on providing models that give a best fit to the data and on providing adequate short term forecasts. It also focuses on statistical, extrapolative time-series methods rather than engaging with the socio-economic and epidemiological factors that may be causing mortality improvements.

In this SA0 thesis I contribute to the existing literature on modelling mortality rates with focus given to the forecasting ability of models and to the value of socio-economic and epidemiological information. I introduce a new extrapolative forecasting model which improves upon the existing extrapolative models in terms of the fit quality and more importantly the forecasting ability. I also develop a dynamic factor model of mortality rates adopted from the economic literature which shows surprisingly good performance when compared to the current mortality models. In chapter 4 of the thesis

I statistically identify a weakness of the current extrapolative models. Namely, the acceleration in mortality improvement occurring since the early 1970's, and the inability of current models to adequately capture this.

The second half of the thesis focuses on socio-economic and epidemiological data and in particular the usefulness of this type of data in helping to explaining the changes that we see in mortality rates. In chapter 5 I consider the correlation between mortality rates and socio-economic variables across Northern Ireland and give some surprising findings in terms of which socio-economic variables correlate well with mortality rate changes. Finally, in chapter 6 I take socio-economic and epidemiological information and use this to inform future forecasts of mortality rates using an adaptation of the Girosi and King model.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Since the very earliest years people have had an interest in the human lifespan. In modeling and predicting longevity and mortality rates and in expected future lifetimes. Initially as a topic of general interest, later as a necessity in pricing insurances relating to lifetime and later still to quantify some of the major risks facing the ageing developed world. Edmund Halley, better known as an astronomer, in 1693 wrote his text *An estimate of the degrees of mortality of mankind* and was probably the first to consider a scientific approach to pricing annuity products using data on numbers of deaths. So was born the mortality table.

Early mortality tables were deterministic and static in nature assuming no further improvement in mortality rates over time and treating all lives as homogeneous with respect to mortality. In more recent years, and as populations age more rapidly and more stake-holders enter the market for mortality risk, greater attention has been given to the modelling and forecasting of mortality and in particular to the uncertainty surrounding mortality rates. Longevity risk is now seen as one of the world's most pressing finan-

cial risks as there appears to be no slow down in trend of improving life expectancy. In 2010 the Office for National Statistics published, as part of the National Population Projections[1], projections of period and cohort life expectancy until 2035. For males, the cohort life expectancy projection figure had improved to 94.2 for a new born in 2035, whilst the female equivalent figure had increased to 97.2. See table 1.1 for further figures. These numbers are more astonishing when you consider that only 30 years ago life expectancy for males and females in the U.K. were 71 and 77 respectively. In straight terms this means life expectancy has increased by 5.5% per decade for males and 4.5% per decade for females over the last 3 decades. More importantly, the impact that increasing life expectancy has on individuals in terms of pension provision and on society as a whole in terms of cost for the elderly is staggering. The International Monetary Fund recently gave some food for thought on increasing life expectancy warning that adding three years to average life expectancy could cost the UK economy 750bn and the eurozone 4.6 trillion. Given that in the next 25 years the O.N.S. is projecting life expectancy for a new born to increase by around 5 years from 2010 to 2035 the issue of accurately forecasting mortality rates suddenly becomes clear.

Mortality rates are very volatile and vary from year to year and from age group to age group as displayed by figures 1.1 and 1.2. As can be seen from the first figure. Mortality rates have declined for each age group over time but that rate of decline has varied from year to year and is not the same across age groups. From the second figure we can see the profile of mortality across age with snapshots taken at 1950,1965,1980 and 2005. What we can initially see here is that the shape of mortality curve has changed little over time with perhaps the only adjustments being that the accident hump has become more pronounced in more recent years and the gradient of the senescence or *later life* mortality declining slightly. These two changes in

[1]The O.N.S. report can be found at `http://www.ons.gov.uk/ons/dcp171776_253938.pdf`

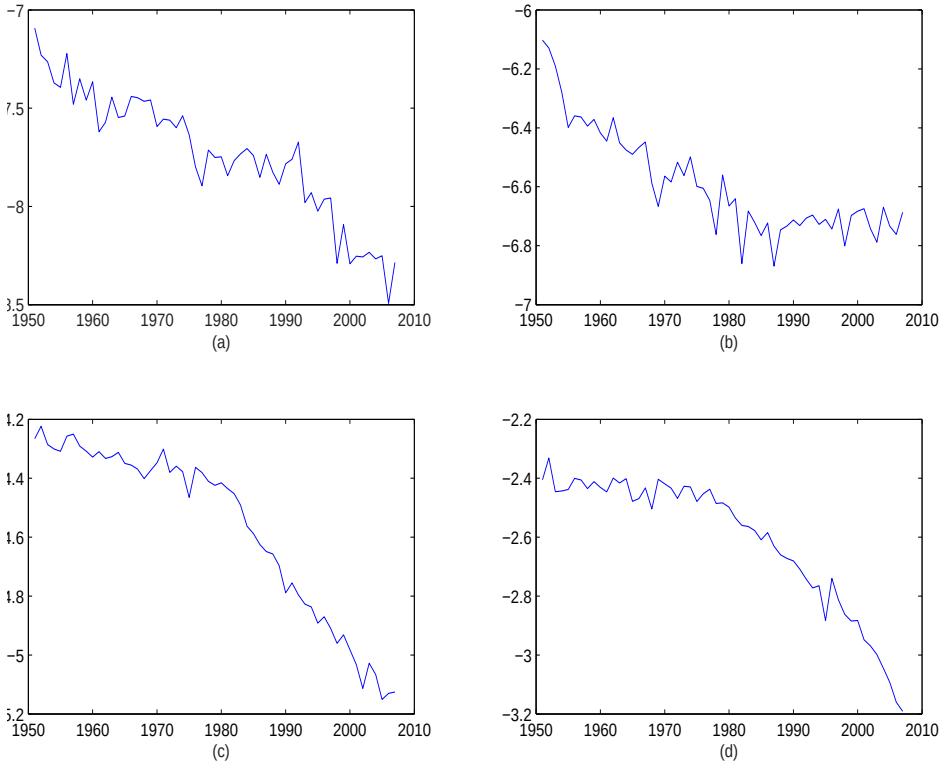

Figure 1.1: Logarithm of mortality by year for U.K. males aged (a) 15, (b) 35, (c) 55, and (d) 75.

the mortality profile suggest individuals are perhaps leading more riskier lives in their younger years but that having survived to the older ages they are now experiencing lower probabilities of death, perhaps due to medical advances.

What is clear from the diagrams is that mortality in the U.K. has been declining over time and this will have an adverse effect on financing in the future. What is of interest to academics however, is how to model the variation in mortality rates in such a way that we capture the dynamics adequately and are able to then forecast accurately. Answering the questions; (i) Why are mortality rates declining? and (ii) How can be better get to grips with estimating future mortality rates? In the following subsections we give a potted summary of some of the key contributions to mortality modelling

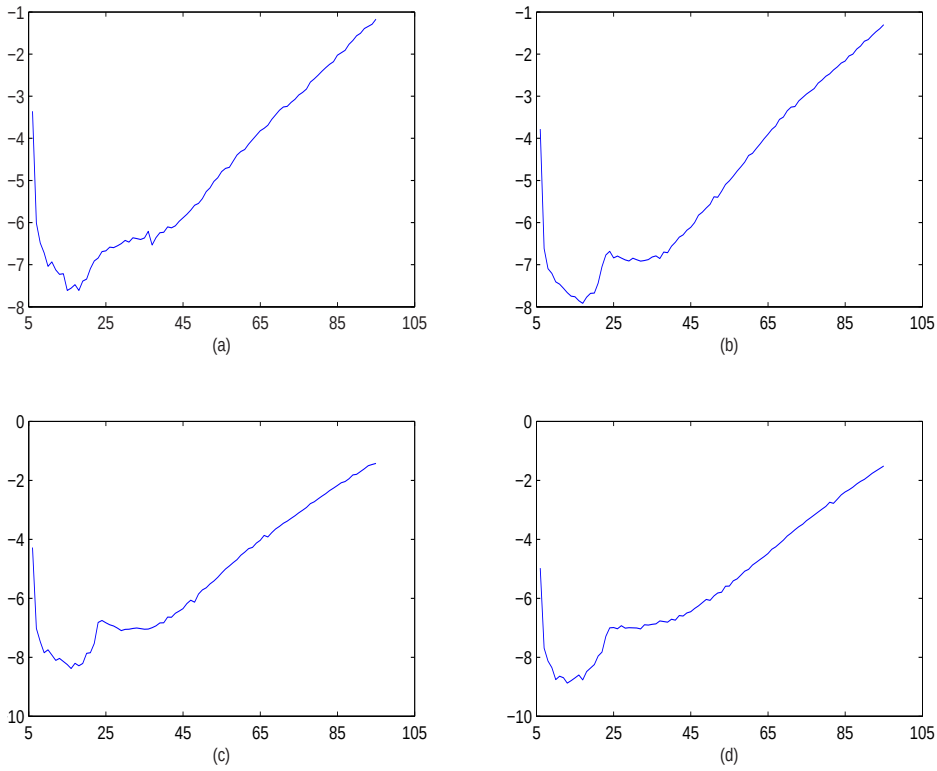

Figure 1.2: Logarithm of mortality for U.K. males during the years (a) 1950, (b) 1965, (c) 1980, and (d) 2005.

leading up to the current thinking on stochastic models of mortality rates.

1.1.1 Early mortality modelling

Early attempts to model mortality did not take account of potential future improvements in mortality rates. Rather, they took current and past data and assumed that future mortality would behave in the same way. An important step towards the development of early age-continuous mortality models came from the early mortality laws originating from the fitting of a mathematical formulae to the mortality data. Probably the first attempt to mathematically model mortality with a continuous formulae was

proposed in 1725 by Abraham De Moivre, who suggested

$$l_x = k\Big(1 - \frac{x}{86}\Big) for 12 \leq x \leq 86$$

where l_x is the number of individuals still alive at age x last birthday from an original pool, l_0, of individuals, and k is a normalizing constant, the assumption in this model being that all individuals will have died by age 86. However, as Haberman (1996) noted, a new era for the actuarial science started in 1825 with the law proposed by Benjamin Gompertz, the pioneer of a new approach to survival modelling. As it is well known, Gompertz's ideas can be properly expressed in terms of what we now call a "force" of mortality. Denoting by μ_x the force of mortality, Gompertz's law is as follows:

$$\mu_x = \alpha \exp(\beta x)$$

where α and β are positive parameters and x is the age.

Gompertz's law constitutes one of the most influential proposals in the early times of survival modelling. Actually, many contributions in the field of mortality laws, throughout the latter half of the 19th century, generalize or proceed from Gompertz's ideas. Remarkable examples are given by the laws proposed by William Makeham in 1860, Wilhelm Lazarus in 1867, Thorvald Thiele in 1867 and Ludvig Oppermann in 1870.

The Gompertz model of mortality focused on older ages (beyond infant and young adult periods were accidental deaths have a major contribution to mortality rates). This was seen as a weakness of the model and focussing on the problem of representing the mortality over the whole lifetime span, Thiele proposed the following function as the

force of mortality:

$$\mu_x = \alpha_1 \exp(-\beta_1 x) + \alpha_2 \exp(-\beta_2 (x - \eta)^2) + \alpha_3 \exp(\beta_3 x)$$

where all parameters are positive (non-negative) real numbers. The model can be broken down into it's three constituent parts. The first term on the right hand side represents the (decreasing) mortality at very young ages (improvements in mortality after overcoming the risks around birth). The second term represents the mortality hump at young-adult ages (recognising that mortality at these ages has less to do with natural deaths and more to do with lifestyle and accidental effects). The third term (which coincides with Gompertz's law) represents the mortality at adult and old ages. Combining the three components together Theile hoped to capture the variability in mortality rates across the whole lifespan. With $\alpha_1 = \beta_1 = \beta_2 = 0$ we obtain a specific case of the Theile model known as Makeham's law, which can be written as follows:

$$\mu_x = \gamma + \alpha \exp(\beta x)$$

where $\alpha, \beta > 0$ and $\gamma > 0$. Clearly this generalizes Gompertz's model and is generalized by the Theile model. The two terms on the right side can be interpreted as the background mortality γ which is independent of age and the senscent mortality $\alpha \exp(\beta x)$ dependent on the age.

Fitting Thiele's formula to experienced mortality is not a trivial matter (particularly at the time it was proposed). It is worth noting that, a century later, Heligman and Pollard (1980) proposed the same structure to model mortality odds rather than rates:

$$\frac{q_x}{p_x} = A^{(x+B)^C} + D \exp\left[-E(lnx - lnF)^2\right] + GH^x$$

In 1932, W.F. Perks proposed a family of survival models, represented by the following formula:

$$\mu_x = \frac{\alpha \exp(\beta x) + \gamma}{\epsilon \exp(-\beta x) + \delta \exp(\beta x) + 1}$$

Setting $\epsilon = \delta = 0$, we find Makeham's law, which is then also a member of the Perk's family. Perk's models constitute the first examples of heterogeneous modelling in the mortality arena. It should be noted that, for usual values of the parameters, the Perks intensity departs from the Makeham (and the Gompertz) intensity as x increases. In particular, whilst in Makeham's law (as well as in Gompertz and Thiele) we have

$$\lim_{x \to \infty} \mu_x = \infty$$

whereas with the Perks model we have:

$$\lim_{x \to \infty} \mu_x = \frac{\alpha}{\delta}$$

Thus with the Makeham and Gompertz models of mortality mortality rates tail off to zero over time, whereas with the Perks model mortality rates never reach zero but approach a positive rate $\frac{\alpha}{delta}$. The asymptotic behaviour of mortality is a very important issue when analyzing the mortality pattern at very old ages. It has been recently observed that the force of mortality is slowly increasing at very old ages, approaching a rather flat shape. In other words, the exponential rate of mortality increase at very old ages is not constant, as for example in Gompertz's law, but declines. This goes against the view that life expectancy should have a biological limit. However, many serious academics, e.g. James Vaupel (2005), see figure 1.3 would argue that there is much cause to believe that life expectancy could increase indefinitely. It is with this

level of uncertainty about the direction of future longevity that we see the proliferation of models of mortality in the last two decades.

The linear-regression trend in figure 1.3 is depicted by a bold black line (slope = 0.243) and the extrapolated trend by a dashed grey line. The horizontal black lines show asserted ceilings on life expectancy, with a short vertical line indicating the year of publication. The dashed red lines denote projections of female life expectancy in Japan published by the United Nations in 1986, 1999, and 2001. Notice the altered UN projection between 1999 and 2001. As can be seen life expectancy has regularly overshot forecasts.

It is also important to note at this point that increasing life expectancy is not restricted to being a U.K. only issue, although the presence of a well developed annuity market in the U.K. has made the problem more visible. Life expectancy around the world has been increasing in a similar way to the U.K. albeit starting off from a different position. The O.E.C.D. published a report into life expectancy at birth `http://www.oecd.org/social/familiesandchildren/47697608.pdf`, across O.E.C.D. countries (see figures 1.4 and 1.5) which showed that life expectancy is high and is increasing across all developed countries.

The proliferation of models that have been developed in the last two decades have been helpfully categorised by Booth and Tickle (2008) into one of three types of model.

- Extrapolative models - taking past data and extrapolating identified trends
- Explanatory models - modelling mortality as a dependent variable explained by socio-economic, biological and environmental factors
- Expectations models - taking advantage of the expert knowledge of actuaries and demographers and targeting future life expectancy at some expert held belief.

In this thesis we focus on the extrapolative and explanatory approaches to mod-

elling mortality leaving the expectations method to the actuarial experts. In the past the actuarial profession has relied on the expectations method of forecasting mortality rates valuing highly their history and expertise in the area but they are now moving more towards the extrapolative method (Continuous Mortality Investigations Bureau 2002,2005, 2006, 2007) as computing power increasing and methods for analysing large amounts of data become more accessible. Although interestingly the latest CMI model, CMI2011,[2] retains some expectations element allowing the actuary to use his or her expert judgement to decide on a long term reduction factor for mortality rates.

The most successful approach to modelling mortality in recent decades has been the extrapolative method which relies heavily on data which has become more and more available in recent years. Chapters 2 and 3 of this thesis adopt this approach to modelling mortality rates and contribute to the existing literature on extrapolative methods. In chapter 4 we identify a particular weakness of the extrapolative approach using statistical methods to identify the presence of structural breaks which extrapolative methods cannot deal with. A particular reason why the extrapolative approach has become so popular becomes clear when you consider the end users of the forecasts. The main users of mortality forecasts are actuaries and finance professionals who require the predictions to be able to price life insurance related products. This has meant that naturally they are less keen to need an answer to the question; *Why are mortality rates improving?* and have more of a tendency to want the answer to the question; *Where are mortality rates going next?*. The explanatory approach has suffered in this respect with limited work having been done academically in the area. Chapters 5 and 6 in this thesis will consider the potential of socio-economic data to contribute to the

[2]The CMI 2011 mortality model was created by the Institute and Faculty of Actuaries and can be found at `http://www.actuaries.org.uk/research-and-resources/pages/cmi-working-paper-55` and a user guide for applying it can be found at `http://www.actuaries.org.uk/research-and-resources/documents/user-guide-cmi-mortality-projections-model-cmi2011`

explanation of future mortality. In the remainder of this section we briefly summarize the progress made in the extrapolative and explanatory literature.

1.1.2 Extrapolative modelling

In the early 1990's researchers began to look at modelling mortality using time series to extrapolate the time trend based on historic mortality experience. These sorts of models make the implicit assumption that past trends identified in the data will continue into the future. They do not make any allowance for structural changes in those trends and the consequences that might have for mortality however, they are the best attempts that we have to date for modelling mortality. The first and most recognized of these types of models is the Lee Carter mortality model which models the time trend using a one factor stochastic model.

The Lee and Carter model, published in 1992, was the first attempt to model longevity data in a stochastic fashion by fitting the past mortality data and modelling the time trend as a stochastic process. The benefit of this for an actuary or an end user of mortality forecasts is that the uncertainty associated with mortality forecast can also be visualised as well as the mean or expected value. The Lee Carter model takes no account of cause of death or any explanatory modelling of mortality. Instead it models the data as a stochastic time series. It has become the baseline model against which all stochastic models of mortality have since been compared.

The model was developed as a simple one factor model which ensured the plausibility of projected age patterns. According to Fopuy and Haberman (2004), the Lee Carter model was a trade off between on the one hand, the separate age-specific projections that lacked consistency and plausibility and on the other hand, the rigidity of models projecting the parameters of a baseline model over all ages such as the Heligman-Pollard model.

The Lee-Carter model has a relatively simple formulation as a one-parameter family of life tables, one for each age x. The model postulated by Lee and Carter is given by:

$$ln(m_{x,t}) = a_x + b_x k_t + \epsilon_{x,t}$$

Where $m_{x,t}$ is the central rate of mortality for a life aged x for the year t. The model is made up from a component a_x, which is independent of time and the product of a second age dependent component, b_x, and a time dependent parameter k_t. The a_x term can be considered to represent the underlying mortality profile independent of time effects and the b_x term can be considered as a modulating factor adjusting the common time improvement effect κ_t for each specific age x. The $\epsilon_{x,t}$ are error terms representing age-specific influences that are not captured by the model. They are assumed to be normally distributed with mean zero and standard deviation σ_ϵ.

The Lee-Carter model, similar to many stochastic mortality models which followed it, is known to suffer from an identifiability problem. For example if b_x and k_t where to solve the model then the we can see that another perfectly acceptable solution to the model would be given by:

$$a_x^{(1)} = a_x + \alpha b_x \qquad b_x^{(1)} = \frac{b_x}{\beta} \qquad k_t^{(1)} = \beta(k_t - \alpha).$$

To avoid this problem, and to force a unique solution we need to impose two constraints on the parameters. With the Lee Carter model the natural choice of constraints, and that used in the paper by Lee and Carter is to set the following conditions on the

parameters:

$$\sum_{k=0}^{n} k_t = 0 \ \sum_{k=0}^{n} b_x = 1$$

The first of these conditions implies that for each x the value of a_x will be approximately equal to the average value over time of the $log(m_{x,t})$. The second condition ensures that the values of a_x and b_x are unique for each x. When the model is fit using Ordinary Least Squares (OLS) the parameters can be interpreted as follows:

- a_x exactly equals the average trend of $ln(m_{x,t})$ averaged over the time variable.
- b_x represents the age-specific patterns of mortality change. It indicates the sensitivity of the logarithm of mortality to changes in the time index k_t
- k_t represents the time trend. The forces of mortality change according to the overall mortality index k_t modulated by the age specific parameters b_x.

The model was fitted to age-specific rates of mortality for U.S. population between the years of 1933 to 1987. The data which included numbers of deaths, $D_{x,t}$, and numbers of lives exposed to death, $E_{x,t}$,for lives aged x last birthday during year t were grouped into 5 year age bands for the purpose of the fitting an forecasting of mortality rates. Grouping ages in this way reduced the volatility of the data points and ensured that all age groups had credible amounts of data in them.

Whilst the Lee Carter model opened up the door to stochastic modelling of mortality it had several, quite major flaws which have added to the development of a range of alternative models. The simple nature of the model reduces its flexibility and means that there is a simple correlation structure between projected mortality rates for different age groups. Since all age groups are projected using a single value for k_t this means that mortality improvements for the younger generations would be linked to

mortality improvements for the older generations via the modulation coefficients b_x. Also, the Lee Carter model requires a significant number of parameters. for modelling age groups in 5 year age bands up to the 85+ aged band we require 36 parameters for a_x and b_x. The Lee Carter model also ignores additional effects later identified in the data for several countries such as the cohort effect.

In 2002 Brouhns *et al* proposed an alternative method to fit the Lee Carter model using Poisson error terms rather than the Gaussian error terms used in the original Lee Carter approach. The advantage of this becomes apparent when you consider data at the older ages where the logarithm of the observed force of mortality at older ages is much more variable. This is because of the smaller absolute number of deaths at older ages making the Gaussian assumption an unrealistic one (Brouhns *et al* 2002).

The fact that the number of deaths can be thought of as a counting variable makes the poisson assumption a plausible one for the error terms. The model is re expressed in this form:

$$D_{x,t} \sim Poisson(E_{x,t} m_{x,t})$$

with $m_{x,t} = exp(a_x + b_x k_t) + constant$. We can then estimate the parameters a_x, b_x, and k_t using a maximum likelihood function approach. The likelihood function is given by:

$$L(\mathbf{a}, \mathbf{b}, \mathbf{k}) = \sum_{x,t} D_{x,t}(a_x + b_x k_t) - E_{x,t} exp(a_x + b_x k_t) + constant$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_M), \mathbf{b} = (b_1, b_2, \ldots, b_M), \mathbf{k} = (k_1, k_2, \ldots, k_N)$ and there are M data points for each calender year and N calender years of data.

Again, after fitting the a_x,b_x and k_t an ARIMA model is used to forecast the time

trend k_t, the forecast k_t then being used to forecast mortality rates. The parameter estimation by this method shows a similar trend to that when using the classical Lee Carter approach to estimating a_x, b_x and k_t. It is also found that the Poisson error approach is able to account for slightly more of the variability than the SVD method.

Around the same time that Brouhns had proposed the poisson approach to modelling mortality rates, Renshaw and Haberman were also producing a paper that looked at developing the Lee Carter approach further. In their paper they proposed several adjustments to the original Lee Carter model including; (i) Lee Carter with the addition of the second SVD component, (ii) Poisson linear model, (iii) Poisson linear with hinge, (iv) Poisson bilinear (as in the Brouhns et al paper), and (v) Poisson double bilinear. In their analysis they also concluded that the only significant difference between the traditional Lee Carter approach and the Poisson error approach occurs in the older age groups where the Poisson error structure explains more of the variability than the traditional Lee Carter approach. They also explored the Generalised Linear Model approach to mortality modelling and developed several models in this framework.

The use of a penalised spline model was proposed in 2004 (Currie 2004, CMI 2006). A P-splines model takes the form:

$$ln(m_{x,t}) = \sum \theta_{i,j} B_{i,j}(x,t)$$

where the $B_{i,j}$ are pre-specified basis functions with regularly-spaced knots and the $\theta_{i,j}$ are parameters to be estimated. It is well known that the use of splines leads to over parameterised models that fit the data "too well" and the resulting mortality surfaces are therefore extremely lumpy. This is solved by penalising the roughness in the parameter estimates $\theta_{i,j}$. Smoothing in this way itself introduces additional complications since too much smoothing in the time space leads to systematic over or

under estimation of future mortality rates. This led Currie (2007) to develop a P-splines model which incorporates period shocks.

In terms of fitting, the P-splines model is shown to fit the mortality data better (Currie *et al* 2004) due to the local nature of the parameters it is able to adapt more readily to variability in the mortality rates. It does this with fewer parameters than the Lee Carter model, about 64 parameters for the P-splines model vs 231 for the Lee Carter model (Currie et al 2004 page 292). The P-splines model also provides a much lighter forecast of future mortality than the Lee Carter model.

Several multi-factor models have been developed that incorporate age and period effects and which extend mortality modelling in the vain of Lee and Carter. In Renshaw and Haberman (2003) a model was proposed which had two dependent time effects. Their model is formulated:

$$ln(m_{x,t}) = \beta_x^1 + \beta_x^2\kappa_t^2 + \beta_x^3\kappa_t^3$$

where κ_t^2 and κ_t^3 are dependent period effects.

When comparing this model to the single factor Lee Carter model using either a gaussian or poisson error structure Renshaw and Haberman's finding are inconclusive.

In 2006, Cairns Blake and Dowd decided to focus on mortality at the older ages 60-89 on the basis that this would be where most of the interest would lie in terms of uncertainty. They developed a two factor model of mortality, (the CBD model) with the first factor affecting mortality at all ages equally and the second factor affecting mortality in a way proportionate to the age. The model itself is a stochastic version of

the Perks (1932) model:

$$\mu_x = \frac{\alpha \exp(\beta x) + \gamma}{\epsilon \exp(-\beta x) + \delta \exp(\beta x) + 1}$$

and is specified as:

$$\tilde{q}(x,t) = 1 - p(t+1, t, t+1, x) = \frac{\exp A_1(t+1) + A_2(t+1)(x+t)}{1 + \exp A_1(t+1) + A_2(t+1)(x+t)}$$

In this model $p(t+1, t, t+1, x)$ is a probability function that has been developed along parallel lines to the interest-rate terminology. In this case the function $p(t, T_0, T_1, x)$ denotes the probability at time t that an individual, aged x at time 0 and alive at time T_0 will still be alive at time T_1.

$A_1(t)$ and $A_2(t)$ are estimated for each t using ordinary least squares on the transformed mortality rates

$$log(\frac{q_t}{p_t}) = A_1 + A_2 t + \epsilon$$

They are then projected forward by modelling the future distribution of $A(t) = (A_1(t), A_2(t))$ as a two dimensional random walk with drift. Specifically,

$$A(t+1) = A(t) + \mu + CZ(t+1)$$

The model displays some of the signs of a biologically reasonable model of mortality as defined by Cairns Blake and Dowd (2009). In particular the trend value $\mu_1$is negative in the fitted model indicating that mortality, as predicted by the model, is

generally improving. The positive fitted value of μ_2 indicates that mortality rates at higher ages are improving at a slower rate. The model does predict that morality rates beyond age 113 are deteriorating which may be though of as an undesirable feature in the model but the authors ignore this for the purpose of their analysis as the turning point happens at such a late age.

The authors also choose to fit the two factor model using data over two differing periods. Firstly, they fit the model using data from the years 1961 - 2002. They then repeat the exercise fitting the model using data from 1982 - 2002. In this way they demonstrate a steepening in the mortality trend μ after 1982 suggesting that mortality improvements have been accelerating over more recent history. Looking at the cohort aspect of the mortality model a biologically reasonable model should also demonstrate increasing mortality rates for older generations than for younger generations. In other words for fixed time t the function $\tilde{q}(t, x)$ should be increasing in x. For the two factor model this reduces to requiring that $A_2(t)$ remains positive. Theoretically A_2 could become negative but Cairns et al conclude that this would not be the case, at least for the data set on which they base the paper, since the starting point for A_2 is relatively high and the drift factor μ_2 is positive. The CBD model has also been extended several times, firstly to allow for the cohort effect in a level way, secondly to include a quadratic term in the age parameter and finally to allow for a cohort effect which diminishes over time.

In Plat (2009) sought to bring together the positive aspects of each of these extrapolative models in a unifying extrapolative model. The model maintains the good aspects of the existing models whilst leaving out the weaker features. The result was a four factor model which took its beginnings from the Lee-Carter model and which added factors to capture the second age-period effect, as per the Cairns *et al.* (2006) model and the cohort effect, as per the Renshaw and Haberman (2006) model. The

innovation in the Plat model was to then add a further period factor affecting only the lower ages and designed to allow the model to fit to the whole age range. The model specification is given by:

$$\ln(m_{x,t}) = a_x + \kappa_t^1 + \kappa_t^2(\bar{x} - x) + \kappa_t^3(\bar{x} - x)^+ + \gamma_{t-x} + \epsilon_{x,t}, \tag{1.1}$$

where the a_x is similar to that of the Lee-Carter model and makes sure that the overall shape of the mortality curve by age is reasonable, the κ_t^1 and κ_t^2 model the mortality rates as in the Cairns *et al.* (2006) model and the κ_t^3 models the effects specific to the lower ages only where $(\bar{x} - x)^+$ takes the value $(\bar{x} - x)$ when this is positive and zero otherwise. Finally the γ_{t-x} models the cohort effect.

A particular problem with each of these types of model, and one which we highlight in chapter 3 of this thesis is that of the presence of structural changes in the identified trends in the mortality data. In many of the models above the forecasting plausibility boils down to the first time factor, κ_t or κ_t^1, which contributes in a major way to the forecast robustness and reasonableness. In each of the models above this is fitted using Box-Jenkins methods and ultimately results in a fitted random walk with appropriate drift. The problem with this however is that it suggests that (removing random fluctuations) the mortality improvement factor will change in a linear fashion going forwards with the gradient of that improvement being determined by the start and end dates of the sample data. Cairns *et al* (2006) identified this when creating their 2 factor model which when fitted to differing time periods showed a steepening of the mortality improvement factor after 1982. In chapter 3 we statistically investigate this for several of the above models using the methods of Bai and Perron (2003) to identify structural breaks in the models.

1.1.3 Explanatory models

The explanatory approach to modelling mortality is an under developed method, perhaps owing to the ease with which extrapolative models can be developed. They are based on structural or causal models of all cause mortality or specific cause mortality and use either known risk factors associated with specific casues of death or more general social or economic environmental factors. Thus use is made of medical, economic and socio-demographic knowledge and information on behavioural and environmental change. The main advantage of the explanatory approach is that the modelling approach can be used to identify significant risk factors which may then feed into policy and process in order to influence future mortality profiles. If we can identify the risk factors influencing longevity this might also provide useful insights into a deeper understanding of longevity. The explanatory approach to forecasting has yet to be fully developed as the relationships between risk factors and mortality are imperfectly understood but there has been some progress.

Many of the models used in explanatory forecasting are regression-based and therefore fit within the G.L.M. framework. They differ from regression-based extrapolative models in that they incorporate explanatory variables or risk factors, which are either lagged or forecast. When structural models are based exclusively on exogenous lagged risk factors, forecast horizons are limited to the shortest lag time. For example, in King and Soneji (2011) the authors lag the smoking variable in their structural model by 25 years and then forecast mortality rates 25 years into the future. In chapter 6 we carry out a King and Soneji approach but applying exogenous variables contemporaneously to overcome this issue.

The King and Soneji (2011) approach extends the work of Girosi and King (2008) that developed a method of modelling mortality rates across ages, years and countries. that method used a Bayesian hierarchical approach to information pooling. Their ob-

jective in doing this was to make use of beliefs that data across neighbouring ages, years or countries should show similar characteristics. For example, we might expect that the mortality rate experienced by a 20 year old in a given year should be similar to that experienced by the 21 year old or the 19 year old in the same year. Similarly, the mortality rate in say 2000, for a given age should be similar to the mortality rate for that same age in 1999 or in 2001. The hierarchical approach allows the smoothing of mortality rates for a single country across ages and time and so produces realistic forecasts of mortality that do not break norms in terms of age and time going forward (for example, mortality rates increasing with age and improving in time). Considering the logarithmically transformed mortality rate during year t for life aged x as $m_{x,t}$ they set out the following model specification:

$$m_{x,t} \sim N\big(\mu_{x,t}, \frac{\sigma_x^2}{b_{x,t}}\big) \tag{1.2}$$

$$\mu_{x,t} = Z_{x,t}\beta_{x,t}$$

The specification only differs from a standard linear regression model in the $b_{x,t}$ weighting that is applied to the variance and in the approach to defining the parameters β_x and σ_x^2. The specification above provides the basic building block of the Bayesian hierarchical approach in which we now interpret the coefficients β_x and standard deviations σ_x^2 as random variables with their own prior distributions. The prior on the coefficients β_x which depends on its own "hyper-parameter" θ is denoted $P(\beta|\theta)$ with prior on the hyper-parameter $P(\theta)$. The prior for the variance random variable $\sigma$is denoted $P(\sigma)$. The functional form of the priors is chosen to be tractable and diffuse so as not to influence the results with a gamma or inverse gamma density function being used.

The prior for the coefficient β is chosen to reflect the "similarity" belief across cross

sections. This is formalised by introducing a density function for the prior defined as:

$$P(\beta|\theta) \propto exp\big(-\frac{1}{2}H^{\beta}[\beta,\theta]\big) \tag{1.3}$$

where

$$H^{\beta}[\beta,\theta] \equiv \frac{1}{2}\sum s_{i,j} \parallel \beta_i - \beta_j \parallel_{\theta}^{2} \tag{1.4}$$

where the notation $\parallel \beta_i - \beta_j \parallel_{\theta}^{2}$ denotes a weighted Euclidean norm and where the symmetric matrix s is called the adjacency matrix.

It's entries reflect the "proximity" of cross section i to cross section j and hence the weight put on the relationship between the coefficients of cross section i and cross section j. Using this approach the fitted model shows forecasts that are smooth in the age and time dimension and that do not violate the smoothness beliefs across age and time that "may" be violated by using multiple regression methods.

Linear regression provides a useful framework for including potentially informative covariates, either a 'cohort effect' (e.g., a cohort's earlier smoking patterns) or a 'period effect'. Further by doing this within their model they also incorporate the empirical regularities of smoothing by age and time imposed in this set-up. The approach of King and Soneji (2011) was to develop a model with exogenous covariates by first identifying the links between mortality rates and lagged covariates, specifically smoking habits and obesity. They argued against using contemporaneous relationships in favour of lagged relationships and from the literature determined the optimal lag period to be 25 years in the case of smoking. They also argue that the additional forecasting step required to project the exogenous variables would lead to additional uncertainty in the model. In chapter 6 we adopt this approach but with the inclusion of more contemporaneous variables such as current GDP, health expenditure, alcohol consumption

or diet.

In the following section we outline in more detail each of the chapters of the thesis.

1.2 Overview and outline

In this thesis we look at various aspects of the modelling of mortality. We introduce several new models including the O'Hare and Li (2012) model in chapter 2 and the O'Hare and French, dynamic factor model in chapter 3. In chapter 4 we investigate a weakness of extrapolative mortality models, that is the inability of these models to account for structural breaks in the underlying data. We identify the structural breaks in a range of existing models and demonstrate the improvement in forecasting as a result of allowing for such structural breaks. In chapters 5 and 6 we consider the explanatory approach to modelling mortality and shift the focus onto identifying exogenous variables which may impact upon mortality rates. In chapter 5 we carry out an analysis of Northern Ireland mortality rates, linking the variation we see in mortality rates to the exogenous measures that we have for each region. In chapter 6 we look at a wider range of countries and identify explanatory exogenous variables for each of the countries considered, forecasting these factors to inform the future direction of mortality rates. In the subsections below we give further details on each of the chapters.

1.2.1 Chapter 2 - modelling non-linearity of mortality

The last twenty years has seen a vast literature on stochastic mortality models. However, when the models are fitted to a wider age range 5-89 rather than 20-89 or 50-89, the results are not satisfactory. This is mainly because the linear innovation are not flexible enough to capture the non-linear dynamics at the lower ages, the so called "lifestyle" mortality (accidents, drug abuse) profile. In this chapter we argue that a

modification to the forth factor is needed to provide a better fit whilst not losing any of the positive features of the existing models. The empirical results show that the proposed model has a better fit to the actual data, is robust, and also has a better forecasting ability.

We propose a model that combines the nice features from existing models as the Plat (2009) did and also allows for a better fit at the lower ages. More specifically, the model fits historical data very well, is applicable to a wider age range ages (5-89), fits the lower ages in a non-linear way, captures the cohort effect, has a non-trivial (but not too complex) correlation structure. The model has no robustness problems and it's structure remains relatively simple.

1.2.2 Chapter 3 - Dynamic factor modelling of mortality rates

In chapter 3 we apply a dynamic factor approach taken from the economic literature and apply it to mortality data focusing on improving forecasting ability. We focus on the short and longer term forecasting results and present improved accuracy against a range of published models. Using techniques from macroeconomic forecasting we propose a dynamic factor model of mortality that fits and forecasts mortality rates parsimoniously. We compare the forecasting quality of this model and of existing models and find that the dynamic factor model generally provides superior forecasts when applied to international mortality data. We also show that existing multifactorial models have superior fit but their forecasting performance worsens as more factors are added. The dynamic factor approach used here can potentially be further improved upon by applying an appropriate stopping rule for the number of static and dynamic factors.

1.2.3 Chapter 4 - Identifying structural breaks in mortality models

In chapter 4 we address statistically one of the main weaknesses of extrapolative models. namely their inability to address structural changes in mortality rates. Many stochastic models of mortality identify linear trends in mortality rates by time, age and cohort and forecast these trends into the future using standard statistical methods. The modelling approaches used fail to capture the effects of any structural changes in the trend and thus potentially produce incorrect forecasts of future mortality rates. Here we consider a range of published stochastic models of mortality and tests for structural breaks in the fitted mortality improvement factor κ_t or κ_t^1. We carry out the analysis using data from 1950 to 2000 for a range of developed countries and using the models of Lee and Carter, Carins, Blake and Dowd, Plat, and O'Hare and Li. We find that in almost all cases structural breaks in the time series are present and when allowing for these the resulting forecasts are significantly different.

1.2.4 Chapter 5 - Spatial variability of mortality in Northern Ireland

A concern for actuaries in using national mortality rates to price insurance and pensions products is the inherent heterogeneity that is present in populations. In this chapter we explore this in terms of the Northern Ireland region.

Mortality rates are known to vary by geographical location and to depend on socio-economic factors. Demographic, ethnic and socio-economic mortality factors vary by geographical location. Regions that are in close proximity to one another are expected to have similar mortality because of similar socio-economic factors and demographic characteristics. In this chapter the spatial variability of Northern Ireland mortality is

assessed using a spatial model encompassing explanatory risk factors including age, income, employment, health and education. We include a measure of the variability of the density of loyalist / nationalist populations and applying a similar technique to the spatial model we employ a ”social” model using political density as an explanatory variable for mortality variability. Data is split by geographical region based on super output area (SOA) and comes from the Northern Ireland Statistics and Research Agency (NISRA).

Regressions are estimated using an hierarchical Bayes model with Markov Chain Monte Carlo methods for mortality rates in 890 super output areas in Northern Ireland using data from the year 2008. We demonstrate the improved fitting achieved by using a spatial or social models explain mortality variation particularly when limited data is available for socioeconomic factors. Deprivation factors, which also vary spatially, reduce the need for spatial models for mortality. The modelling has implications for pricing and risk management in life insurance companies.

1.2.5 Chapter 6 - Forecasting death rates using exogenous determinants

Mortality models used for forecasting are predominantly based on the statistical properties of time-series and do not generally incorporate an understanding of the forces driving trends. In this chapter, we identify explanatory variables for mortality variation in a number of developed countries using statistical techniques developed in macroeconomics and finance. We consider whether the space spanned by the latent factor structure in mortality data can be adequately described by developments in GDP, health expenditure and lifestyle-related risk factors. These covariates are then shown to improve forecasts when we compare the forecasting performance of our variant of

the Bayesian hierarchical modelling approach of King and Soneji to forecast mortality against the benchmark mortality models in the field.

1.3 Further Discussions

The chapters presented in this thesis contribute to the literature in mortality modelling by developing a new extrapolative mortality model, widening the age range that it may be applied to and by applying modelling methods from the field of economics to introduce a new model to the mortality sphere. It also provides statistical evidence of the inability of some of the leading extrapolative models of mortality to adequately capture the structural breaks present in mortality rates. Particularly the acceleration in mortality improvement seen in the early 1970's due to improved mortality at older ages. In the later chapters of this thesis attention is drawn to the ability of socio-economic variables explain some of the variability seen in mortality rates. At a within-country level we identify the correlations between mortality rates and quality of housing environment. At a global level we identify the ability of G.D.P., smoking habits and alcohol intake to explain mortality rates.

Further research could be done to look at the temporal effects of socio-economic factors at a within-country level. It is also clear that different countries are at different positions on the *mortality spectrum*. Further work could be carried looking at mortality rates between countries to identify the common drivers of mortality improvements be they medical, social or economic and to track these between countries. This may be able to provide an opportunity to plot the future mortality path for current lagging countries on the mortality spectrum.

Table 1.1: U.K. life expectancy from age in the year 2010, 2011, 2021, 2031 and 2035

	Males					Females				
From age	2010	2011	2021	2031	2035	2010	2011	2021	2031	2035
	Period									
0	78.5	78.6	81	82.8	83.4	82.4	82.6	84.7	86.5	87
15	63.9	64.1	66.5	68.2	68.7	67.9	68.1	70.1	71.8	72.3
60	22.1	22.3	24.4	25.8	26.3	24.9	25.1	27	28.5	29
65	18.1	18.3	20.3	21.7	22.1	20.7	20.9	22.7	24.1	24.6
75	11.1	11.3	13.1	14.3	14.6	12.9	13.1	14.7	16	16.4
85	6	6.1	7.3	8.4	8.7	6.9	6.9	8.1	9.2	9.5
	Cohort									
0	90.2	90.3	92	93.5	94.2	93.7	93.8	95.2	96.6	97.2
15	73.4	73.6	75.2	76.7	77.3	77.1	77.3	78.6	80	80.6
60	25.5	25.6	26.9	28.2	28.7	28.5	28.6	29.8	31	31.5
65	21	21.1	22.4	23.5	24	23.7	23.8	25	26.1	26.6
75	12.8	13	14.3	15.3	15.6	14.7	14.9	16.2	17.1	17.5
85	6.5	6.6	8	8.8	9.1	7.3	7.4	8.9	9.8	10.1

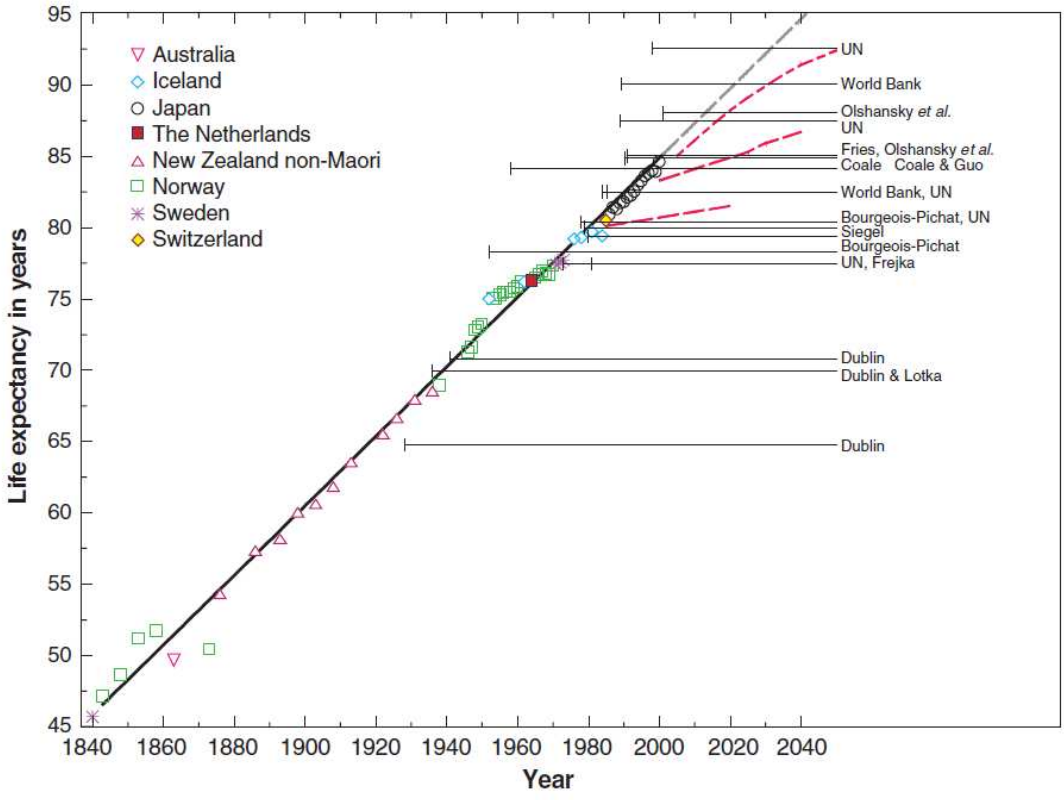

Figure 1.3: Record female life expectancy from 1840 to the present.

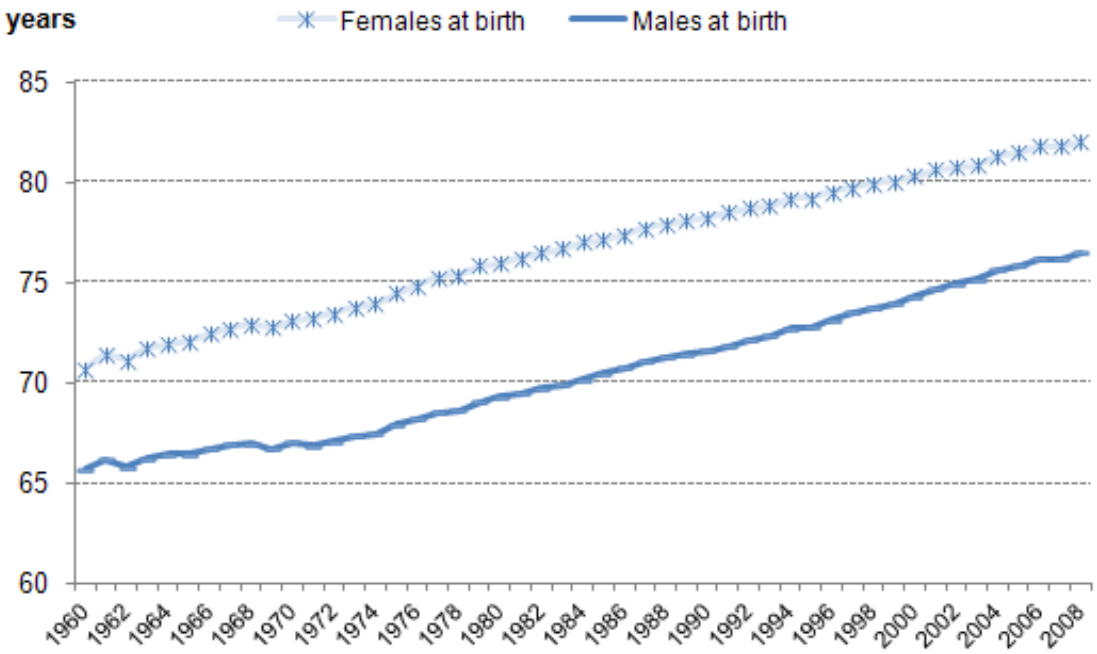

Figure 1.4: Trends in OECD life expectancy at birth: 1960-2008, female and males

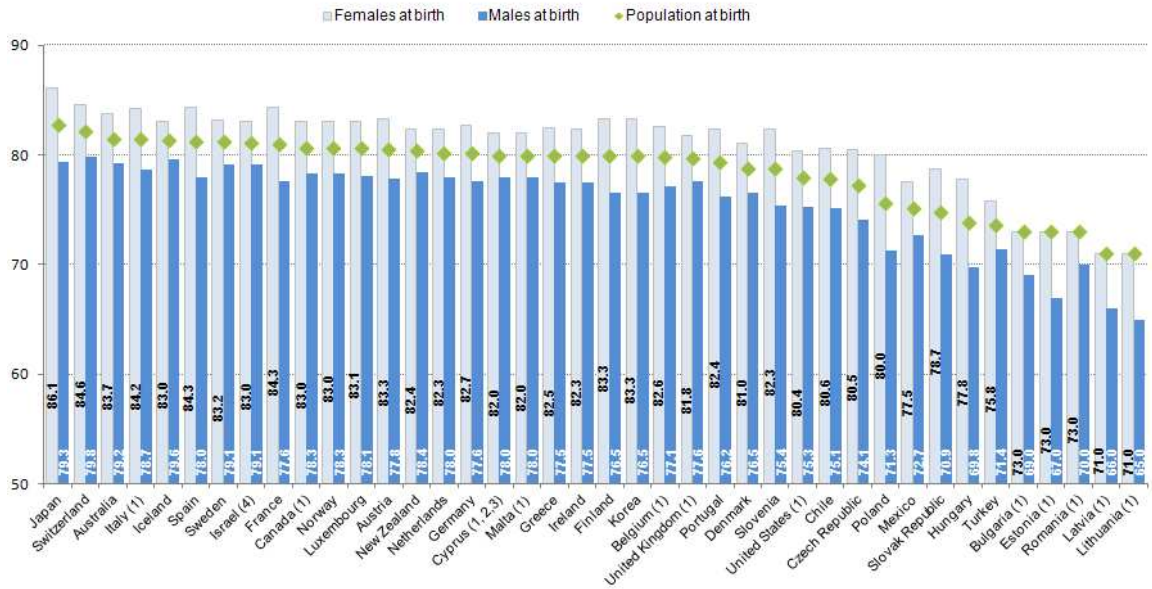

Figure 1.5: Life expectancy at birth, in years, females and males, 2008

Chapter 2

Explaining young mortality

2.1 Introduction

In recent years there has been an increasing amount of attention put on the modelling of mortality risk as a significant risk that pension providers and insurance firms are exposed to. These development have been driven in part by the introduction of more stringent regulation and historically low rates of interest and inflation. The later has exposed longevity risk as being a significant risk in its own right and the development of innovative hedging products has allowed risk holders to unbundle longevity risk from the interest and inflation risks.

There is a significant amount of literature on stochastic modelling of mortality rates. The impetus for the rapid development in stochastic mortality modelling started with the often used model of Lee and Carter (1992) who modeled US male data using a one factor time series approach. Many innovations of the Lee-Carter model have been developed since including, Booth *et al.* (2002), Brouhns *et al.* (2002), Girosi and King (2005), Renshaw and Haberman (2006), Cairns *et al.* (2006), Currie *et al.* (2004), Currie (2006), Hári *et al.* (2008), Tulijapurkar (2008), and Plat (2009).

Many papers propose that mortality in advanced ages is influenced by the mortality experiences at the younger age range and it is clear that the average life expectancy of a population will be affected by experience at all ages. This cumulative effect means that experience at the younger ages is important to consider when modelling the mortality experience of a population. From a demographic viewpoint it is also clear that being able to model and forecast mortality at all ages is important. Hauser and Weir (2011) and Weir (2010) state that greater attention must be given to study designs that allow early-life exposures, experiences, and characteristics to be included in the analysis of outcomes in later life. Cohort effects[1] have been identified as an important component in a mortality model and yet existing models are missing significant information on the most recent cohorts by excluding the younger ages from their models. When we fit existing models to a wider age range starting from age 5 rather than age 20 or 50, the results are not satisfactory[2] since the linear innovations are not flexible enough to capture the non-linear dynamics at the lower ages, the so called "lifestyle" mortality (accidents, drug abuse) profile. In this paper we propose a mortality model that aims to improve upon the fit quality of existing models on a wider age range whilst at the same time not losing sight of the positive aspects of existing models. In particular, using a wider age range introduces a non-linear profile of mortality and we aim to capture this in a better way.

Using the data of a range of developed countries' from 1950 - 2006 we find that the proposed model fits the data very well, is applicable to a fuller age range and captures the cohort effect. It also has a non-trivial correlation structure, captures the non-linear effects at lower ages, has no robustness problems and can take into account parameter

[1]The cohort effect was identified in reports by the Government Actuary's Department (1995, 2001, 2002). These reports highlighted that the generations born between 1925 and 1945 (centered on the generation born in 1931) experienced more rapid improvement than earlier and later generations. This feature had been noted for both males and females in the UK.

[2]We show later in the paper that fitting errors more than double in some cases when a wider age range is fitted. See tables 2.6 and 2.7 model M9 for example.

risk, while the structure of the model remains relatively simple.

The remainder of the paper is organized as follows. First, in Section 2 the background to stochastic mortality modelling is reviewed. In Section 3 an empirical comparison of existing models is conducted which further motivates the paper. In Section 4 a modification of the Plat (2009) model is proposed and its fitting and forecasting performance is assessed using the mortality data of 7 different countries. Conclusions are drawn in Section 5.

2.2 Background

Due to the increasing focus on risk management and measurement for insurers and pension funds, the literature on stochastic mortality models has developed rapidly during the last twenty years. A need to measure the performance quality of these models led to the development of a range of criteria against which models could be assessed. In this section we discuss the background to stochastic mortality modelling starting with the criteria. We follow this with an overview of existing stochastic mortality models up to and including the Plat (2009) model.

In order to assess the quality of a model (from both a fitting and a forecasting perspective) we need to have a range of metrics on which we can quantify the performance of the model. A good set of criteria should allow us to quantify the performance of a mortality model against a range of aspects considered to be "good qualities" for a model of mortality rates. Cairns *et al.* (2011) proposed criteria against which a model can be assessed. For example, the model must fit the existing data well, the model must produce biologically reasonable forecasts etc. Using these criteria we can determine how good a particular model is at fitting and forecasting mortality.

Stochastic mortality models either model the central mortality rate or the initial

mortality rate (see Coughlan *et al.*, 2007). Let $D_{x,t}$ be the number of people with age x that died in year t, and $E_{x,t}$, the exposure being the average population with age x in the year t, the central mortality rate[3] $m_{x,t}$ is defined as:

$$m_{x,t} = \frac{D_{x,t}}{E_{x,t}}, \tag{2.1}$$

The first and most well known stochastic mortality model is that of Lee and Carter (1992):

$$\ln(m_{x,t}) = a_x + b_x\kappa_t + \epsilon_{x,t}, \tag{2.2}$$

where a_x and b_x are age effects and κ_t is a random period effect.[4] Applying the necessary constraints the a_x are given by

$$a_x = \frac{1}{N}\sum_{t=1}^{N} \ln m_{x,t}. \tag{2.3}$$

The bilinear part $b_x\kappa_t$ was then determined as the first singular component of a singular value decomposition (SVD), with the remaining information from the SVD considered to be part of the error structure. The κ_t are estimated and refitted to ensure the model maps onto historic data and the subsequent time series κ_t is used to forecast mortality rates using normal time series forecasting techniques.

Among many discussions of the Lee-Carter model, Cairns *et al.* (2006, 2009, and 2011) summarized the main disadvantages of the model. The model has one factor, resulting in mortality improvements at all ages being perfectly correlated (trivial correlation structure). For countries where a cohort effect is observed in the past, the model

[3]The initial mortality rate q_x is the probability that a person aged x dies within the next year. The different mortality measures are linked by the approximation: $q_x \approx 1 - e^{-m_x}$.

[4]This model was fitted to US mortality data for ages 0-110 between the years of 1933 and 1987.

gives a poor fit to historical data. The uncertainty in future death rates is proportional to the average improvement rate b_x which for high ages can lead to this uncertainty being too low, since historical improvement rates have often been lower at high ages. Also, the model can result in a lack of smoothness in the estimated age effect b_x.

Despite the weaknesses of the Lee-Carter model it's simplicity has led to it being taken as a benchmark against which other stochastic mortality models can be assessed. There has been a significant amount of literature developing additions to, or modifications of, the Lee-Carter model. For example Booth *et al.* (2002), Brouhns *et al.* (2002), Lee and Miller (2001), Girosi and King (2005), De Jong and Tickle (2006), Delwarde *et al.* (2007) and Renshaw and Haberman (2003, 2006).

Mortality data is 2 dimensional with deaths and exposures being recorded by year and by age. We can therefore consider the data from three different viewpoints, the age profile (or how mortality changes from age to age), the time profile (how mortality rates for a specific age change over time), and more recently identified, the cohort profile (how mortality for a specific cohort of the population - those born in a particular year - changes in relation to other cohorts). The Lee Carter model identified the interaction between age and time through the one bilinear factor $b_x\kappa_t$. Many of the modifications since the Lee Carter model have sought to capture additional age-period effects or cohort effects and they can be grouped as such.

2.2.1 Cohort effect additions

Renshaw and Haberman (2006) modified the Lee-Carter model by simply adding a factor γ_{t-x} to capture effects that could be attributed to the year of birth $(t-x)$,

$$\ln(m_{x,t}) = a_x + b_x^1\kappa_t + b_x^2\gamma_{t-x} + \epsilon_{x,t}, \tag{2.4}$$

where κ_t is defined as before and γ_{t-x} is a random cohort effect.

The model does have a much better fit for countries such as the UK where a cohort effect has been identified, however it suffers from a lack of robustness perhaps due to the presence of more than one local maximum in the likelihood function. Among others, for instance Currie (2006) noted that if the model was fitted using data from 1961-2000 then the parameters showed qualitatively different characteristics to those obtained when fitting to data from 1981-2000. Furthermore, as noted by Currie (2006), although the model incorporates the cohort effect, for most of the simulated mortality rates the correlation structure is still trivial with the simulated cohort parameters only being relevant for the higher ages at the far end of the projection.

Following this analysis Currie (2006) applied a simplified age-period-cohort model of Clayton and Schifflers (1987) to mortality which removed the robustness problem but at the expense of the fitting quality:

$$\ln(m_{x,t}) = a_x + \kappa_t + \gamma_{t-x} + \epsilon_{x,t}. \tag{2.5}$$

2.2.2 Age-period effect additions

Cairns *et al.* (2006) observed that for England & Wales and United States data, the fitted cohort effect appeared to have a trend in the year of birth. This suggested that the cohort effect was compensating for the lack of a second age-period effect, as well as trying to capture the cohort effect in the data. This led them to introduce a two factor model of mortality,

$$\text{logit}(q_{x,t}) = \kappa_t^1 + \kappa_t^2(x - \bar{x}) + \epsilon_{x,t}, \tag{2.6}$$

where $\bar{x}$ is the mean age in the sample range and (κ_t^1, κ_t^2) are assumed to be a bivariate random walk with drift. The two factors in this model were both period factors with no cohort effect allowed for. This was rectified in Cairns *et al.* (2009), namely capturing the cohort effect as an additional effect on top of the two age-period effects. All these models have multiple factors resulting in a non-trivial correlation structure which mirrors the reality that improvements in mortality rates are different for different age ranges. A further adaptation was also created allowing for the cohort effect to diminish over time. The main problem with these models arises from the fact that they were designed for higher ages and so ignored the modelling of mortality at the lower ages (for example the accident hump). Cairns *et al.* (2009) argue that the significant cost associated with mortality is at the older ages and thus their modelling focused on those ages. When using these models for full age ranges, the fit quality is relatively poor and the projections are biologically unreasonable.

2.2.3 Age-period and cohort additions combined

Plat (2009) wanted to develop a model which maintained the good aspects of the existing models whilst leaving out the weaker features. The result was a four factor model which took its beginnings from the Lee-Carter model and which added factors to capture the second age-period effect, as per the Cairns *et al.* (2006) model and the cohort effect, as per the Renshaw and Haberman (2006) model. The innovation in the Plat model was to then add a further period factor affecting only the lower ages and designed to allow the model to fit to the whole age range. The model specification is given by:

$$\ln(m_{x,t}) = a_x + \kappa_t^1 + \kappa_t^2(\bar{x} - x) + \kappa_t^3(\bar{x} - x)^+ + \gamma_{t-x} + \epsilon_{x,t}, \tag{2.7}$$

where the a_x is similar to that of the Lee-Carter model and makes sure that the overall shape of the mortality curve by age is reasonable, the κ_t^1 and κ_t^2 model the mortality rates as in the Cairns *et al.* (2006) model and the κ_t^3 models the effects specific to the lower ages only where $(\bar{x} - x)^+$ takes the value $(\bar{x} - x)$ when this is positive and zero otherwise. Finally the γ_{t-x} models the cohort effect.

The range of existing models described above meet most of the criteria set out by Cairns *et al.* (2011) and the Plat model meets all of the criteria by it's very design. However, when the age range is widened to allow for the non-linear characteristics of young mortality experience then as far as we are aware, none of the existing models meet the above criteria adequately (although some are close). This is the starting point of this paper.

2.3 Empirical comparison of existing models

In this section we empirically compare the existing models to see their performance when the age range is widened to allow for the non-linear mortality experience at lower ages. For ease of notation we will use the naming convention established by Cairns *et al.* (2009). Table 2.5 in the appendix sets out the names we will use for each of the models.

We fit the models to different countries and to different age ranges for each country. The data sets[5] used are: Male mortality data during 1950-2006 for age ranges 5-89, 20-89 and 50-89 of Great Britain (GB), England & Wales (E&W), Scotland (SCO), United States (US), Australia (AUS), New Zealand (NZ), and The Netherlands (NL). Although a longer history is available for some of the countries, we have used the period 1950 - 2006 for all the countries as this data is more reliable and will allow a valid

[5]The data consists of numbers of deaths $D_{x,t}$ and the corresponding exposures $E_{x,t}$ and is extracted from: www.mortality.org, see HMD 2004.

comparison with the results of Cairns *et al.* (2009 and 2011), and with Plat (2009) who used the period 1960 - 2006. The model fit is compared using the Mean Average Percentage Error (MAPE) measure and the Bayesian Information Criterion(BIC) measure.

The MAPE measures the average difference in absolute value between $\hat{m}_{x,t}$, the estimate of $m_{x,t}$, and $m_{x,t}$ itself, it is defined by:

$$\text{MAPE} = \frac{1}{NM} \sum_{x,t} \frac{\|\hat{m}_{x,t} - m_{x,t}\|}{m_{x,t}}. \tag{2.8}$$

where we have N time dimensions (in this case N=57) and M age dimensions (in this case M=70). The BIC measure provides a trade-off between fit quality and parsimony of the model and it is defined as:

$$\text{BIC} = L(\hat{\phi}) - \frac{1}{2} K \ln(P), \tag{2.9}$$

where $L(\hat{\phi})$ is the log-likelihood of the estimated parameter $\hat{\phi}$, P is the number of observations and K is the number of parameters being estimated.

Table 2.1 gives a comparison of the fitting results (in terms of MAPE) to the age range 5-89. Tables 2.2 and 2.3 show the fitting results to ages 20-89 and 50-89. We see from tables 2.1 and 2.2 that when a wide age range is used (5-89 or 20-89), the Plat model M9 is not the best fitting model, however, if we exclude model M2, which suffers from robustness issues, the Plat model is confirmed to be the best fitting model over the age range 20-89. When fitting to the age ranges 5-89 and 20-89 it is important to note that the models of Cairns *et al.* (2006, 2009) do not perform very well for these age ranges, since they were designed for higher ages only. For comparison we also fit the existing models to data between 1950 and 2006 for ages 50-89 only. Table 2.3

shows that the Plat model still outperforms other models.

Table 2.1: The MAPE for the model fit to ages 5-89 (%)

Model	M1	M2	M3	M5	M6	M7	M9
GB	6.14	**3.91**	6.96	21.83	16.71	12.88	7.56
E&W	6.38	**4.16**	7.08	21.83	16.87	13.03	7.64
SCO	10.97	**9.28**	12.76	19.99	18.74	15.76	14.72
US	4.58	**2.96**	5.43	16.08	15.59	15.20	5.65
NL	8.99	**7.01**	7.91	23.57	17.82	12.95	7.22
AUS	7.45	**6.44**	8.80	23.86	20.46	18.52	9.61
NZ	12.32	**11.86**	13.66	27.42	25.46	23.84	13.74

Table 2.2: The MAPE for the model fit to ages 20-89 (%)

Model	M1	M2	M3	M5	M6	M7	M9
GB	14.45	**3.19**	14.53	16.53	9.93	7.60	3.27
E&W	14.34	**3.39**	14.42	16.82	10.09	7.73	3.50
SCO	15.67	**6.31**	15.70	16.45	10.32	8.81	**6.31**
US	12.47	**2.46**	12.53	14.07	7.92	6.30	2.76
NL	12.54	**4.16**	12.62	16.14	11.20	8.03	4.22
AUS	5.67	**4.56**	5.84	17.10	10.99	8.40	5.25
NZ	9.57	**8.57**	9.26	19.20	15.32	12.06	9.19

We also look at the fitting results based on the BIC. Tables 2.6, 2.7, and 2.8 in the appendix show the BIC measures for the seven countries, based on fitting to the full age 5-89, the 20-89 age range, and the 50-89 age range, respectively. We see from the tables that it is unclear which model is the best performing using a BIC measure with the Renshaw-Haberman model, M2, showing some good fitting performances, but with models M3, M5, M6, and M9, all performing well on some countries data sets. A particular point to note at this stage (and to motivate the discussion further), is that by widening the age range from 20-89 to 5-89 we can see that for the Plat model for example, the fit quality moves from 3.27% on the 20-89 age range to 7.56% on the 5-89 age range.

To understand why the Plat model does not perform very well for the wider age

Table 2.3: The MAPE for the model fit to ages 50-89 (%)

Model	M1	M2	M3	M5	M6	M7	M9
GB	2.86	1.75	2.00	3.87	1.93	1.53	**1.36**
E&W	2.94	1.87	2.13	4.03	2.02	1.62	**1.48**
SCO	4.05	3.33	3.17	4.57	3.29	3.14	**2.82**
US	2.21	1.47	1.61	2.52	2.09	1.87	**1.41**
NL	4.05	2.39	2.59	4.01	2.19	2.16	**2.07**
AUS	3.18	3.22	3.78	3.62	2.94	2.65	**2.59**
NZ	5.64	5.46	5.94	6.35	5.81	5.78	**5.37**

range and to motivate our further analysis, we look at male data from GB and US. At first, it might be informative to split the data into the period effect and the age effect. Figures 2.1 and 2.2 plot the time effect for GB and US males at ages 15, 35, 55 and 75 with each graph showing the natural logarithm of mortality between the years 1950 and 2006. We see from figures 2.1 and 2.2 that the logarithm of mortality for both GB and US shows a markedly downward trend over time for each of the age ranges, and the mortality looks more volatile at the younger ages, in this case the 15 and 35 year old samples. This might be attributed to the small numbers of deaths at those ages and the fact that deaths at the lower ages are due to a wider range of causes influenced by "lifestyle" choices and so are not linked to general deterioration due to ill health and old age.

Focusing on specific years and looking at the mortality effect for the whole age range, in figures 2.3 and 2.4, we can see that a linear pattern does emerges beyond age 25 or so, however, looking at the mortality below that age we see a very clear non-linear pattern arising. Again this is due to "lifestyle" factors and in order to model these effects we require more flexibility in the factors than the existing model allow.

Looking at the 4 factor model of Eq. (4.6), the design innovation was to include the additional factor $\kappa_t^3(\bar{x} - x)^+$. This factor adds, in a linear way, an additional flexibility for ages less than the mean of the data set. In the case of Plat this would be for ages

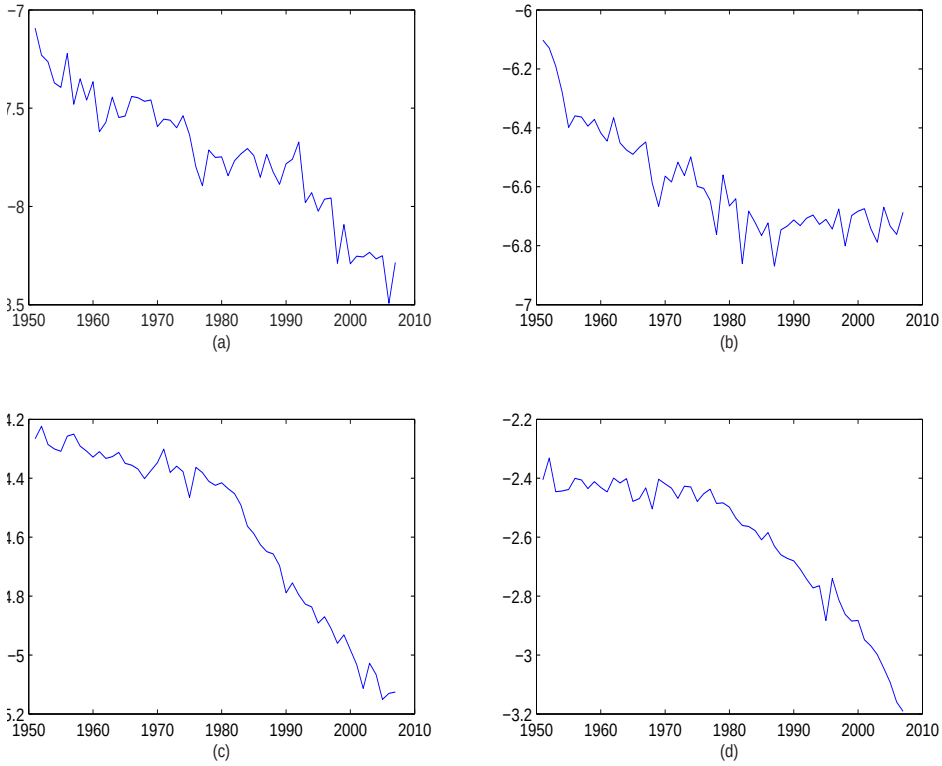

Figure 2.1: Logarithm of mortality by year for GB males aged (a) 15, (b) 35, (c) 55, and (d) 75.

less than 55. Figures 2.3 and 2.4 show clearly that the logarithm of mortality for ages below the mean of 55 are far from linear.

As we have seen from tables 2.1 - 2.3 whilst the Plat model performs relatively well when fit to the data set from age 20, its performance dips somewhat when fitted to the larger data set. In terms of the MAPE when looking at tables 2.1 and 2.2 we find that when the wider age range is fitted the percentage error more than doubles across all countries for which we have fit the model. This implies that the addition of a fourth linear factor is inadequate when modelling mortality at lower ages. In the following section we propose a modification to the Plat model which introduces some additional flexibility into the model allowing it to be more adequately fitted to a wider age range.

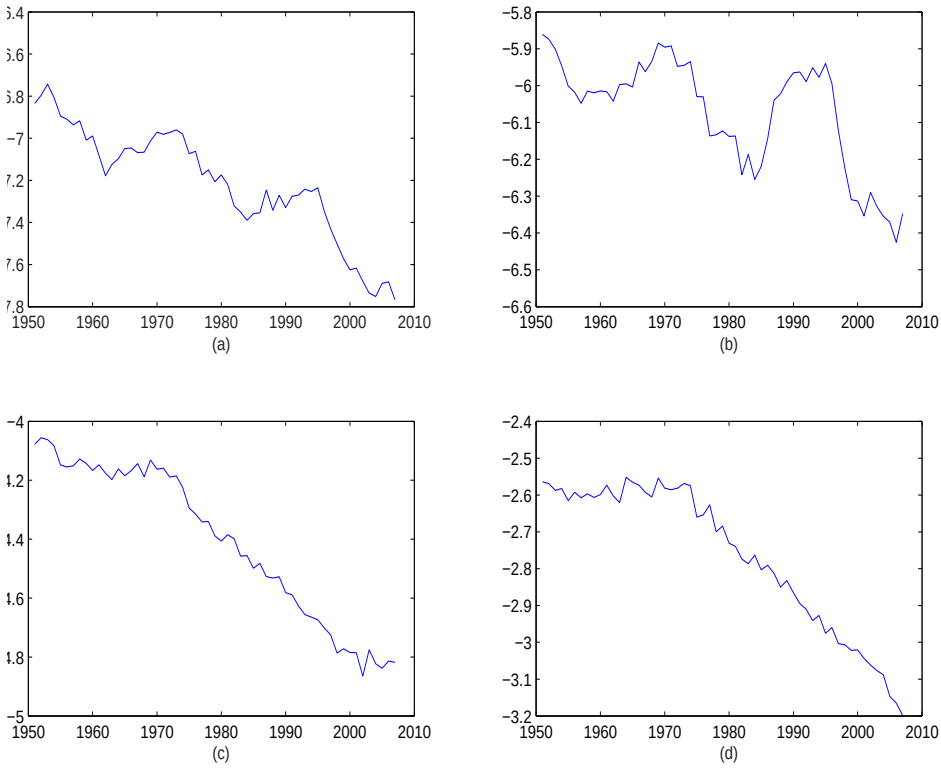

Figure 2.2: Logarithm of mortality by year for US males aged (a) 15, (b) 35, (c) 55, and (d) 75.

2.4 A modification to the Plat model

In this section we incorporate the non-linear features of mortality at younger ages into an adaptation of the Plat model proposing an alternative better fitting model. We show the quality of the fit of the proposed model with that of the existing models by fitting to data from a range of countries for the age ranges 5-89, 20-89 and 50-89 and for years 1950-2006.

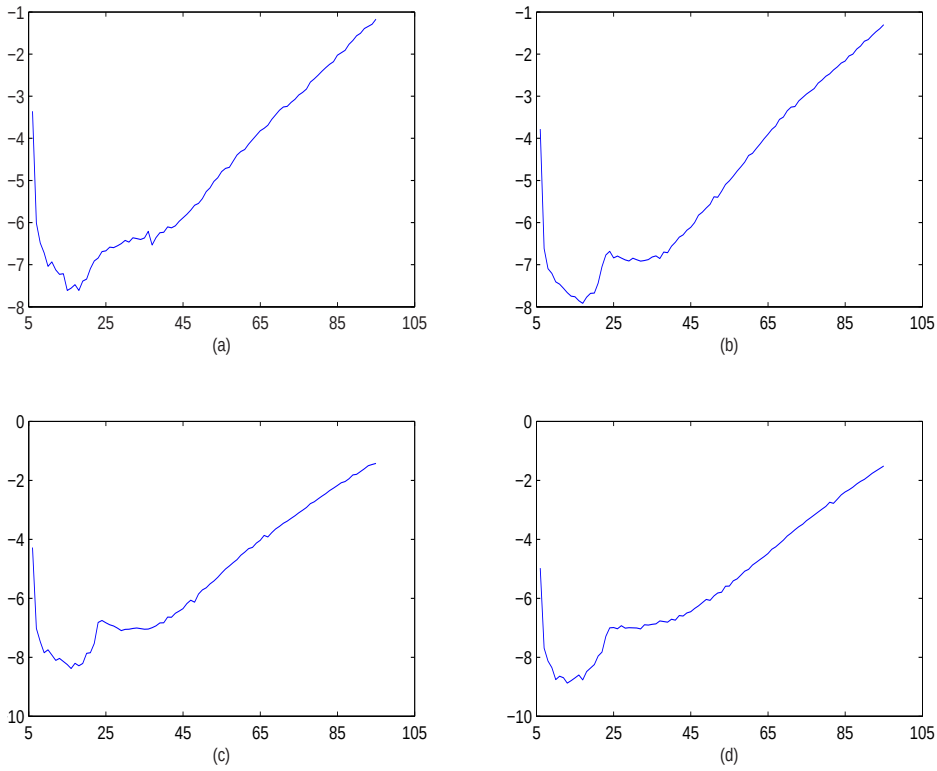

Figure 2.3: Logarithm of mortality for GB males during the years (a) 1950, (b) 1965, (c) 1980, and (d) 2005.

2.4.1 The model

We model the central mortality rate $m_{x,t}$ as:

$$\ln(m_{x,t}) = a_x + \kappa_t^1 + \kappa_t^2(\bar{x} - x) + \kappa_t^3\big((\bar{x} - x)^+ + [(\bar{x} - x)^+]^2\big) + \gamma_{t-x} + \epsilon_{x,t}, \quad (2.10)$$

where a_x makes sure that the basic shape of the mortality curve over ages is in line with historical observations as in the Lee-Carter model (4.1) and the κ_t^1 factor represents changes in the level of mortality for all ages. Following the reasoning in Cairns *et al.* (2006), the (long-term) stochastic process for this factor should not be mean reverting. The κ_t^2 factor allows changes in mortality to vary between ages reflecting the historical

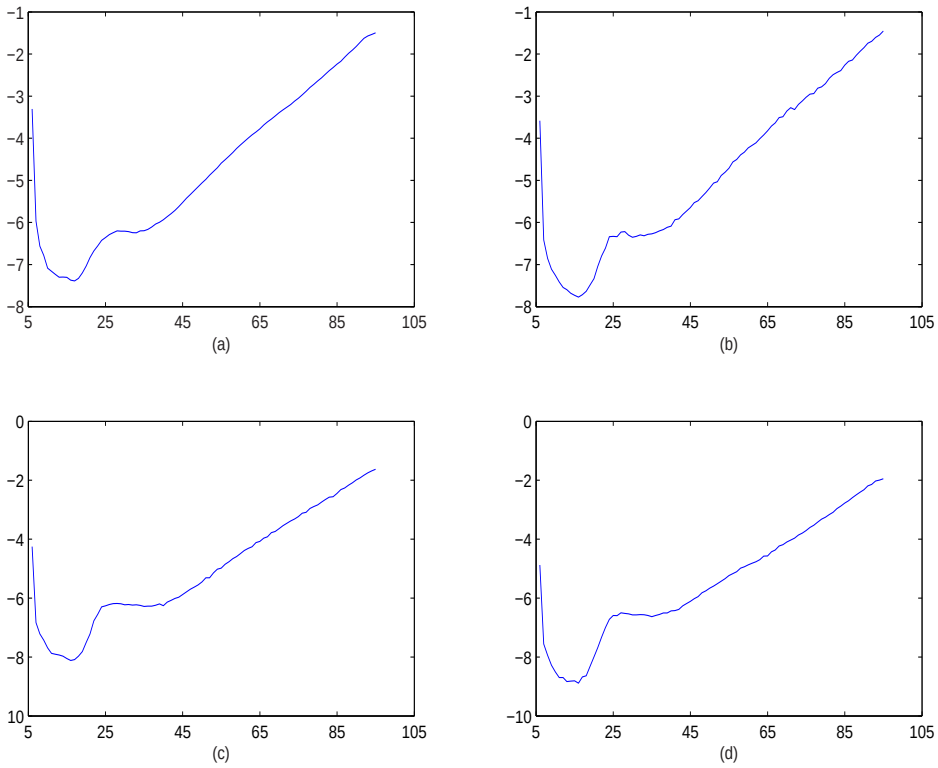

Figure 2.4: Logarithm of mortality for US males during the years (a) 1950, (b) 1965, (c) 1980, and (d) 2005.

observation that improvement rates can differ for different age classes and κ_t^3 models the effects specific to the lower age only as in the Plat model (4.6). The adjusted coefficient of κ_t^3 is designed to capture some of the non-linear effects observed at the lower ages, the "quadratic lower age effect"[6]. Finally the γ_{t-x} models the cohort effect in the same way as the models of Currie (2006) and Cairns *et al.* (2009) and Plat (2009). The proposed model (2.10) has 4 stochastic factors, and so has a relatively simple structure similar to the Plat (2009), Currie (2006) and the Cairns *et al.* (2006)

[6]We also look at the more general case $\ln(m_{x,t}) = a_x + \kappa_t^1 + \kappa_t^2(\bar{x} - x) + \kappa_t^3\big((\bar{x} - x)^+ + a[(\bar{x} - x)^+]^2\big) + \gamma_{t-x} + \epsilon_{x,t}$ where the parameter "a" was included to test a range of different quadratic coefficients. However, we found that the fit quality did not vary much, on both BIC and MAPE for non-zero values of "a", and we therefore focus on a model with a parameter $a = 1$. Results of general "a" are available on request.

models.

Historical data indicates that the dynamics of mortality rates at lower ages (up to age 40 / 50) whilst still showing a downward trend over time does show much more variation around the trend. This can be attributed in part to the small number of deaths and in part to the nature of deaths at these ages, the so called "lifestyle" mortality factors (smoking, drug abuse, alcohol abuse, car accidents and violence) for example. In the Plat model the factor κ_t^3 was added. In model (2.10) we modify the coefficient of κ_t^3 to capture the non-linear dynamics observed in the historical data.

The factor κ_t^1 shows a trend and is fitted with a non-stationary ARIMA process. The factors κ_t^2 and κ_t^3 allow the model to have a non-trivial correlation structure between ages. Fitting a non-stationary ARIMA-process for these factors could result (in some scenarios) in projected scenarios where the shape of the mortality curve over ages is not biologically reasonable. Therefore, a stationary (mean reverting) process will be assumed for these factors. The process for the cohort effect factor γ_{t-x} should not have a trend since we should not expect cohort effects to improve year on year. Therefore, a trendless mean reverting process will be assumed for γ_{t-x}.

As with all stochastic mortality models, the mortality model proposed above has an identifiability problem, meaning that different parameterizations could lead to identical values for $\ln(m_{x,t})$. However, this can be resolved by setting identifiability constraints. As the model has the same time series structure to that of the Plat (2009), following an approach of Cairns *et al.* (2009, model M6), we have

1. $\sum_{c=c_0}^{c=c_l} \gamma_c = 0$
2. $\sum_{c=c_0}^{c=c_l} c\gamma_c = 0$
3. $\sum_t \kappa_t^3 = 0$

where c_0 and c_1 are the earliest and latest year of birth to which a cohort effect is fitted,

and $c = t - x$. These constraints are the same as for the Plat(2009) model as is the rationale behind the choice of the constraints.

Fitting methodology — The original method by which to fit such a stochastic model was to use SVD as used in Lee and Carter (1992). Brouhns *et al.* (2002) described an alternative fitting methodology for the Lee-Carter model in which the number of deaths $D_{x,t}$ is modeled as a Poisson distribution with parameter $(E_{x,t}m_{x,t})$ where $m_{x,t}$ is the mortality rate we are estimating. The main advantage of the Brouhns *et al.* (2002) approach over the SVD approach is that it accounts for the heteroscedasticity of the mortality data for different ages. Indeed this method has been used more commonly, see for example Renshaw and Haberman (2003, 2006) and Cairns *et al.* (2009), Plat (2009). We adopt this approach and model the number of deaths by $D_{x,t} \approx \text{Poisson}(E_{x,t}m_{x,t})$. The parameters of model (2.10) are estimated by maximizing the log-likelihood function[7]:

$$L(\phi; D, E) = \sum_{x,t} D_{x,t} \ln[E_{x,t}m_{x,t}(\phi)] - E_{x,t}m_{x,t}(\phi) - \ln(D_{x,t}!). \tag{2.11}$$

Besides estimates for a_x, the fitting procedure described above leads to time series of estimations of κ_t^1, κ_t^2, κ_t^3, and γ_{t-x}. After fitting the model we take the fitted values for the time series and fit suitable ARIMA-processes.

2.4.2 Comparison of fit quality with existing models

To evaluate whether the proposed model fits historical data well, we fit the model to the data sets described in Section 3. We also fit the model to the three different age ranges, 5-89, 20-89 and 50-89 to show the flexibility of the proposed model. The fitting quality

[7]We used an adaptation of the R-code of the software package "Lifemetrics" which is an open source toolkit for measuring and managing longevity and mortality risk, designed by J.P. Morgan, see http://www.lifemetrics.com and http:// www.r-project.org/.

for each of the countries, using a MAPE and BIC measure are presented in table 2.4.

Table 2.4: MAPE and BIC results for model M10

Country	MAPE			BIC		
	5-89	20-89	50-89	5-89	20-89	50-89
GB	5.88	2.77	1.29	-28326	-21887	-13148
E&W	6.05	2.79	1.37	-27549	-21634	-13074
SCO	11.88	2.79	2.76	-18586	-15768	-10096
US	4.14	2.31	1.35	-50487	-29998	-18727
NL	6.11	2.67	1.99	-19586	-16729	-10601
AUS	8.14	2.65	2.48	-22466	-18327	-10956
NZ	13.65	2.69	5.39	-17272	-14714	-9433

When a wider age range is used the logarithm of mortality is no longer relatively linear. However, when comparing the results with table 2.1 (Excluding the results from model M2,the Renshaw-Haberman model, because of robustness problems) we see that this non-linearity is captured adequately by the quadratic lower age effect in the proposed model. Across all countries considered in this paper the proposed model fits the data better than the previous best fitting models. Looking at the results when compared with table 2.2 the performance of the model is still very good when compared with the leading stochastic models of mortality. Comparisons with the results of table 2.3 show that the model still outperforms the existing stochastic models for the age range 50-89. As the improved specification has been done within a 4 factor framework this model has a similar structure to the previously best performing model on a fitting measure, namely the Plat model. Thus the model remains relatively parsimonious and this is reflected in the BIC measures in table 2.4 when compared with tables 2.6, 2.7, and 2.8 in the appendix.

The goodness of fit of stochastic mortality models can be evaluated by analyzing residuals of the models, Dowd *et al.* (2010a) applied the t-test, variance ratio test, and the Jarque-Bera test among others to six stochastic mortality models (M1, M2, M3, M5, M6 and M7) using the English and Welsh male mortality data. We carried out

similar tests for model M10 using US and GB data. Results[8] of these tests show that the proposed M10 model performs adequately when compared to those in the Dowd *et al.* (2010a).

Fitting the ARIMA processes — In the remainder of this subsection, we focus on the populations of GB and US males and on fitting to the age range 5-89. After fitting the model to the population data the next step is to select and fit suitable ARIMA-process to the time series' of κ_t^1, κ_t^2, κ_t^3, and γ_{t-x}. The fitted parameters κ_t^1, κ_t^2, κ_t^3, and γ_{t-x} for GB males are given in figure 2.5 and for US males are given in figure 2.7. The estimates for the α_x parameters are given in figure 2.6 and figure 2.8. The figures shows that the pattern of the important parameter κ_t^1 is well-behaved. The patterns of the other parameters all reveal some autoregressive behavior. Since the factor κ_t^1 drives a significant part of the uncertainty in mortality rates, its relatively regular behavior (for this particular dataset) will also show in the relatively narrow confidence intervals.

The parameters for the Plat model are plotted in the appendix as figures 2.11, 2.12, 2.13 and 2.14 for comparison purposes. They show that the qualitative characteristics of the parameters κ_t^1, κ_t^2, κ_t^3, and γ_{t-x} remain unchanged with the more general model specification.

It is commonly assumed that the time series driving the dynamics, namely κ_t^1 should be fitted with an ARIMA(0,1,0) time series. For the other parameters, which show some autoregressive behavior, we have fit them with ARIMA(1,0,0) processes as in Plat (2009). It is also commonly assumed (see Renshaw and Haberman (2006), CMI (2007) and Cairns *et al.* (2011)) that the process for γ_{t-x} is independent of the other processes, so the parameters of this process can be fitted independently using Ordinary Least Squares. The other processes can be fitted simultaneously using Seemingly Unrelated Regression.

[8] The results are available upon request.

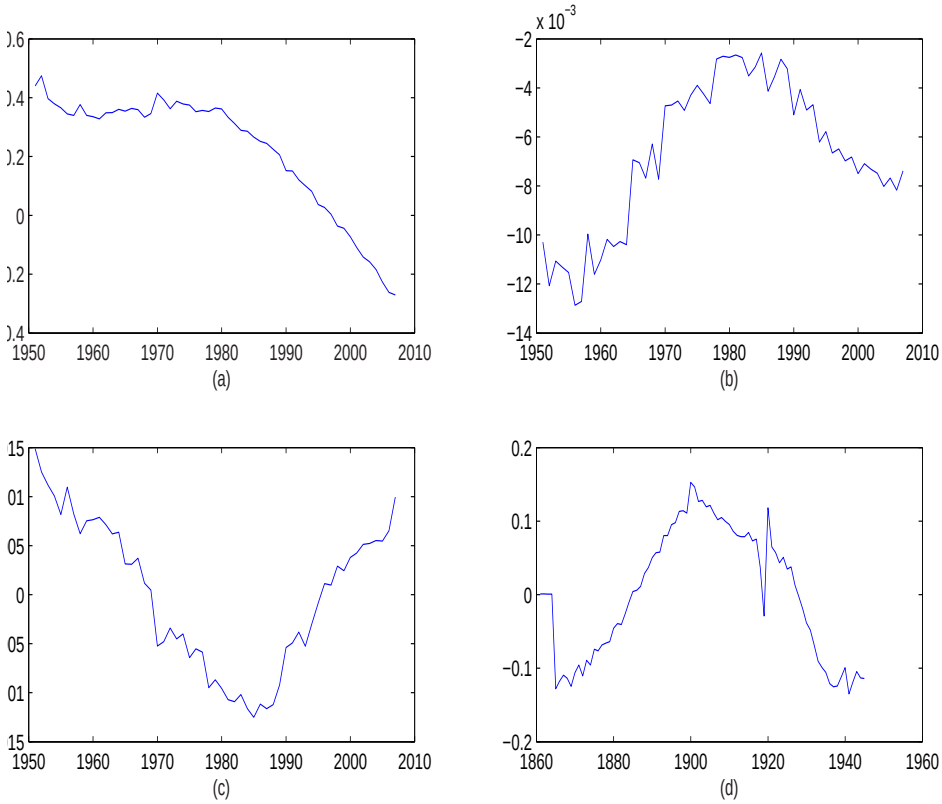

Figure 2.5: Estimated values of (a) κ_t^1, (b) κ_t^2, (c) κ_t^3, and (d) γ_{t-x} based on GB males aged 5-89 between years 1950 and 2006.

2.4.3 Forecasting

This section shows the simulation results and results of robustness tests for the proposed mortality model.

Using the fitted ARIMA processes and the fitted values for a_x and γ_{t-x} (see figures 2.5, 2.6, 2.7 and 2.8[9]), future mortality rate scenarios can be constructed using Monte Carlo simulation. Figures 2.9 and 2.10 show simulation results for ages 15, 35, 55 and 75 for GB males and US males.

For higher ages, the widths of the confidence intervals are broadly similar as the

[9]The fitted values for a_x and γ_{t-x} for England & Wales, Scotland, Netherlands, Australia and New Zealand are available in the appendix in figures 2.15 and 2.16.

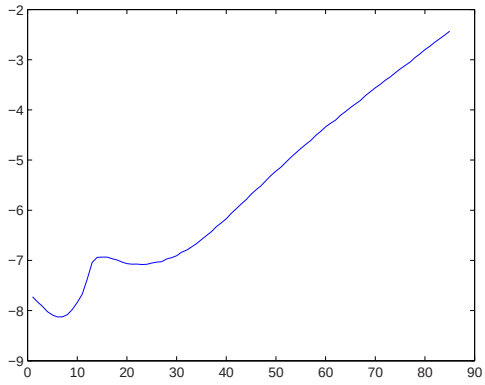

Figure 2.6: Estimated values of a_x based on GB males aged 5-89 between years 1950 and 2006.

models of Plat (2009) and Cairns *et al.* (2011) confirming the results are biologically plausible. The results for younger ages (15 and 35) also seem plausible, where the observed historical variability is reflected in the wider confidence intervals.

Recall that some models suffer from a lack of robustness, for instance the Renshaw-Haberman model is not robust for changes in range of years. The model proposed in this paper is tested for robustness by fitting the model to data from 1975-2006. In doing this we are looking to observe that the qualitative characteristics of the fitted parameters have not changed because the fitting period is different. We are not looking to show that the trend direction is unchanged, or that the actual forecasts are unchanged. It is a characteristic of these sorts of models that the forecasted trend will to an extent be dependent on the period over which the model has been fitted to the data. Given that it is likely the trend forecast will be different when fit to the period 1975-2006 compared to 1950-2006, it is inevitable, for all models, that the simulation results will be somewhat different.

Figures 2.17 and 2.18 in the Appendix plot the fitted parameters for GB and US data from 1975-2006. The illustrations show that the estimated parameters do not show significantly different qualitative characteristics when fitted to a different data set. The

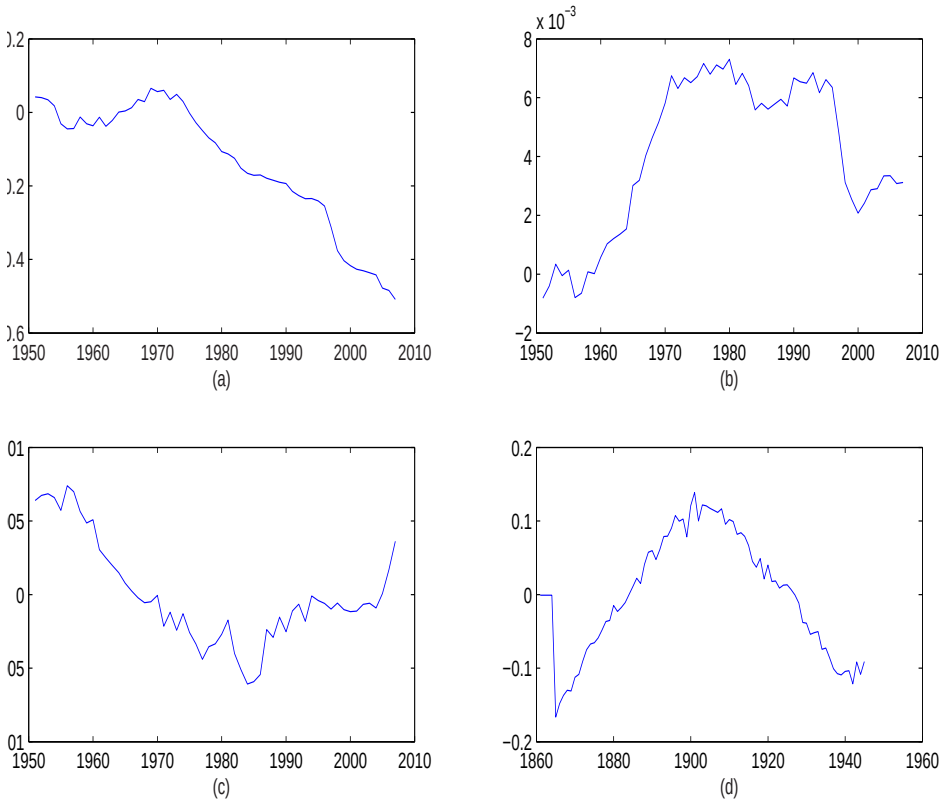

Figure 2.7: Estimated values of (a) κ_t^1, (b) κ_t^2, (c) κ_t^3, and (d) γ_{t-x} based on US males aged 5-89 between years 1950 and 2006.

conclusion is that the proposed model is robust for the fitting periods given above.

Furthermore, backtesting (as in Dowd *et al.* 2010) of the model has been carried out, meaning that the model is fitted to historical data, 1950-1999 in this case, and the forecast results are compared with the actual observations for the period 2000-2006. The results are illustrated in figure 2.19 in the Appendix where we can see that the proposed model performs adequately.

We have shown so far that the proposed model produces plausible results and they seem robust. Plat (2009) came to the same conclusion for model M9 and Cairns *et al.* (2011) came to the same conclusion for the models of Currie (2006) and Cairns *et al.* (2006, 2009), M7. The models of Cairns *et al.* (2006, 2009) are designed for higher

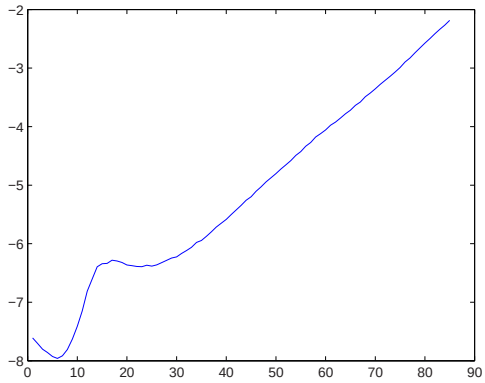

Figure 2.8: Estimated values of a_x based on US males aged 5-89 between years 1950 and 2006.

ages, so will not produce plausible results for lower ages. Compared to those models the proposed model has the advantage that it does produce plausible results for a full age range.

Compared to the model of Currie (2006) the proposed model has the advantage that it has a non-trivial correlation structure. This is important because often insurers and pension funds have different type of exposures for younger or middle ages (term insurance, pre-retirement spouse option) than for higher ages (pensions, annuities). Aggregating these different types of exposures can only be done sufficiently if the model has a non-trivial correlation structure. Assuming an almost perfect correlation between ages, as in the Currie (2006) model, will possibly lead to an overstatement of the diversification benefits that arise when aggregating these exposures. Compared to the model of Plat (2009) the proposed model produces plausible forecasts for the lower age range (below age 20) for which the Plat model was not designed.

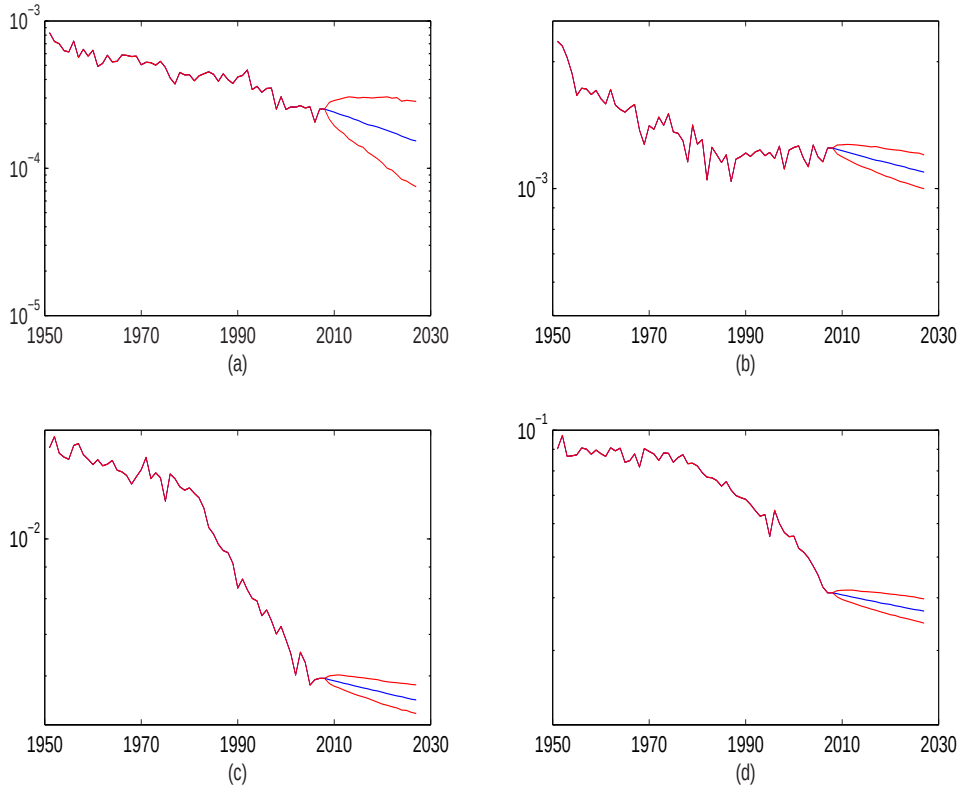

Figure 2.9: Log mortality rates from 1950-2006 followed by forecasting results 2006 - 2026 (mean and 95% confidence intervals) for ages (a) 15, (b) 35, (c) 55, and (d) 75 for GB males.

2.5 Conclusions

In this paper we identify and address a limitation of the Plat (2009) model and previous stochastic mortality models. This limitation is in the inability of existing models to adequately fit mortality rates at the lower ages due to the non-linear dynamics at the lower ages, the so called “lifestyle” mortality profile. We believe that it is important to be able to factor in such mortality rates into a single mortality model because of the cumulative nature of mortality and from a demographic viewpoint it is clearly important to be able to model and forecast mortality rates at all ages. The proposed model has the additional flexibility to fit to the mortality rates of a wider age range,

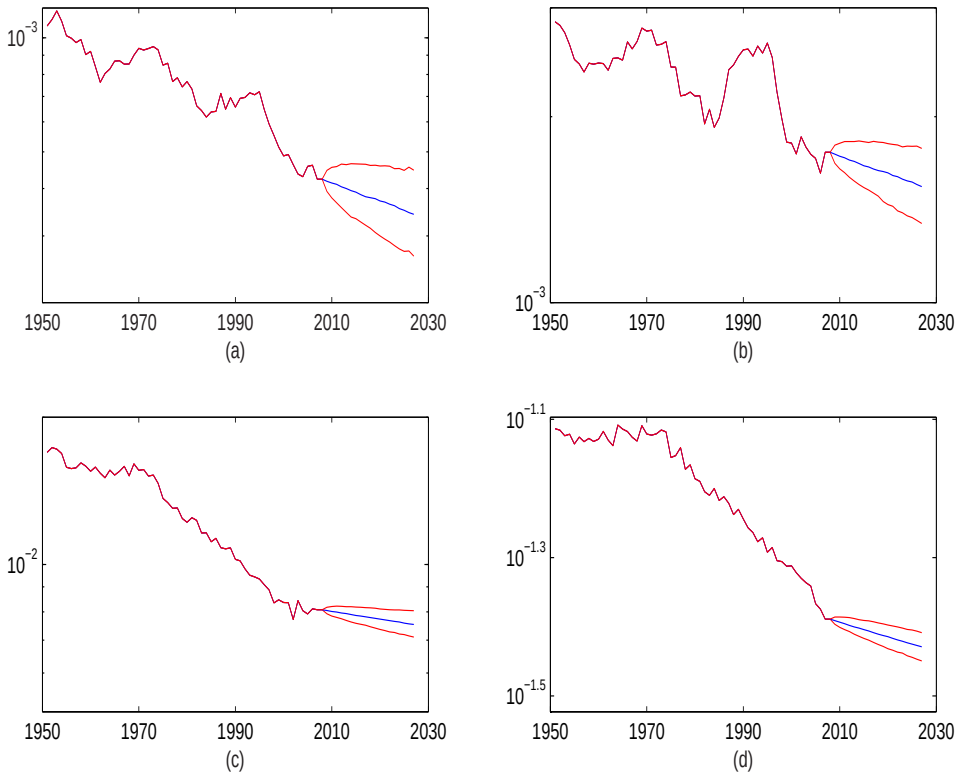

Figure 2.10: Log mortality rates from 1950-2006 followed by forecasting results 2006 - 2026 (mean and 95% confidence intervals) for ages (a) 15, (b) 35, (c) 55, and (d) 75 for US males.

5-89. In particular, the model captures the non-linear profile of mortality at lower ages. We show that the model has a better fit for the range of countries considered in this study. We have also shown that the model does not lose any of the benefits of the previous stochastic models.

The results of this analysis have exposed the weakness of previous models when trying to fit to non-linear features of the data and shows that a more non-linear flexibility is needed to capture the mortality profile, particularly at lower ages. To develop this area further we now need to address the “lifestyle” factors affecting mortality rates in this age range. These may be affected by policy, social, environmental and economic

pressures suggesting that a future approach may be to model the underlying causes rather than by trend forecasting.

2.6 Appendix: Additional Figures and Tables

Table 2.5: The names of stochastic mortality models

Name	Model and Name
M1	Lee and Carter (1992)
	$\ln(m_{x,t}) = a_x + b_x\kappa_t + \epsilon_{x,t}$
M2	Renshaw and Haberman (2006)
	$\ln(m_{x,t}) = a_x + b_x^1\kappa_t + b_x^2\gamma_{t-x} + \epsilon_{x,t}$
M3	Currie (2006)
	$\ln(m_{x,t}) = a_x + \kappa_t + \gamma_{t-x} + \epsilon_{x,t}$
M5	Cairns *et al.* (2006)
	$\text{logit}(q_{x,t}) = \kappa_t^1 + \kappa_t^2(x - \bar{x}) + \epsilon_{x,t}$
M6	Cairns *et al.* (2009) with cohort effect
	$\text{logit}(q_{x,t}) = \kappa_t^1 + \kappa_t^2(x - \bar{x}) + \gamma_{t-x} + \epsilon_{x,t}$
M7	Cairns *et al.* (2009) with cohort and quadratic age effect
	$\text{logit}(q_{x,t}) = \kappa_t^1 + \kappa_t^2(x - \bar{x}) + \kappa_t^3((x - \bar{x})^2 - \sigma_x^2) + \gamma_{t-x} + \epsilon_{x,t}$
M9	Plat (2009)
	$\ln(m_{x,t}) = a_x + \kappa_t^1 + \kappa_t^2(\bar{x} - x) + \kappa_t^3(\bar{x} - x)^+ + \gamma_{t-x} + \epsilon_{x,t}$
M10	Quadratic effect model
	$\ln(m_{x,t}) = a_x + \kappa_t^1 + \kappa_t^2(\bar{x} - x) + \kappa_t^3\big((\bar{x} - x)^+ + [(\bar{x} - x)^+]^2\big) + \gamma_{t-x} + \epsilon_{x,t}$

Note: The model M4 and M8 are not included in our analysis. The M4 is a P-splines model developed in Currie (2006), it is of a structurally different nature to the remaining stochastic models. The M8 in Cairns *et al.* (2009) with diminishing cohort effect is a modification of the M5, it was primarily designed for ages over and above 50. The M10 is the model that we propose in this paper.

Table 2.6: The BIC for the model fit to ages 5-89

Model	M1	M2	M3	M5	M6	M7	M9
GB	-38228	**-28854**	-34181	-150856	-96992	-57365	-33640
E&W	-36686	**-28315**	-32984	-136658	-88876	-53211	-32295
SCO	-20960	**-20633**	-20787	-29413	-25752	-22671	-20998
US	-72612	**-43997**	-69820	-552628	-271679	-258323	-66989
NL	-24914	**-22122**	-22516	-55568	-37711	-26912	-22178
AUS	-24340	**-23217**	-25594	-82648	-44443	-29848	-25692
NZ	**-17842**	-18288	-18154	-31360	-23208	-22149	-18284

Table 2.7: The BIC for the model fit to ages 20-89

Model	M1	M2	M3	M5	M6	M7	M9
GB	-33926	**-24684**	-27031	-91937	-58540	-39770	-24921
E&W	-32516	**-24236**	-26558	-83889	-54516	-37511	-24551
SCO	-18326	-17904	**-17575**	-22881	-19994	-18897	-17689
US	-64565	**-37863**	-56350	-368252	-143067	-97548	-43425
NL	-20928	-19012	-18980	-32420	-26601	-21424	**-18778**
AUS	-20833	**-19909**	-21449	-57681	-30301	-23740	-20697
NZ	**-15282**	-15714	-15484	-22794	-18177	-16394	-15818

Table 2.8: The BIC for the model fit to ages 50-89

Model	M1	M2	M3	M5	M6	M7	M9
GB	-20246	-15074	-16493	-26013	-15756	-14874	**-14834**
E&W	-19591	-14891	-16301	-25094	-15524	-14817	**-14759**
SCO	-11394	-11247	**-11050**	-11675	-11111	-11146	-11259
US	-29981	**-20581**	-22653	-35419	-27749	-22562	-21598
NL	-13344	-11902	-11910	-13009	**-11684**	-11786	-11909
AUS	-12337	**-12187**	-12864	-12798	-12324	-12306	-12278
NZ	**-9534**	-9873	-9801	-9698	-9847	-9993	-9984

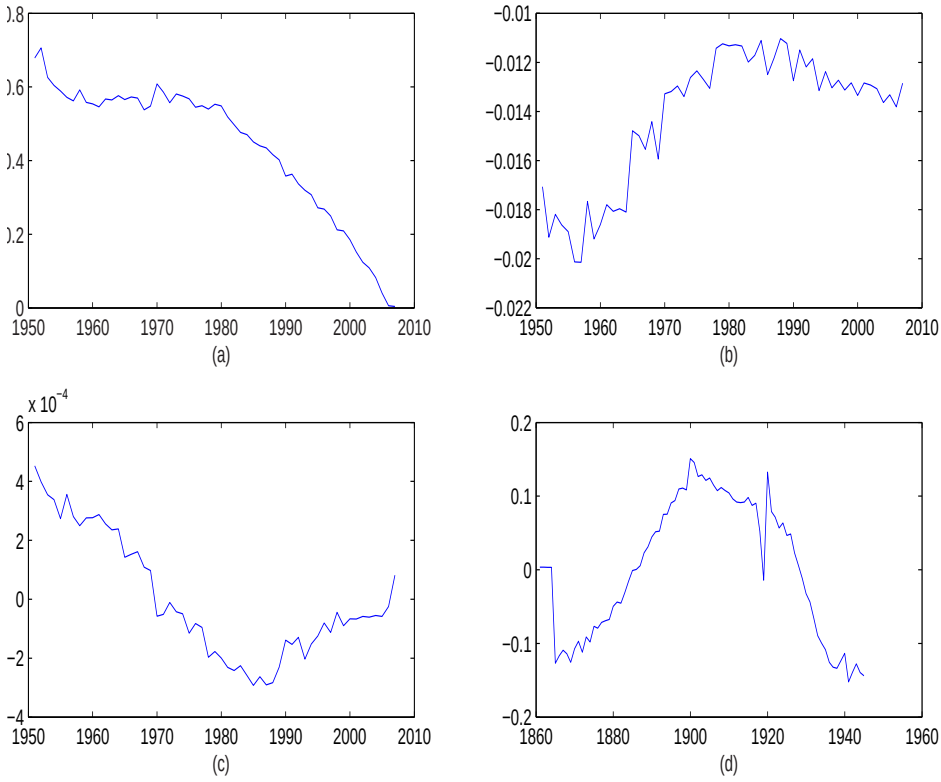

Figure 2.11: Estimated values of (a) κ_t^1, (b) κ_t^2, (c) κ_t^3, and (d) γ_{t-x} based on GB males aged 5-89 between years 1950 and 2006 for the Plat model.

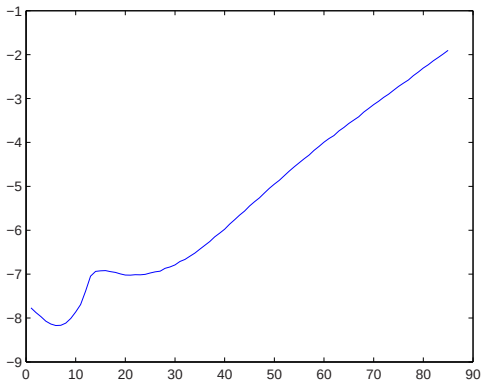

Figure 2.12: Estimated values of a_x based on GB males aged 5-89 between years 1950 and 2006 for the Plat model.

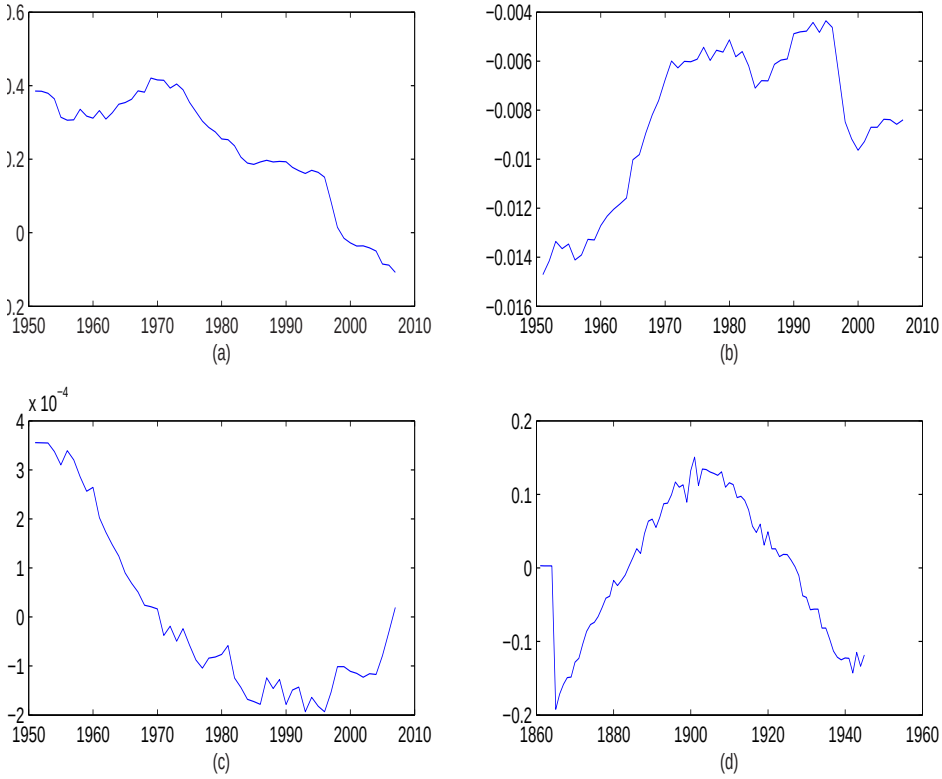

Figure 2.13: Estimated values of (a) κ_t^1, (b) κ_t^2, (c) κ_t^3, and (d) γ_{t-x} based on US males aged 5-89 between years 1950 and 2006 for the Plat model.

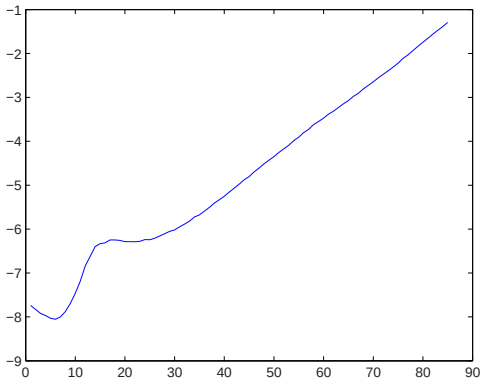

Figure 2.14: Estimated values of a_x based on US males aged 5-89 between years 1950 and 2006 for the Plat model.

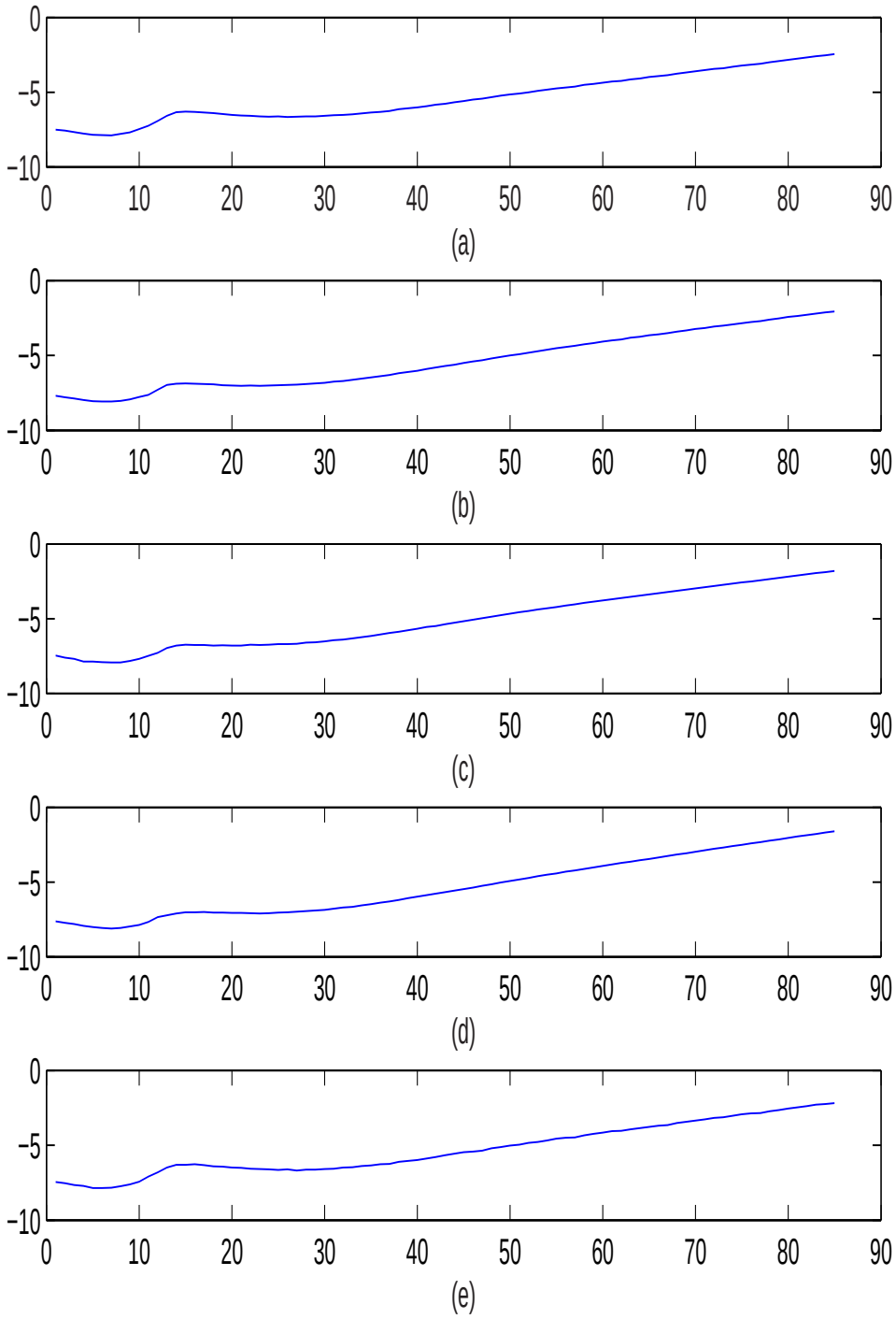

Figure 2.15: Estimated values of a_x based on ages 5-89 for countries (a) Australia, (b) England and Wales, (c) Scotland, (d) New Zealand and (e) Netherlands.

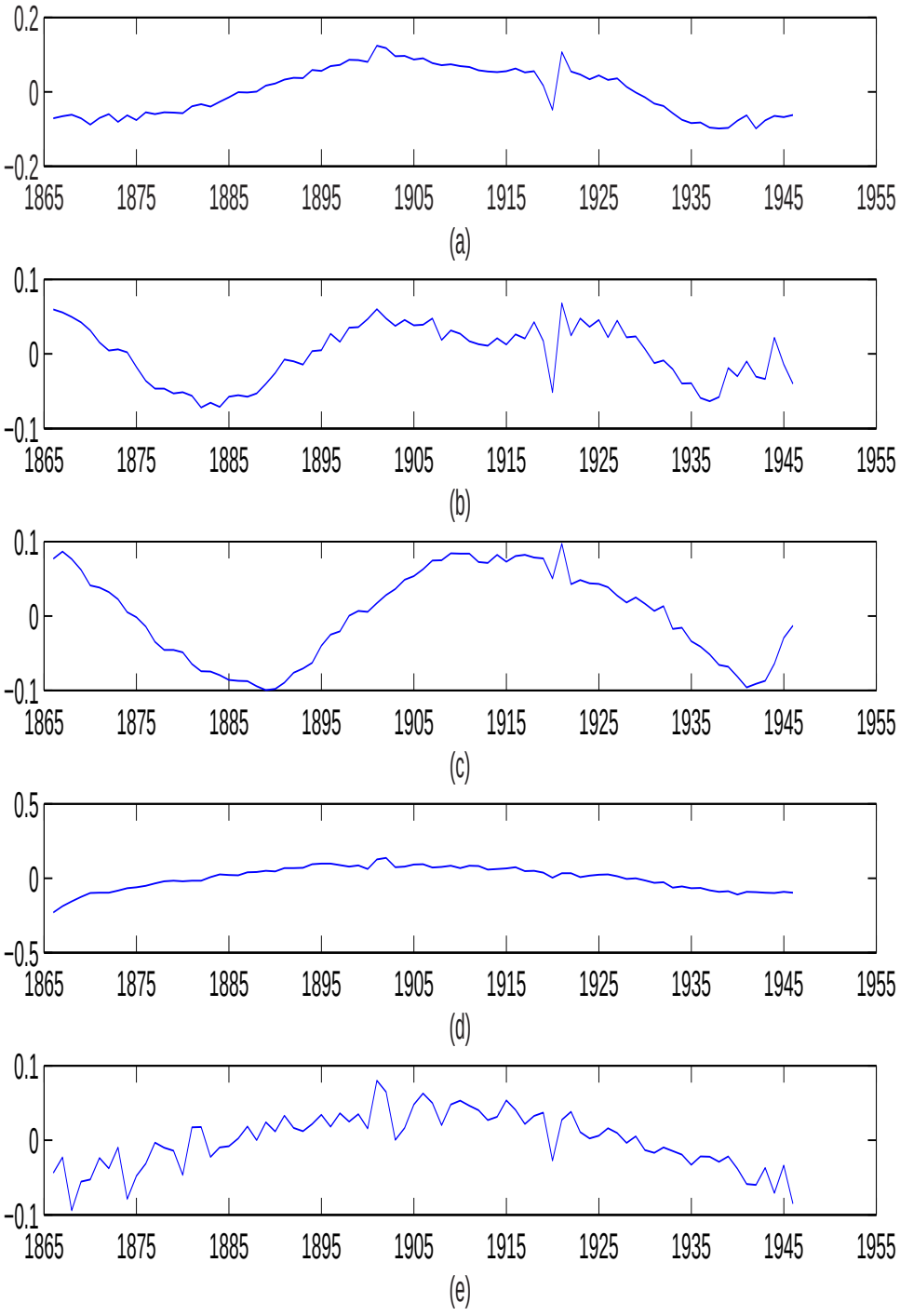

Figure 2.16: Estimated values of γ_{t-x} based on year of birth 1865-1955 for countries (a) England and Wales, (b) Scotland, (c) Netherlands, (d) Australia and (e) New Zealand.

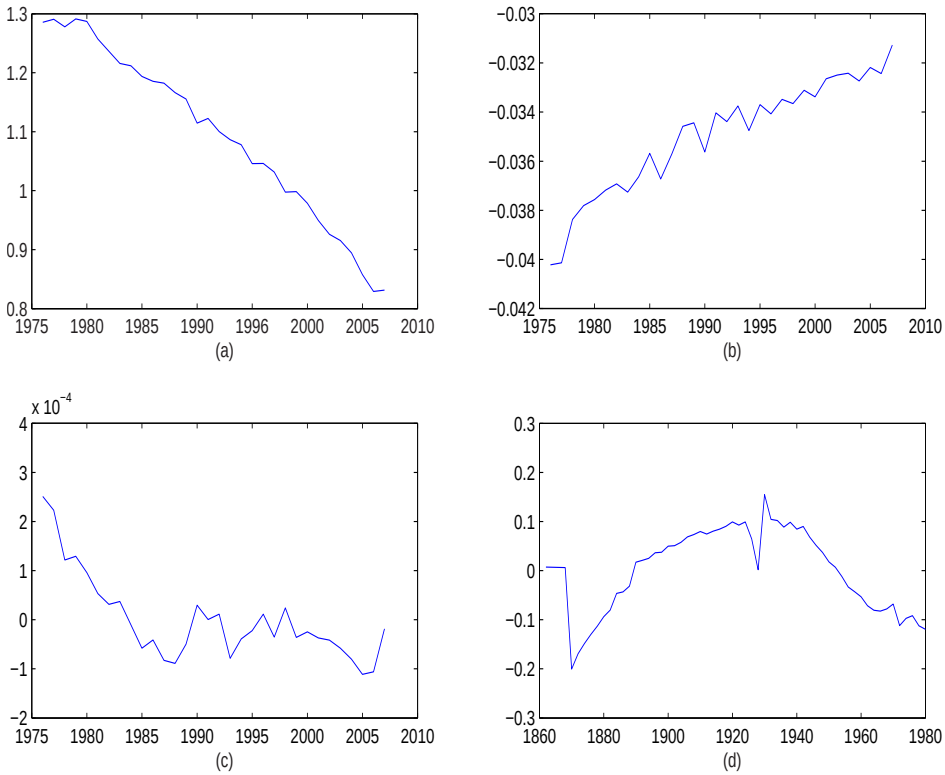

Figure 2.17: GB fitted parameters (a) κ_t^1, (b) κ_t^2, (c) κ_t^3, and (d) γ_{t-x} with data from 1975-2006.

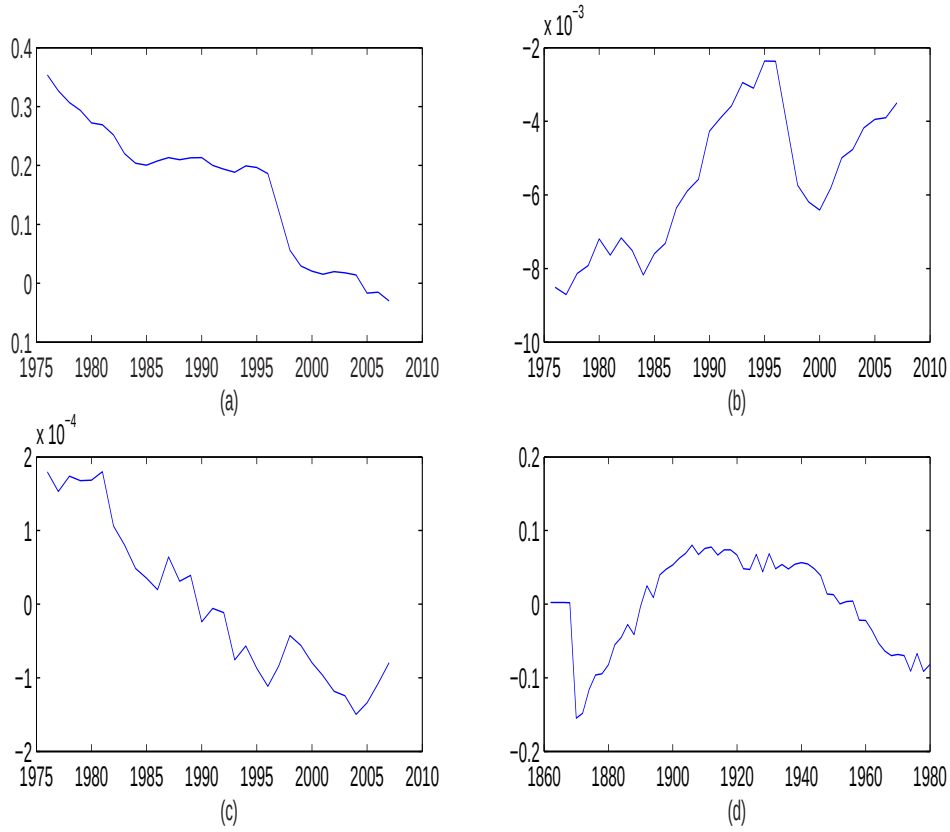

Figure 2.18: US fitted parameters (a) κ_t^1, (b) κ_t^2, (c) κ_t^3, and (d) γ_{t-x} with data from 1975-2006.

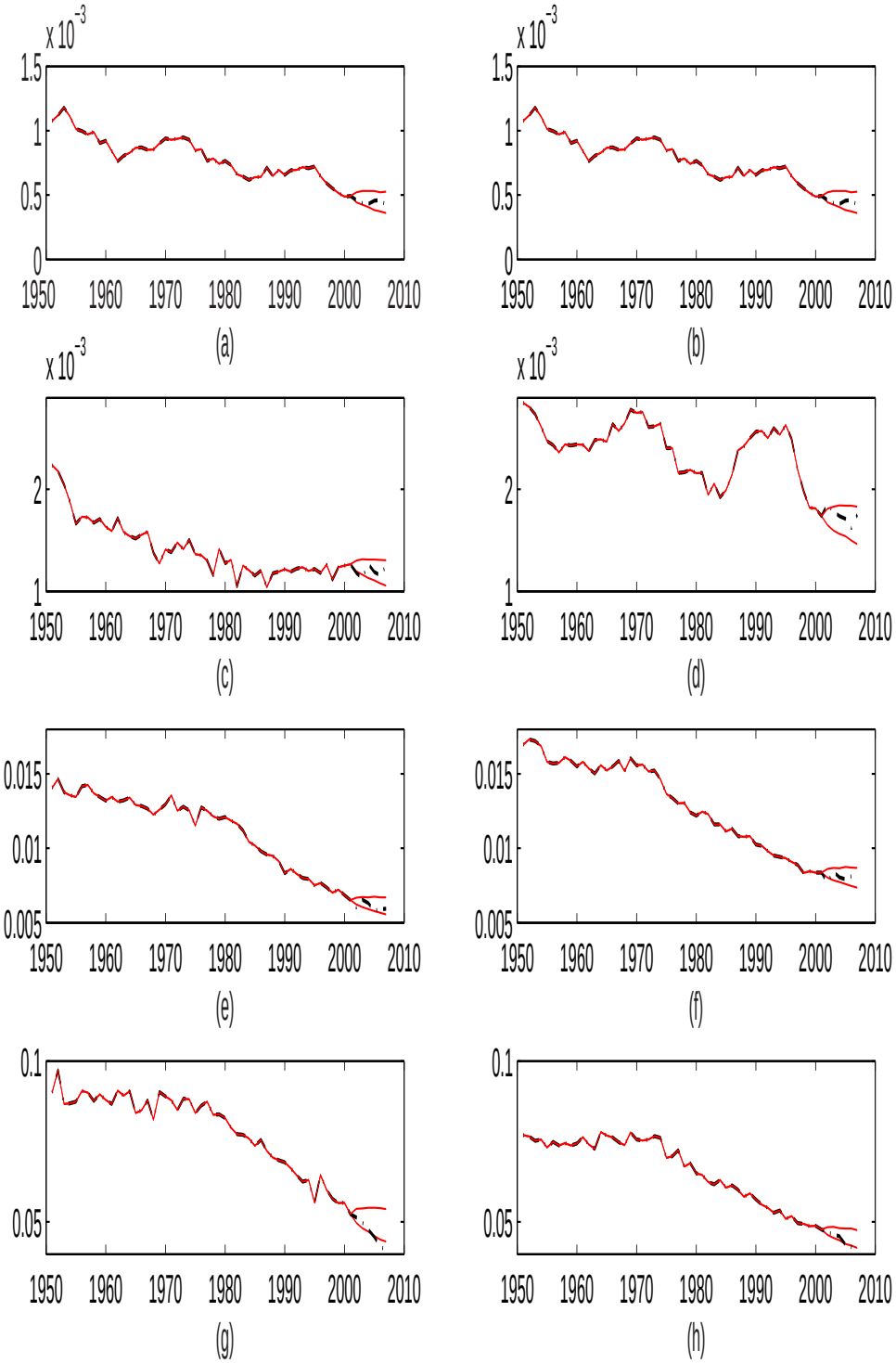

Figure 2.19: Log mortality rates from 1950 - 2006 plotted with 95% confidence intervals from 2000-2006 based on fitting from 1950-2000. Plots show ages (a) and (b) 15, (c) and (d) 35, (e) and (f) 55, and (g) and (h) 75 for countries GB and US respectively.

Chapter 3

A dynamic factor approach to mortality modelling

3.1 Introduction

Defined benefit pension schemes receive contributions from individuals over their working life. In exchange they provide a pension in retirement for the lifetime of those members. The pension scheme trustees will invest contributions to generate a return and draw from their total funds to make pension payments as and when they are necessary. There are three major risks faced by a pension provider when offering such a product which are interest rate risk, inflation risk and longevity risk. An annuity product is a promise to pay a series of payments to an individual for their lifetime and it is priced incorporating each of these risks (see Dickson *et al*,2009). The pricing of longevity risk relies on accurate mortality forecasts. In this paper, following Forni et al. (2005) we propose a dynamic factor model of mortality which provides more accurate forecasts than commonly used benchmark models.

Interest rate risk exists since members' benefits are defined in advance and so the

contributions received plus any investment return have to be sufficient to cover these benefits. If interest rates fall below a forecasted level then the funds of the scheme may not grow sufficiently large to be able to cover the benefits promised and the trustees may no longer be able to pay those benefit. Inflation is a risk since many defined benefit pension schemes offer benefits that are protected in real terms. This means that during periods of high inflation the benefits that must be paid will be higher. The risk absorbed by the pension provider is that the assets held in the fund will grow equally as the benefits during periods of high inflation. Financial innovation in the areas of interest rate and inflation derivative markets have led to interest rate and inflation risk becoming less of an issue for pension providers since those risks can now be successfully hedged against using appropriate investment strategies. Since pensions are paid for the lifetime of a member the amount of funds paid out to an individual will be directly associated with the lifetime of that individual. Longevity risk is the major unhedged risk faced by many providers of annuities and pensions. Due to the increased visibility of longevity risk and secular trends in life expectancy, there has been a proliferation of mortality models in the literature (e.g. Lee and Carter, 1992 ; Cairns et al., 2006 ; Currie, 2006 ; O'Hare and Li, 2012).

The significance of longevity risk and recognised capacity issues within the traditional financial institutions has resulted in the process of transferring longevity risk through the traditional route becoming prohibitively expensive. This has led to the capital markets, who have historically had little involvement with longevity risk, recognising longevity risk as a potential market. Longevity risk is also seen to be relatively uncorrelated with the more traditional financial risks managed by the capital markets and so it offers diversification opportunities. The potential size of the market in longevity risk is huge. For example, recent deals involving longevity include Rolls Royce who entered a £3bn longevity swap deal with Deutsche Bank in 2011 ; British Airways

who entered a £1.3bn deal with Goldman Sachs in 2010 and again in 2011 and Pilkington Glass who entered a £1bn with Legal and General in 2012. For a review of the development of the international longevity market see Cairns *et al* (2006a), Barrieu *et al*(2010), Blake *et al* (2011).

Longevity hedging products are priced according to forecasts of mortality rates. Therefore underpinning the development of a successful market in longevity is the development of accurate and robust models of mortality rates that fit the data well and that can be used to forecast mortality rates accurately. The existing literature in the field is limited in this respect as many papers do not address the forecasting aspect of their models adequately. In this paper we propose a model of mortality rates that outperforms the existing models in the literature on several measures of forecastability. We fit the model to a range of countries using data from 1950-1990 and forecast the model rates from 1991-2006 testing forecasts against actual mortality rates. We repeat this for a range of stochastic models that are standard in the literature for comparison.

The remainder of the paper is organized as follows. In Section 2, we discuss longevity risk and the development of hedging products. We also discuss the existing stochastic models in the literature. In section 3, we discuss the data that we are using to carry out our analysis. Section 4 discusses the dynamic factor modelling methodology in the context of modelling mortality rates. In Section 5 we explain why this approach is appropriate for forecasting mortality and fit this model to mortality data from 6 different countries (USA,UK, Netherlands, France, Australia and Japan). We examine the forecast performance of the dynamic factor model unlike many other studies which emphasize the fit quality of their models. We compare results from this model with results from the seminal Lee Carter mdoel, a Lee Carter variant with a second factor for older ages and a Lee Carter variant with a cohort effect as well as additional factors. We then attempt to develop an ex-ante rule for selecting the numbers of dynamic and

static factors to improve performance. Conclusions and the implications for pricing are discussed in Section 6.

3.2 Context

3.2.1 Longevity risk

The risks associated with lifetime can be broken down into two types. Unsystematic mortality risk is the risk that the actual number of deaths in a given population will deviate from the anticipated number of deaths. This risk can be partly diversified away by increasing the population size if we assume that individual lifetimes are independent of each other. However, catastrophic mortality shocks due to violent storms or epidemics cannot be diversified away. The second type of risk, systematic mortality risk or longevity risk, is the risk that future mortality rates evolve in a different way to that anticipated.

Quantifying these two different types of risk is important from a risk management perspective for different stakeholders. For example, insurance firms writing large numbers of life policy business will take a keen interest in potential catastrophic events and the associated catastrophic mortality risk as this may directly affect their outgo. In contrast, given the persistently improving trend in expected lifetimes, life annuity writers will be particularly interested in minimising longevity risk.

Innovations in longevity risk started in the early 2000's with the theoretical development of the longevity bond by Blake and Burrows (2001) and with an adaptation of the catastrophe bond, traditionally applied in the inflation linked securities space, to mortality data. In the former, the longevity bond is designed to mitigate the longevity risk faced by an annuity provider i.e. the risk that future longevity trends have been mispriced into the annuity product. In the latter, the product innovation is designed

with a relatively short lifespan, say three years, and the insured is covered for the event of a catastrophic change in mortality experience over the short period, for example, from a pandemic that wipes out a significant proportion of life assurance policyholders.

At the same times as these theoretical developments were taking place trades involving the transfer of longevity risk were beginning to take place in the market. The first of these were longevity bond based (the EIB / BNP Parabas longevity bond) and were less successful. However, catastrophe bonds deals were successful (Swiss Re 3 year catastrophe bond) and mortality swap trades have also been relatively successful. The size of these trades has been significant and the increasing frequency with which they are taking place all point towards an appetite to innovative around this risk. Developments by several parties (J.P.Morgan, Lifemetrics; Goldman Sachs, QxX.LS; Deutsche Bank, Xpect Age) are now focused on standardising products so that a fluid market in longevity risk can be created. A more detailed summary of the developments in this respect can be found in Blake *et al (2011).*

3.2.2 Existing stochastic mortality models

Stochastic mortality models either model the central mortality rate, $m_{x,t}$ or the initial mortality rate, $q_{x,t}$ (see Coughlan *et al.*, 2007)[1]. Let $D_{x,t}$ be the number of people with age x that died in year t, and $E_{x,t}$, the exposure being the average population with age x in the year t, the central mortality rate $m_{x,t}$ is defined as:

$$m_{x,t} = \frac{D_{x,t}}{E_{x,t}}, \tag{3.1}$$

[1]The initial mortality rate q_x is the probability that a person aged x dies within the next year. The different mortality measures are linked by the approximation: $q_x \approx 1 - e^{-m_x}$.

The seminal stochastic mortality model is that of Lee and Carter (1992) which is given by:

$$\ln(m_{x,t}) = a_x + b_x\kappa_t + \epsilon_{x,t}, \tag{3.2}$$

where a_x and b_x are age effects and κ_t is a random period effect. Applying identifying constraints the a_x are given by:

$$a_x = \frac{1}{N}\sum_{t=1}^{N} \ln m_{x,t}. \tag{3.3}$$

The bilinear part $b_x\kappa_t$ was then determined as the first singular component of a singular value decomposition (SVD), with the remaining information from the SVD considered to be part of the error structure. The κ_t are estimated and refitted to ensure the model maps onto historic data and the subsequent time series κ_t is used to forecast mortality rates using ARIMA time series forecasting techniques.

Cairns *et al.* (2006, 2009, and 2011) summarized the main disadvantages of the Lee-Carter model. The model has one factor, resulting in mortality improvements at all ages being perfectly correlated (trivial correlation structure). For countries where a cohort effect is observed in the past, the model gives a poor fit to historical data. Also, the model can result in a lack of smoothness in the estimated age effect b_x. Despite the weaknesses of the Lee-Carter model its simplicity has led to it being taken as a benchmark against which other stochastic mortality models can be assessed.

Mortality data is two dimensional with deaths and exposures being recorded by year and by age. We can therefore consider the data from three different viewpoints, the age profile (or how mortality changes from age to age), the time profile (how

mortality rates for a specific age change over time), and the more recent approach, the cohort profile (how mortality for a specific cohort of the population - those born in a particular year - changes in relation to other cohorts).

Several survey papers comparing the various models of mortality rates have been published, see for example: Booth and Tickle (2008), Cairns *et al* (2011). In particular, these identify the extrapolative approach of Lee Carter and its variants as having been the most successful. Variants of the Lee Carter model allow for cohort effects and add additional factors to account for the behaviour of mortality among particular age groups (Cairns et al., 2006 ; Plat, 2009). The structure of these models is outlined in table 3.1. For a more thorough review of these models and other variants, the reader is referred to O'Hare and Li (2012).

We have benchmarked our model against these models since they have been the most successful strand of extrapolative mortality modelling and between them capture the young and older age effects as well as the cohort effects.

3.3 Data

The data that we use in this paper comes from the Human Mortality Database.[2] The data available for each country includes number of deaths $D_{x,t}$ and exposure to death $E_{x,t}$ for lives aged x last birthday during year t. We can use this to gain a proxy for the central mortality rate for lives aged x during year t as:

$$m_{x,t} = \frac{D_{x,t}}{E_{x,t}} \tag{3.4}$$

[2]This can be found at http://www.mortality.org/. The database is maintained in the Department of Demography at the University of California, Berkeley, USA, and at the Max Planck Institute for Demographic Research in Rostock, Germany.

Due to the exponential nature of mortality rates we model the logarithmically transformed central mortality rates. We defined this as $y_{x,t} = \ln m_{x,t}$.

Data is available going back to the mid nineteenth century in some cases but we have restricted this study to data from 1950-2006 and to six countries representing a geographical spread around the globe. Specifically, we fit the models to data between 1950-1990 and forecast in sample from 1991-2006 reporting the performance on U.K., U.S., Netherlands, France, Japan, and Australia. We focus on the age range 20-89 for several reasons. Firstly, the papers and models upon which we have based our comparisons are also fitted to this age range. Secondly, and as identified by Currie (2011), data at the older ages provide additional problems in terms of the reliability of the data. Indeed in several cases mortality rates determined using older data appear to fall sharply beyond age 95.

3.4 Methodology

The method we propose is a simplified dynamic factor model which is equivalent to using a Lee-Carter model but estimating the common factor in two steps : (i) by estimating the common covariance among age-specific mortality rates in the frequency domain and then (ii) by down-weighting age-specific mortality rates not following common patterns. This method is consistent under much more realistic conditions than allowed for in Lee-Carter.

The Lee-Carter model can be treated as a Principal Component model with one component and other researchers have used this understanding to generalise the model by extracting a larger number of components (Yang et al., 2010). We propose to generalise the model instead by taking account of the dynamic structure of mortality data using dynamic factor modelling (Forni *et al*, 2005; hereafter FHLR). The approach has

been traditionally used in forecasting macroeconomic indicators and is increasingly also used in macroeconomic analysis (Breitung and Eickmeier, 2006). This approach represents a generalisation of the traditional static factor model, building on the dynamic exact factor models of Geweke (1977) and Sargent and Sims (1977); the static approximate factor model of Chamberlain and Rothschild (1983) and the dynamic approximate factor model of Stock and Watson (2002).

Static factor analysis is characterised by extraction of contemporaneous co-movements in a vector of standardised data $\mathbf{y_t} = [\mathbf{y_{1t}}, \ldots, \mathbf{y_{Nt}}]'$ and is of the form

$$\mathbf{y_t} = \mathbf{\Lambda F_t} + \xi_\mathbf{t} \tag{3.5}$$

where $\mathbf{F_t} = [\mathbf{F_{1t}}, \ldots, \mathbf{F_{rt}}]'$,the idiosyncratic component $\xi_\mathbf{t} = [\xi_\mathbf{1t}, \ldots, \xi_\mathbf{Nt}]'$ and the factor loadings $\mathbf{\Lambda} = [\lambda_\mathbf{1}, \ldots, \lambda_\mathbf{N}]'$ for $\lambda_\mathbf{i} = [\lambda_\mathbf{i1}, \ldots, \lambda_\mathbf{iN}]'$. The estimator for Λ is consistent for a fixed number of variables, N and $T \to \infty$ as long as $E(\xi_t \xi_t') = \mathbf{\Sigma} = \sigma^\mathbf{2}\mathbf{I}$.The model is static in the sense that y_t is related contemporaneously with F_t but F_t itself can be a dynamic process i.e. $\mathbf{F_t} = \mathbf{B(L)f_t}$ where $\mathbf{B(L)}$ is a vector of dynamic factor loadings (Bai and Ng, 2007).

The FHLR approach to dynamic factor analysis uses a *dynamic approximate* factor model by which we mean (i) an approximate factor model where the idiosyncratic errors $\xi_\mathbf{t}$ are allowed to be weakly correlated and (ii) a dynamic factor model where the common factors are loaded onto the variables through a lag structure which is assumed finite. The model is:

$$\mathbf{y_t} = \mathbf{B(L)f_t} + \xi_\mathbf{t} \tag{3.6}$$

where $\mathbf{f_t} = [\mathbf{f_{1t}}, \ldots, \mathbf{f_{qt}}]'$ and $\mathbf{B(L)}$ is a matrix whose (i, j) entry is $b_{ij}(L)$, a polynomial in the lag operator. The dynamic model therefore is similar to a static factor

model where the set of r static factors are lags of q dynamic factors. In the empirical work carried out in this paper, we use the smallest number of factors possible i.e. $r = 1$ and $q = 1$. In this case, the model reduces to $\mathbf{y_t} = \mathbf{BF_t} + \xi_\mathbf{t}$ for $\mathbf{B} = [\mathbf{b_1}, .., \mathbf{b_N}]'$ which has one parameter for each age as in the static factor Lee Carter model. However, in the static case, factors are estimated by an eigenvalue decomposition of the data covariance matrix, while the dynamic factor approach uses an eigenvalue decomposition of the data spectral density matrix leading to more efficient estimates.

First, the lag-k covariance matrices for the common component,$\mathbf{\Gamma^\chi_k}$, and the idiosyncratic component,$\mathbf{\Gamma^\xi_0}$, are estimated using a frequency-domain analysis i.e. explaining $\mathbf{y_t}$ by cycles of different frequency. The factor space is then estimated by a process equivalent to extracting static factors from data transformed by down-weighting variables with large idiosyncratic components. Using the contemporaneous idiosyncratic covariance $\mathbf{\hat{\Gamma}^\xi_0}$ estimated in the first step, standard principal components are extracted from the transformed vector $\mathbf{\tilde{y}_t} = (\hat{\mathbf{\Gamma}}_\mathbf{0}^{\ \xi})^{-\mathbf{1/2}}\mathbf{y_t}$. A fuller description is given in the appendix and in FHLR.

The estimation procedure is consistent under the following assumptions:

(i) The idiosyncratic errors $\xi_\mathbf{t} = [\xi_\mathbf{1t}, \ldots, \xi_\mathbf{Nt}]'$ can be serially correlated but must be stationary.

(ii) They may also be heteroscedastic and weakly cross-correlated (i.e. "finite clusters of correlation").

(iii) The factor $\mathbf{f_t}$ and idiosyncratic errors $\xi_\mathbf{t}$ may also be correlated.

To apply this approach to the context of mortality modelling, let $D_{x,t}$ denote the number of deaths in a population at age x and period t and $E_{x,t}$ the corresponding

exposure. We define mortality rate

$$y_{x,t} = \ln m_{x,t} = \ln(\frac{D_{x,t}}{E_{x,t}}). \tag{3.7}$$

To make the data stationary we take the first difference $\Delta y_{x,t}$ and standardize $z_{x,t} = \frac{\Delta y_{x,t} - \mu_x}{\sigma_x}$. Forecasts are then generated by cumulatively summing the estimated change in (log) mortality as

$$\hat{y}_{x,T+h|T} = y_{x,T} + \hat{\Delta} y_{x,T+1|T} + \ldots + \hat{\Delta} y_{x,T+h|T} \tag{3.8}$$

Similar to the assessment of forecasting in GAD (2001), the prediction accuracy at horizon h is evaluated using the following three measures:

- A measure of overall bias: the mean percentage error over all ages from 20 to 89 inclusive

$$E1_y^h = \frac{1}{70 * h} \sum_{x=20}^{x=89} \sum_{k=1}^{h} \frac{\hat{y}_{x,T+h|T} - y_{x,T+h}}{y_{x,T+h}} \tag{3.9}$$

- A measure of the magnitude of the error : the mean absolute percentage error (MAPE) over all ages from 20 to 89 inclusive

$$E2_y^h = \frac{1}{70 * h} \sum_{x=20}^{x=89} \sum_{k=1}^{h} \frac{|\hat{y}_{x,T+h|T} - y_{x,T+h}|}{y_{x,T+h}} \tag{3.10}$$

- A measure of the standard deviation of the error : the root mean square of the percentage error.

$$E3_y^h = \sqrt{\frac{1}{70 * h} \sum_{x=20}^{x=89} \sum_{k=1}^{h} \Big(\frac{\hat{y}_{x,T+h|h} - y_{x,T+h}}{y_{x,T+h}}\Big)^2} \tag{3.11}$$

Percentage errors are preferred in each measure so as to make the error size independent of the size of the mortality rate.

In the context of mortality modelling, the FHLR dynamic approximate factor model represents an improvement over simply extracting static factors from the data as

(i) By being approximate it allows the component of age-specific mortality rates not explained by common factors to be correlated across a subset of ages (cross-correlation) and across time (serial correlation). This is reasonable under the plausible assumption that there are unidentified drivers affecting mortality only for subsets of the population e.g. improvements in geriatric care for the very elderly or alcohol consumption among young men.

(ii) By being dynamic it allows for a small number of core factors that are driving mortality changes for every age but with possibly different lags. This would capture cohort effects due to early life effects, educational improvements or the young being quicker to adapt behaviour.

(iii) By projecting on an estimate of the common component it fully exploits the dynamics of the data by using all contemporaneous and lagged covariances in death rates across the ages considered.

(iv) By using generalised principal components to extract the common factor it down-weights data which is not associated with the common trend and this should provide more efficient estimates.

3.5 Results

In this section we discuss the characteristics of data that we would expect a dynamic factor model to be able to capture, demonstrate the presence of these characteristics

in our mortality data and compare forecasting results of existing models against the dynamic factor model. Models are fitted to the years 1950 - 1990 and ages 20-89 and forecast from 1991 to 2006 for male death rates in United Kingdom (UK), United States (US), Australia (AUS), France (FRA), the Netherlands (NTH) and Japan (JPN).

The data are first transformed to be stationary. Panel Unit root tests are provided in Table 3.2. For all countries it is clear that the null hypothesis of non-stationarity is rejected when the data is first-differenced. In levels, tests indicate non-stationarity except when a trend term is added to the ADF regression but no lagged terms are added to account for error autocorrelation. It therefore seems reasonable to first difference the data to ensure stationarity as required by FHLR.

D'Agostino and Giannone (2007) state that a dynamic factor model would give a reasonable representation of the data if the data displayed the following characteristics,

(i) strong co-movements in the data,

(ii) a rich dynamic structure, and

(iii) variation in the amount of idiosyncratic variance.

In the first instance therefore, we will examine U.S. mortality data for those aged 20-89 over the period 1950-1990 for these patterns taking each one in turn. All of the mortality models reviewed in section 2 contain a common factor component (e.g.κ_t in Lee-Carter or κ_t^1,κ_t^2,and κ_t^3 in Plat) reflecting the consistent finding that death rates for different ages tend to follow common secular trends. It should therefore be clear that there should be co-movements in the data over time. This can also be seen in the second row of table 3.3 where a standard principal components analysis (PCA) has been performed on the data which has first been transformed to be stationary[3]. If the data are contemporaneously cross-correlated then extracted factors should capture much of

[3]The natural logarithm of the data was first-differenced and standardised.

the variation in the data. The percentage of variation explained is given for models with one through to ten factors and in all cases the amount of variance explained by the common factors is substantial. The first principal component explains 29% of the variation in the data while ten factors explain 72% of the variation. Mortality at each age is therefore strongly associated with trends over time that are common to all ages.

To check if there is a rich dynamic structure we also give the percentage variation explained by one to ten dynamic factors in the third row[4]. It can be seen, first of all, that a much smaller number of dynamic factors are required to capture the variation in the data e.g. three dynamic factors capture approximately the same amount of variance as nine static factors. If we extracted factors until 95% of the total variance of the data was explained[5] we would require ten dynamic factors and twenty-eight static factors. This indicates that it may be possible to parsimoniously represent large numbers of static factors by a linear combination of lags of a small number of dynamic factors justifying the representation in (11).

Mortality rates for each age have different levels of association with these common factors. Allowing for one dynamic factor, the percentage of the variation in mortality rates not explained by the common factor is calculated for each age. The results are graphed in figure 3.1. The variation is clear: while 57% of the variation in mortality rates at age 52 is unexplained only 23% of the variation at age 77 is not captured by the dynamic factor. The variation in signal to noise observed indicates that the FHLR dynamic factor approach of correcting estimates of common factors by suppressing information from those variables with the highest idiosyncratic errors will provide more efficient results. This is analogous to corrections for heteroscedasticity in ordinary least squares by using weights inversely proportional to the error variance.

[4]As in D'Agostino and Giannone (2007), we use $\frac{trace(\hat{\Gamma}_0^{\chi})}{trace(\hat{\Gamma}_0)}$ as our measure of the percentage of variance explained where $\hat{\Gamma}_0^{\chi}$ is estimated by the first q dynamic factors

[5]This rule is advocated by Joliffe (1986) to determine when to stop extracting factors.

Generally, the simplest dynamic factor model provides a good fit to US male mortality data. The correlation coefficient between the actual and forecast first differences of log mortality rates over ages 20-89 is 0.52. However, as we have argued above, an exact fit to the data is not necessarily helpful in improving the accuracy of predictions.

Forecasts and actual mortality rates at all ages are compared in figure 3.2 below. An overprediction is indicated by a value greater than 1 and an underprediction by a value less than 1. The general pattern is that younger age mortality rates are overpredicted, middle age mortality rates are underpredicted and older age mortality forecasts are quite accurate. Progressing from left to right, forecasts get worse the further into the future we try to predict. For example in 2006, the forecast for age 32 is 53% higher than the actual mortality rate in that year and the forecast mortality rate for age 49 is 16% lower than the actual mortality rate.

Other models also have large forecast errors. Looking at mortality rates for all ages together, the measures of forecasting quality E1-E3 described in section 4 are calculated below in tables 3.4-3.6 using data from all countries. The most accurate forecast is highlighted in bold. Considering the E1 measure of forecasting bias which measures the average forecast error, we see that in all cases except Japan forecasting models overestimate mortality rates. The dynamic forecasting model is best in three out of the six countries considered while Lee-Carter is best in the USA and UK and Plat is best for Japan. In the latter case the Plat model which incorporates a second age-period effect, cohort effects and younger age-period effect is vastly superior to all others. The additional younger age period effect here appears to be modelling a dramatic improvement in younger age mortality which is not correlated to general trends in mortality. The Cairns model is the most biased in all countries.

Considering the E2 measure of forecast error which measures the magnitude of difference between forecasts and actual rates, the magnitude of the forecast error can

be seen to be roughly an inverse function of population size and is consequently largest for Australia and the Netherlands. With this measure, the dynamic model is best for all countries except the USA. The improvement in percentage error from using the dynamic model instead of the Lee-Carter model for the UK is substantial at almost 5%. For Japan, the improvement is more slight. The Cairns and Plat models are generally worse with the former being wildly inaccurate. The low level of forecast error in the Lee-Carter model for the UK and the Plat model for Japan indicated by E1 disguises the true level of inaccuracy of forecasts seen in E2 indicating underestimates and overestimates are negating each other to give a low bias estimate. Whereas E2 measures the average absolute error, E3 measures the root mean square of the error. On this measure, the dynamic model is best for most countries.

Taking all error measures together across all countries, the more parameterised models (Cairns and Plat) generally give worse forecasts. Much of the justification for adding additional parameters to models in the literature reviewed above relies on the goodness of fit over the period of estimation. The results given in this paper highlight the importance of testing the ability to forecast instead of goodness of fit particularly when the development of long term longevity hedging products is dependent on accurate forecasts. The dynamic factor model gives superior forecasts everywhere. For the Netherlands, France and Australia the dynamic factor model is both less biased and more accurate (regardless of the measure). In the US, UK and Japan the dynamic factor model is generally more accurate than all other models with lower average absolute error and root mean square error but is more systematically biased. When developing hedging products based on forecasts of mortality rates the issue of low bias and high absolute error (for example caused by recurring over and underestimation of mortality rates), as measured by a low E1 and high E2 measure, is a particularly serious one. For example, a longevity swap priced using a mortality model which may be unbiased but

which produces estimates which fluctuate around the actual mortality profile would lead to larger transactions and costs associated with the larger errors between actual and predicted mortality. The alternative of a more bias model which has a low absolute error, as measured by a high E1 and low E2, will result in a lower level of transactions and associated costs. This issue partly contributes to the lack of appeal in long term hedging of longevity risk and potentially rules out the use of models which suffer from large absolute errors.

To investigate short term and longer term forecast accuracy we follow the rolling origin approach proposed by Shang *et al* (2011). The fitting period is initially set to 1950-1990, and then we compute one-step-ahead (i.e. forecasting the 1991 mortality rate) and ten-step-ahead point forecasts (i.e. the 2000 mortality rate). Forecast errors E1 - E3 are computed as before. We then increase the fitting period by one year (i.e. 1950-1991) and compute one-step-ahead and ten-step-ahead forecasts again. This process is repeated until we have one-step and ten-step ahead forecasts for 2006. The results are shown in tables 3.7 - 3.12. In the short term, the dynamic factor model generally outperforms the comparator models on the E2 and E3 measures. The picture is less clear on the E1 bias measure. In the longer term, the dynamic factor model outperforms the comparator models on all measures for most countries except for the US and Japan where the Lee Carter and dynamic factor forecast accuracy is comparable. This indicates that the superior forecasting accuracy of the dynamic factor model over the 16 year period highlighted above is due to a combination of superior short term and longer term forecasts.

A simple dynamic factor with one dynamic and one static factor surpasses existing models in most countries according to these forecast measures. As a first step, it therefore seems reasonable to advocate using dynamic factor modelling with a default of $q = 1, r = 1$ as generally the most accurate approach to forecasting male mortality

rates. The particular success of the Lee Carter model when fitting to U.S. data has led to its extensive use as a reference model in stochastic mortality modelling. But as we have shown here (and as has been shown by others), the Lee Carter model does not perform particularly well in other countries.

We now examine whether there is a simple ex-ante rule-of-thumb for choosing q and r to improve on the forecasting power of the simplest dynamic factor model. Calculating the MAPE (E2) for the number of dynamic factors $q = 1, ..., 10$ and the number of static factors $r = 1, ..., 10$ and ranking models on the size of the forecast error (see table 3.13) we see that the simplest model is certainly nowhere near the most accurate except in the case of Australia. We can use statistical approaches to terminating factor extraction. Standard stopping rules for determining the number of static principal components,r, in data which were developed in psychological or sociological studies are not applicable in datasets with large number of variables (Breitung and Eickmeier, 2005). Bai and Ng (2002) have developed a statistical test based on information criteria with appropriately chosen penalties to determine the number of static factors in this situation. In later work (Bai and Ng, 2007), they also develop a rule for estimating the number of dynamic factors,q, based on a VAR of finite order in r static factors (the 'restricted dynamic factor model') . Hallin and Liska (2007) develop an alternative test applicable when the VAR has infinite length ('the generalised dynamic factor model'). Applying the Bai and Ng rules to our data leads to the choice of $\hat{q}_{BN} = 1, \hat{r} = 1$ (UK,NTH, AUS) or $\hat{q}_{BN} = 2, \hat{r} = 2$ (USA, FRA,JPN). The Hallin and Liska estimates of q suggest a much richer dynamic structure: $\hat{q}_{HL} = 8$ (USA,NTH) or $\hat{q}_{HL} = 9$ (UK, FRA AUS,JPN).

Error measures are given in table 3.14 for models where the number of factors is chosen ex-ante and compared to results from the simplest dynamic factor model. The best results for each measure are then emboldened. There is little evidence that

selecting the number of factors by these rules consistently improves forecasts. Using $\hat{r}$ as above and $\hat{q}_{HL}$, results are slightly improved for the USA and Japan. For all other countries using the rules is no more accurate than using the simplest model. In conclusion, there does not appear to be a clear way to optimally choose factors to maximize accuracy and the potential improvement anyway appears to be slight. Dynamic factor models based on factor selection rules do not necessarily give the most accurate forecasts due to the fact that these statistical rules are based on the fit of factor models to the data over which the model is estimated and do not necessarily relate directly to predictive power. In an assessment of the ability of factor selection to improve forecast accuracy but in a macroeconomic application, Barhoumi et al (2010) reached a similar conclusion - factor selection tests do not generally improve forecasts from dynamic factor models and the nave model is often best.

3.6 Conclusions

In this paper we have provided an alternative approach to modelling and forecasting mortality rates which gives superior forecasts to many of the standard stochastic mortality models that form the basis of longevity hedging products. We have described problems with developing long term mortality hedging products due to the lack of consideration given to the forecastability of existing models. We focus on the forecasting quality of our model using out of sample comparisons and conclude that existing multifactorial models give poor forecasting performance. Even with the simplest of specifications the dynamic model outperforms the the existing models of mortality.

Further work may be carried out to examine the forecasting power of the approach over specific age ranges rather than the full age range, and over a longer forecast horizon. Identifying the number of static and dynamic factors giving optimal forecasts

needs to be further examined. In keeping with the current direction of research in this area potential directions may be to (i) incorporate exogenous determinants into the dynamic factor model and (ii) to develop hedging products based on the dynamic factor model.

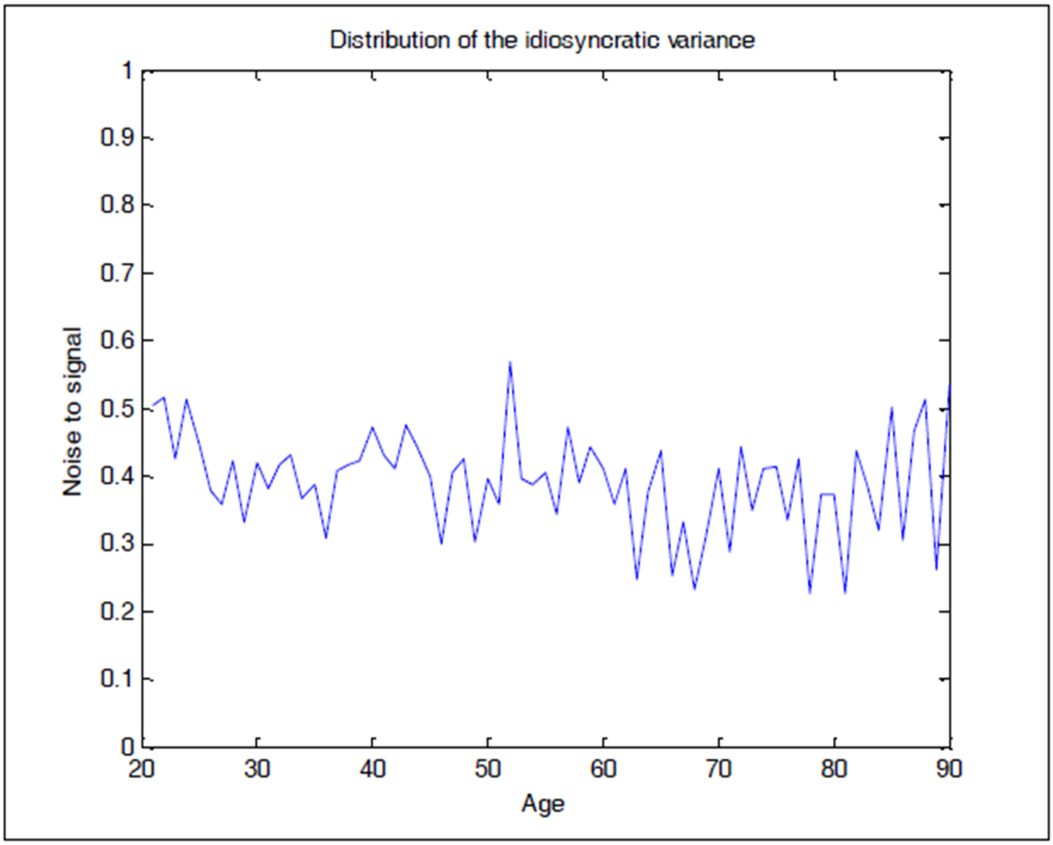

Figure 3.1: Distribution of noise to signal ratio on extracting 1 dynamic factor for US male mortality, ages 20-89

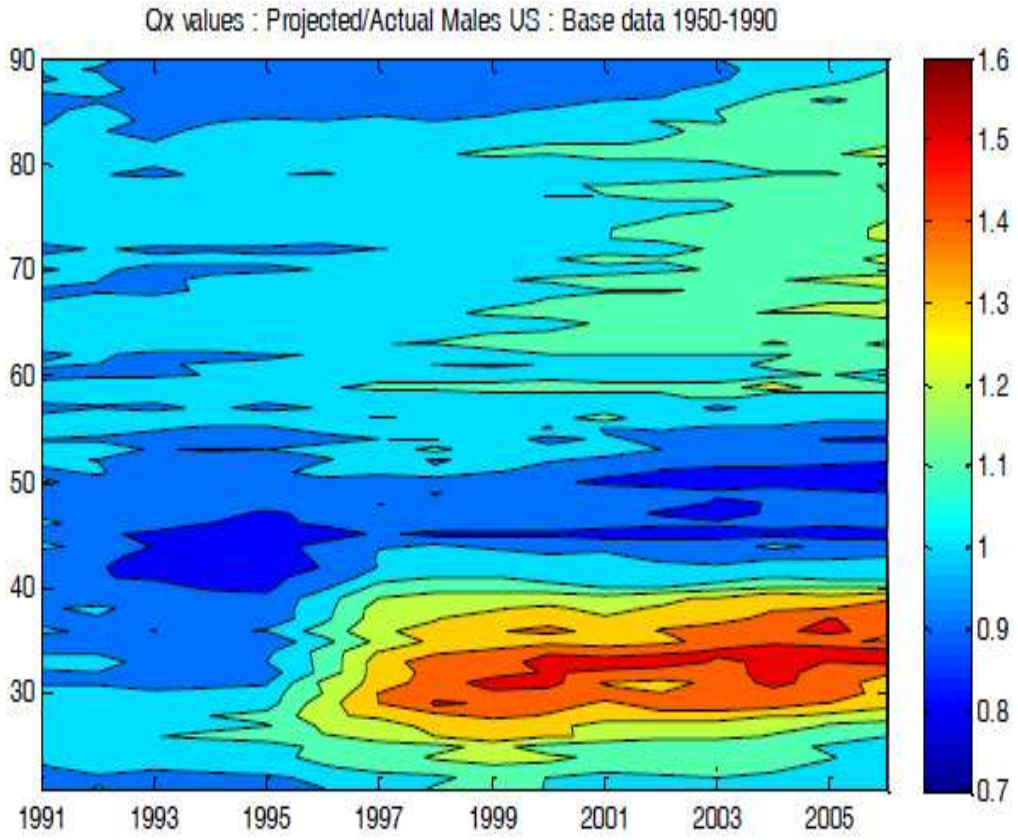

Figure 3.2: Projected / Actual males US: Base data 1950-1990

Appendix 1 - Dynamic Factor modelling

It is assumed that each of the stochastic processes $\mathbf{y_{it}}$ are stationary and that $\mathbf{y_t} = [\mathbf{y_{1t}}, \ldots, \mathbf{y_{Nt}}]'$ can be decomposed as the sum of an idiosyncratic component $\xi_\mathbf{t} = (\mathbf{xi_{1t}}, \ldots, \xi_\mathbf{nt})'$ and a common component $\chi_\mathbf{t} = (\chi_\mathbf{1t}, \ldots, \chi_\mathbf{nt})'$ which are assumed orthogonal i.e.

$$\mathbf{y_t} = \chi_\mathbf{t} + \xi_\mathbf{t}$$

Each $\chi_\mathbf{it}$ is driven by q factors $\mathbf{f_t} = (\mathbf{f_{1t}}, \ldots, \mathbf{f_{qt}})'$ with possibly different lags and coefficients. Using the lag operator to indicate a linear combination of various lags of a factor f_{jt} we have

$$\begin{array}{rrcl} & \chi_{it} & = & b_{i1}(L)f_{1t} + b_{i2}(L)f_{2t} + \ldots + b_{iq}(L)f_{qt} \\ \text{or} & \chi_\mathbf{t} & = & \mathbf{B(L)f_t} \\ \text{then} & \mathbf{y_t} & = & \chi_\mathbf{t} + \xi_\mathbf{t} = \mathbf{B(L)f_t} + \xi_\mathbf{t} \end{array}$$

Estimation of the dynamic factors $\chi_\mathbf{t}$ is performed in the frequency domain. Any stationary variable y_t can be written in the frequency domain as a weighted sum of periodic functions

$$y_t = \mu + \int_0^\pi [\alpha(\omega)\cos(\omega t)d\omega + \delta(\omega)\sin(\omega t)]d\omega$$

The spectral density of y_t is then given by

$$\sigma_{ii}(\omega) = \frac{1}{2\pi}\Sigma_{j=-\infty}^{\infty}\gamma_j e^{-i\omega j}$$

where γ_j is the j^{th} autocovariance (Hamilton, 1994). The frequency domain representation of the factor model is given by $\mathbf{\Sigma^y}(\omega) = \mathbf{\Sigma}^\chi(\omega) + \mathbf{\Sigma}^\xi(\omega)$. The first step of the process provides an estimate of $\mathbf{\Sigma}^\chi(\omega)$ and consequently $\mathbf{\Sigma}^\xi(\omega)$ by extracting the

first q principal components from an estimate of $\mathbf{\Sigma}^{\mathbf{y}}(\omega)$. This is known as *dynamic principal components analysis* (Brillinger, 1981). By taking inverse Fourier transforms of the respective spectral densities we can determine the autocovariance matrices $\mathbf{\Gamma}^{\chi}_{\mathbf{k}}$ and $\mathbf{\Gamma}^{\xi}_{\mathbf{k}}$ and this information is used in the second step to generate a more efficient representation of the factor space than standard PCA. A linear combination of the processes y_{it} is chosen so as to approximate the factor space i.e. since $\mathbf{a}\mathbf{y_t} = \mathbf{a}\chi_{\mathbf{t}} + \mathbf{a}\xi_{\mathbf{t}}$ we choose $\mathbf{a}$ such that $var(\mathbf{a}\xi_{\mathbf{t}})$ is minimised subject to the constraint $var(\mathbf{a}\mathbf{y_t}) = \mathbf{1}$. Using the covariance matrices $\hat{\Gamma}^{\chi}_{0}$ and $\hat{\Gamma}^{\xi}_{0}$ estimated in the first step this problem reduces to a generalised principal components problem of the form

$$\hat{\mathbf{Z}}_{\mathbf{j}}\hat{\mathbf{\Gamma}}^{\chi}_{\mathbf{0}} = \hat{\mathbf{v}}_{\mathbf{j}}\hat{\mathbf{Z}}_{\mathbf{j}}\hat{\mathbf{\Gamma}}^{\xi}_{\mathbf{0}}$$

such that $\hat{\mathbf{Z}}_{\mathbf{j}}\hat{\mathbf{\Gamma}}^{\xi}_{\mathbf{0}}\hat{\mathbf{Z}}_{\mathbf{j}} = \mathbf{1}$ and $\hat{\mathbf{Z}}_{\mathbf{i}}\hat{\mathbf{\Gamma}}^{\xi}_{\mathbf{0}}\hat{\mathbf{Z}}_{\mathbf{j}} = \mathbf{1}$ for $i \neq j$. It can be shown that the generalised principal components of $\mathbf{y_t}$ are equivalent to the standard principal components of the transformed vector $\tilde{\mathbf{y_t}} = \left(\hat{\mathbf{\Gamma}}^{\xi}_{\mathbf{0}}\right)^{\mathbf{1/2}}\mathbf{y_t}$. Therefore this approach provides estimates of the common factors that weights the data in such a way as to suppress information from those variables with the highest idiosyncratic errors. This should provide more efficient estimates. Forecasts are then obtained from the generalised principal components

$$\hat{\mathbf{Z}} = \left(\hat{\mathbf{Z}_{\mathbf{1}}}', \ldots, \hat{\mathbf{Z}_{\mathbf{r}}}'\right)'$$

giving

$$\hat{\chi}_{T+h|T} = \left[\hat{\mathbf{\Gamma}}^{\chi}_{\mathbf{h}}\hat{\mathbf{Z}}'\left(\hat{\mathbf{Z}}\hat{\mathbf{\Gamma}}_{\mathbf{0}}\hat{\mathbf{Z}}'\right)^{-\mathbf{1}}\right]\left[\hat{\mathbf{Z}}\mathbf{x_t}\right].$$

3.7 Appendix: Additional Figures and Tables

Table 3.1: Stochastic mortality models

Model	Structure
Lee Carter (1992)	$\ln(m_{x,t}) = a_x + b_x\gamma_t + \epsilon_{x,t}$
Cairns, Blake and Dowd (2006)	$\text{logit}(q_{x,t}) = \gamma_t^1 + \gamma_t^2(x - \bar{x}) + \epsilon_{x,t}$
Plat (2009)	$\ln(m_{x,t}) = a_x + \gamma_t^1 + \gamma_t^2(\bar{x} - x) + \gamma_t^3(\bar{x} - x)^+ + \kappa_{t-x} + \epsilon_{x,t}$

Table 3.2: Panel Unit Root tests for log mortality in levels and first differences

	Levels						First differences					
	Constant Lags			Constant & Trend Lags			Constant Lags			Constant & Trend Lags		
	p=0	p=1	p=2	p=0	p=1	p=2	p=0	p=1	p=2	p=0	p=1	p=2
USA	13.1	13.1	13.1	4.9	8.0	6.7	**-54.6**	**-29.0**	**-16.7**	**-54.5**	**-27.8**	**-14.6**
UK	3.6	7.5	13.5	**-12.6**	-0.1	5.9	**-77.8**	**-45.4**	**-29.3**	**-79.3**	**-46.7**	**-31.4**
NTH	-7.3	0.8	4.9	**-13.6**	-0.4	5.6	**-79.8**	**-45.9**	**-25.7**	**-82.1**	**-47.2**	**-25.8**
FRA	1.9	7.4	12.6	**-18.8**	-1.1	5.8	**-88.0**	**-48.8**	**-25.9**	**-89.6**	**-48.7**	**-23.7**
AUS	5.9	12.7	15.4	**-6.4**	4.8	9.2	**-75.6**	**-43.8**	**-27.0**	**-77.1**	**-44.2**	**-28.2**
JPN	1.3	6.0	10.7	**-12.5**	**-3.6**	2.8	**-64.6**	**-47.8**	**-29.3**	**-64.7**	**-48.0**	**-28.8**

Note: Following Im, Pesaran and Shin (2003) we use $W_{tbar}(p,\rho) = \frac{\sqrt{N}\left(t-bar_{NT} - \frac{1}{N}\sum_{i=1}^{N} E[t_{iT}(p_i,0)|\beta_i=0]\right)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} Var[t_{iT}(p_i,0)|\beta_i=0]}}$ where $t-bar_{NT}$ is the simple average of the t-tests for the null of unit roots ($\beta_i = 0$) in the Augmented Dickey-Fuller test $\Delta y_{it} = \alpha_i + \delta_i t + \beta_i y_{i,t-1} + \sum_{j=1}^{p} \rho_{i,j}\Delta y_{i,t-j} + \epsilon_{it}$. The lag length p is fixed across all ages. The statistic $W_{tbar}(p,\rho)$ is asymptotically normally distributed. Emboldened results represent a rejection of the null hypothesis using a left-sided test (i.e.$W_{tbar} < -1.64$). Results are also given for ADF specification with constant only ($\delta_i = 0 \forall i$).

Table 3.3: Percentage variance explained by factors

Number of Factors	1	2	3	4	5	6	7	8	9	10
Static	29.3%	42.2%	47.7%	52.4%	56.5%	60.3%	63.8%	67.1%	69.8%	72.3%
Dynamic	41.3%	58.6%	69.3%	77.6%	83.6%	88.2%	91.4%	93.4%	94.8%	95.8%

Table 3.4: E1 - the mean percentage error of projection (overall bias) for males aged 20-89

Country	Lee & Carter	Cairns *et al.*	Plat	Dynamic
USA	**6.14%**	14.74%	14.59%	7.82%
UK	**3.75%**	27.28%	12.63%	6.43%
NTH	16.98%	25.90%	17.39%	**10.72%**
FRA	12.67%	30.85%	20.28%	**10.83%**
AUS	14.90%	36.72%	24.61%	**12.61%**
JPN	-11.90%	12.85%	**-2.93%**	-12.07%

Table 3.5: E2 - the mean absolute percentage error of projection (overall error magnitude) for males aged 20-89

Country	Lee & Carter	Cairns *et al.*	Plat	Dynamic
USA	11.30%	28.70%	18.30%	**10.82**%
UK	17.05%	48.48%	15.97%	**12.18%**
NTH	18.83%	41.35%	20.28%	**14.94%**
FRA	16.91%	46.57%	22.61%	**13.40%**
AUS	20.15%	62.27%	26.89%	**16.40%**
JPN	14.80%	50.31%	16.16%	**14.42%**

Table 3.6: E3 - the root of the squared percentage error of projection (standard deviation of the error) for males aged 20-89

Country	Lee & Carter	Cairns *et al.*	Plat	Dynamic
USA	**14.56%**	34.06%	25.51%	16.25%
UK	21.13%	62.51%	22.28%	**16.36%**
NTH	25.52%	52.28%	27.55%	**20.05%**
FRA	26.79%	63.36%	31.81%	**20.56%**
AUS	26.85%	78.50%	37.86%	**22.06%**
JPN	21.10%	69.00%	20.03%	**19.44%**

Table 3.7: E1 - the mean percentage error of projection (overall bias) of one-step-ahead point forecasts for males aged 20-89

Country	Lee & Carter	Cairns *et al.*	Plat	Dynamic
USA	0.73%	-6.86%	1.35%	**0.61%**
UK	-4.32%	-7.17%	2.40%	**1.12%**
NTH	**0.61%**	-6.93%	3.65%	2.19%
FRA	4.45%	-6.72%	2.21%	**1.62%**
AUS	**0.49%**	-7.79%	4.97%	1.99%
JPN	-6.26%	-8.90%	1.58%	**-1.15%**

Table 3.8: E2 - the mean absolute percentage error of projection (overall error magnitude) of one-step-ahead point forecasts for males aged 20-89

Country	Lee & Carter	Cairns *et al.*	Plat	Dynamic
USA	6.80%	14.40%	7.30%	**2.76%**
UK	7.92%	16.10%	6.82%	**4.50%**
NTH	**6.57%**	14.57%	7.08%	7.18%
FRA	9.55%	14.58%	8.35%	**4.33%**
AUS	8.10%	23.33%	9.89%	**6.67%**
JPN	9.36%	13.40%	5.71%	**4.13%**

Table 3.9: E3 - the root of the squared percentage error of projection (standard deviation of the error) of one-step-ahead point forecasts for males aged 20-89

Country	Lee & Carter	Cairns *et al.*	Plat	Dynamic
USA	9.64%	21.22%	9.81%	**3.62%**
UK	9.84%	24.42%	11.78%	**5.83%**
NTH	**9.00%**	22.85%	9.54%	10.11%
FRA	14.77%	20.32%	12.29%	**5.92%**
AUS	11.20%	31.58%	17.40%	**8.87%**
JPN	13.41%	21.72%	7.19%	**5.42%**

Table 3.10: E1 - the mean percentage error of projection (overall bias) of ten-step-ahead point forecasts for males aged 20-89

Country	Lee & Carter	Cairns *et al.*	Plat	Dynamic
USA	13.29%	20.28%	20.80%	**12.51%**
UK	9.35%	31.87%	18.91%	**9.11%**
NTH	24.73%	33.37%	27.94%	**17.06%**
FRA	23.56%	40.19%	32.91%	**18.73%**
AUS	21.64%	43.06%	33.51%	**16.98%**
JPN	-10.51%	17.00%	**0.18%**	-12.06%

Table 3.11: E2 - the mean absolute percentage error of projection (overall error magnitude) of ten-step-ahead point forecasts for males aged 20-89

Country	Lee & Carter	Cairns *et al.*	Plat	Dynamic
USA	15.46%	31.36%	22.34%	**14.79%**
UK	18.22%	55.56%	21.17%	**14.90%**
NTH	25.46%	48.35%	28.01%	**19.16%**
FRA	23.84%	50.31%	33.50%	**20.06%**
AUS	24.15%	69.51%	33.60%	**19.75%**
JPN	**15.15%**	56.90%	20.22%	16.29%

Table 3.12: E3 - the root of the squared percentage error of projection (standard deviation of the error) of ten-step-ahead point forecasts for males aged 20-89

Country	Lee & Carter	Cairns *et al.*	Plat	Dynamic
USA	**19.18%**	34.30%	28.70%	20.02%
UK	21.52%	61.29%	27.06%	**17.81%**
NTH	29.65%	52.54%	33.47%	**22.67%**
FRA	32.81%	58.57%	42.33%	**27.61%**
AUS	29.00%	75.93%	43.05%	**22.80%**
JPN	20.94%	65.80%	22.80%	**20.35%**

Table 3.13: Rank of simplest dynamic factor model (q=1, r=1) on E2 measure compared to multifactorial dynamic factor models (100 combinations of q=1,...,10 ; r=1,...,10)

Country	Rank of simplest dynamic factor model
USA	92
UK	76
NTH	83
FRA	37
AUS	8
JPN	41

Table 3.14: Rank of simplest dynamic factor model (q=1, r=1) on E2 measure compared to multifactorial dynamic factor models (100 combinations of q=1,...,10 ; r=1,...,10)

Country	Rank of simplest dynamic factor model
USA	92
UK	76
NTH	83
FRA	37
AUS	8
JPN	41

Table 3.15: Error measures for dynamic factor modelling selecting q,r ex-ante

	E1			E2			E3		
	Dynamic	Dynamic (choosing q,r)		Dynamic	Dynamic(choosing q,r)		Dynamic	Dynamic(choosing q,r)	
	(q=1,r=1)	$\hat{r}, \hat{q}_{BN}$	$\hat{r}, \hat{q}_{HL}$	(q=1,r=1)	$\hat{r}, \hat{q}_{BN}$	$\hat{r}, \hat{q}_{HL}$	(q=1,r=1)	$\hat{r}, \hat{q}_{BN}$	$\hat{r}, \hat{q}_{HL}$
USA	7.82%	7.60%	**7.59%**	10.82%	10.66%	**10.61%**	16.25%	16.01%	**15.82%**
UK	**6.43%**	**6.43%**	6.64%	**12.18%**	**12.18%**	12.31%	**16.36%**	**16.36%**	16.51%
NTH	**10.72%**	**10.72%**	10.94%	**14.94%**	**14.94%**	15.02%	**20.05%**	**20.05%**	20.16%
FRA	**10.83%**	10.95%	11.08%	**13.40%**	13.47%	13.55%	**20.56%**	20.65%	20.77%
AUS	**12.61%**	**12.61%**	13.29%	**16.40%**	**16.40%**	16.88%	**22.06%**	**22.06%**	22.49%
JPN	-12.07%	-12.10%	**-12.05%**	14.42%	14.10%	**14.10%**	19.44%	**19.03%**	19.03%

Chapter 4

Identifying structural breaks in mortality data

4.1 Introduction

Over the past recent decades, life expectancy in developed countries has risen to historically unprecedented levels. The prospects of future reductions in mortality rates are of fundamental importance in various areas such as demography, actuarial studies, public health, social insurance planning, and economic policy. Over the last years, significant progress has been made in mortality forecasting (for a recent review see Booth, 2006). The most popular approaches to long-term forecasting are based on the Lee and Carter (1992) model. It describes the time-series movement of age-specific mortality as a function of a latent level of mortality, also known as the overall mortality index, which can be forecasted using simple time-series methods. The method was initially used to forecast mortality in the U.S., but since then has been applied to many other countries (amongst others see Tuljapurkar and Boe, 1998; Lee, 2000; Carter and Prskawetz, 2001; Lee and Miller, 2001; Booth et al., 2002; Brouhns and

Denude, 2002; Renshaw and Haberman, 2003 and Kosso et al., 2005).

The original Lee-Carter model received a number of criticisms (see the discussion in Lee and Miller, 2001) primarily due to its simplistic structure and so inability to fully capture the variations present in mortality data adequately. Particularly the fact that we do not see improvements in mortality rates across ages that are correlated to each other. This has led to several extensions being proposed in the literature (see Booth et al., 2006) to address these inadequacies. One major issue concerns the stability of the model over time. Since the method is usually applied to long time-series there is a risk that important structural changes may have occurred in the past and any neglected structural change in the estimation period may result in forecasts that have a tendency to deviate from the future realizations of the mortality index. This could lead to potentially large long-term forecast errors. In fact, historically, mortality in the U.S. has not always declined in a linear way as depicted in Lee and Carter (1992) for the period 1900-1989. The authors of that paper also re-estimated their random walk with drift model for the mortality index for several shorter and more recent periods and concluded that there was some instability. Other studies also document that there has been a systematic overestimation of the projected mortality rates in many countries (Kosso et al., 2005). In a multi-country comparison of several versions of the Lee-Carter method, Booth et al. (2006) find significant differences in the forecasting performance when alternative fitting periods are used, providing further evidence of different trends in the mortality rate.

In the demographic literature (e.g., Kannisto et al. 1994; Vaupel 1997) it has been observed that, for many developed countries, the reduction in mortality rates has accelerated in the 1970s. Although this observation has important implications for social, health, and research policy, there has been little attempt to test or quantify the existence and effects of such a shift. In this paper we provide evidence for the structural breaks

present in some of the main models of mortality. We examine the Lee Carter model and several of its variants and test the extracted time series for structural breaks. More specifically, having identified and fitted time series in the various models we apply the generalised fluctuations test framework of Kuan and Hornik (1995) to statistically test for and date any structural breaks. Having identified the presence of structural breaks we then refit the models allowing for these and present improved forecasting results. The proposed method is based on recent advances in testing and estimating structural change models.

Consideration has been given to structural changes in mortality trends in the actuarial literature. Li, Chan and Cheung (2011) applied a broken trend stationary model to the extracted mortality trend κ_t of the Lee carter model using the Zivot and Andrews' (1992) procedure. They applied the model to data from the U.S. and from England and Wales and in each case identified break points in the mid 1970's. Their findings confirmed those of Renshaw and Haberman (2003) who also identified an improvement in fitting if an adjustment in the trend was allowed for at 1975. Coelho and Nunes (2011) repeated this analysis using the tests of Harevy *et al* (2009) and Harris *et al* (2009) to identify the presence of and date any structural breaks in the extracted mortality trend κ_t. Their study was wider focusing on 18 different countries in total and focusing on both males and females. Notably they found structural breaks in 16 of the 18 countries for males but in only 5 of the 18 countries for females suggesting that any potential acceleration in mortality improvement has had a greater impact on male mortality than on female mortality. They also found a range of structural break dates from 1955, for Japanese females, through to the year 2000 for Netherlands males. They also forecast with and without an allowance for the identified structural breaks and in the case of Portugal suggest an increase in life expectancy at birth of just over 2 years (80.9 vs. 78.7) when allowing for the break. It is important to note in each of the cases studied

the mortality improvement factor κ_t appears to accelerate after the break suggesting that if there is a structural break identified then the resulting model allowing for this break will predict a higher life expectancy. In particular, using a model which doesn't adequately capture any structural change in the improvement in mortality rates for pricing and reserving may lead to an under provision of reserves or prices.

This paper contributes to the existing literature by considering not only the Lee Carter (1992) model but also a selection of extensions from the Lee carter model. The purpose of this is to test whether the inclusion of additional age, time or cohort effects has any effect on the presence or not of structural breaks in the fitted mortality improvement factor κ_t. The proposed methodology is applied to male data over the period 1950-2006 for a selection of developed countries. Structural changes in the rate of decline in the overall mortality rate are found for almost every country and model considered. By allowing for the structural breaks and showing the improved fit and forecast quality we demonstrate that accounting for a structural change leads to a major impact in mortality estimates.

The paper is organized as follows. Section 2 presents a brief review of extrapolative models such as the Lee-Carter model and its extensions. In section 3 we discuss the data used in this study and in section 4 we discuss the methodology we use to identify the structural breaks. An analysis of the fitted parameters in the Lee Carter model and its variants is discussed in section 5. In section 6 we present the results of our analysis to identify structural breaks and to quantify the impact we demonstrate the fitting and out of sample forecasting results with and without allowance for the identified structural changes. Finally, section 7 concludes with some ideas for further research.

4.2 Lee Carter and its variants

The current leading method for forecasting mortality rates is the stochastic extrapolation approach. In this method data is first transformed (by taking natural logarithms) and then analysed using statistical methods to identify and extract patterns. These patterns are then forecast using well known time series approaches. The resulting forecasts are then used to predict future mortality rates. The first and most well known stochastic mortality model of this type is the Lee and Carter (1992) model. Based on U.S. data the model uses a stochastic, time series framework to identify a single period effect pattern in the natural logarithm of mortality rates. This linear trend over time is extracted and using Box-Jenkins an appropriate ARIMA processes is fitted to the data (a random walk with drift in each case). The random walk with drift is forecast and resulting future mortality rates predicted. Also known as a one factor or one principle component approach the model became a benchmark and underlined a new approach to modelling mortality rates for several reasons; (1) firstly, the model has an extremely simple structure and so is very easy to communicate, and (2) secondly, the use of the random walk with drift enabled the authors not only to predict the expected future mortality rates but also to visualise the uncertainty associated with the predictions. The full model, outlined below includes two age dependent parameters a_x and b_x which respectively represent the intercept and gradient for the log mortality rate at each age and the time or period trend κ_t which is forecast using a random walk with drift:

$$\ln(m_{x,t}) = a_x + b_x\kappa_t + \epsilon_{x,t}, \tag{4.1}$$

where a_x and b_x are age effects and κ_t is a random period effect.[1]

[1]This model was fitted to US mortality data for ages 0-110 between the years of 1933 and 1987.

The model is known to be over parameterised and applying the necessary constraints as in the original Lee and Carter (1992) paper the a_x are given by

$$a_x = \frac{1}{N} \sum_{t=1}^{N} \ln m_{x,t}. \tag{4.2}$$

In the original paper the bilinear part $b_x \kappa_t$ of the model specification was determined as the first singular component of a singular value decomposition (SVD), with the remaining information from the SVD considered to be part of the error structure. The κ_t were then estimated and refitted to ensure the model mapped onto historic data Finally the subsequent time series κ_t was used to forecast mortality rates.

Despite the attractiveness of the models simplicity it has several weaknesses. Among many discussions of the Lee-Carter model, Cairns *et al.* (2006, 2009, and 2011) summarized the main disadvantage of the model as having only one factor, resulting in mortality improvements at all ages being perfectly correlated (trivial correlation structure). They also note that for countries where a cohort effect is observed in the past, the model gives a poor fit to historical data. The uncertainty in future death rates is proportional to the average improvement rate b_x which for high ages can lead to this uncertainty being too low, since historical improvement rates have often been lower at high ages. Also, the model can result in a lack of smoothness in the estimated age effect b_x.

Despite the weaknesses of the Lee-Carter model it's simplicity has led to it being taken as a benchmark against which other stochastic mortality models can be assessed. There has been a significant amount of literature developing additions to, or modifications of, the Lee-Carter model. For example Booth *et al.* (2002), Brouhns *et al.* (2002), Lee and Miller (2001), Girosi and King (2005), De Jong and Tickle (2006), Delwarde *et al.* (2007) and Renshaw and Haberman (2003, 2006).

Renshaw and Haberman (2006) modified the Lee-Carter model by simply adding a factor γ_{t-x} to capture effects that could be attributed to the year of birth $(t-x)$,

$$\ln(m_{x,t}) = a_x + b_x^1 \kappa_t + b_x^2 \gamma_{t-x} + \epsilon_{x,t}, \tag{4.3}$$

where κ_t is defined as before and γ_{t-x} is a random cohort effect.

The model does have a much better fit for countries such as the UK where a cohort effect has been identified, however it suffers from a lack of robustness perhaps due to the presence of more than one local maximum in the likelihood function. Among others, for instance Currie (2006) noted that if the model was fitted using data from 1961-2000 then the parameters showed qualitatively different characteristics to those obtained when fitting to data from 1981-2000. Furthermore, as noted by Currie (2006), although the model incorporates the cohort effect, for most of the simulated mortality rates the correlation structure is still trivial with the simulated cohort parameters only being relevant for the higher ages at the far end of the projection.

Following this analysis Currie (2006) applied a simplified age-period-cohort model of Clayton and Schifflers (1987) to mortality which removed the robustness problem but at the expense of the fitting quality:

$$\ln(m_{x,t}) = a_x + \kappa_t + \gamma_{t-x} + \epsilon_{x,t}. \tag{4.4}$$

Cairns *et al.* (2006) observed that for England & Wales and United States data, the fitted cohort effect appeared to have a trend in the year of birth. This suggested that the cohort effect was compensating for the lack of a second age-period effect, as well as trying to capture the cohort effect in the data. This led them to introduce a two factor

model of mortality,

$$\text{logit}(q_{x,t}) = \kappa_t^1 + \kappa_t^2(x - \bar{x}) + \epsilon_{x,t}, \tag{4.5}$$

where $\bar{x}$ is the mean age in the sample range and (κ_t^1, κ_t^2) are assumed to be a bivariate random walk with drift. The two factors in this model were both period factors with no cohort effect allowed for. This was rectified in Cairns *et al.* (2009), namely capturing the cohort effect as an additional effect on top of the two age-period effects. All these models have multiple factors resulting in a non-trivial correlation structure which mirrors the reality that improvements in mortality rates are different for different age ranges. A further adaptation was also created allowing for the cohort effect to diminish over time. The main problem with these models arises from the fact that they were designed for higher ages and so ignored the modelling of mortality at the lower ages (for example the accident hump). Cairns *et al.* (2009) argue that the significant cost associated with mortality is at the older ages and thus their modelling focused on those ages. When using these models for full age ranges, the fit quality is relatively poor and the projections are biologically unreasonable.

Plat (2009) wanted to develop a model which maintained the good aspects of the existing models whilst leaving out the weaker features. The result was a four factor model which took its beginnings from the Lee-Carter model and which added factors to capture the second age-period effect, as per the Cairns *et al.* (2006) model and the cohort effect, as per the Renshaw and Haberman (2006) model. The innovation in the Plat model was to then add a further period factor affecting only the lower ages and designed to allow the model to fit to the whole age range. The model specification is

given by:

$$\ln(m_{x,t}) = a_x + \kappa_t^1 + \kappa_t^2(\bar{x} - x) + \kappa_t^3(\bar{x} - x)^+ + \gamma_{t-x} + \epsilon_{x,t}, \tag{4.6}$$

where the a_x is similar to that of the Lee-Carter model and makes sure that the overall shape of the mortality curve by age is reasonable, the κ_t^1 and κ_t^2 model the mortality rates as in the Cairns *et al.* (2006) model and the κ_t^3 models the effects specific to the lower ages only where $(\bar{x} - x)^+$ takes the value $(\bar{x} - x)$ when this is positive and zero otherwise. Finally the γ_{t-x} models the cohort effect. In 2012 O'Hare and Li (2012) modified the Plat (2009) model to provide a better fit for a wider age range including ages 5-20.[2]

A weakness of the Lee Cater (1992) model that is not addressed by any of the above models and which is a key assumption in each of these models, is that any identified and extracted patterns in the data will not change over time. It is this assumption that we test in this paper.

4.3 Data

The data that we use in this paper comes from the Human Mortality Database.[3] The data available for each country includes number of deaths $D_{x,t}$ and exposure to death $E_{x,t}$ for lives aged x last birthday during year t. We can use this to gain a proxy for the central mortality rate for lives aged x during year t as:

$$m_{x,t} = \frac{D_{x,t}}{E_{x,t}} \tag{4.7}$$

[2]See chapter 2 of this document for the details.

[3]This can be found at http://www.mortality.org/. The database is maintained in the Department of Demography at the University of California, Berkeley, USA, and at the Max Planck Institute for Demographic Research in Rostock, Germany.

Due to the exponential nature of mortality rates we model the logarithmically transformed central mortality rates.

Data is available going back to the mid nineteenth century in some cases but we have restricted this study to data from 1950-2006 and to the countries U.S., U.K., Netherlands and Australia. These countries were chosen to representing a geographical spread around the globe. Specifically, we fit the models to data between 1950-2000 and forecast between the years 2001-2006 testing our forecasts against actual data over that period. We focus on the age range 20-89 for several reasons. Firstly, the papers and models upon which we have based our comparisons are also fitted to this age range. Secondly, and as identified by Currie (2011), data at the older ages provide additional problems in terms of the reliability of the data. Indeed in several cases mortality rates determined using older data appear to fall sharply beyond age 95. Since we are interested in identifying if structural changes are present in the data which re not picked up by the models we plot graphics of the parameters that are fit in the models and eyeball them for changes in direction.

4.4 Methodology

To identify if there are any structural breaks present in the models, we first need to fit the models to the data and extract the corresponding time series κ_t or κ_t^1 (in the case of Cairns Blake Dowd (2006), Plat (2009) and O'hare and Li (2012)). This time series reflects the average mortality rate improvement factor in each of the models and is the main driver of the forecasts of mortality derived from each of the models. We use Box-Jenkins approach to identify the most suitable ARIMA process to fit to the extracted κ_t which is all cases turns out to be a simple random walk with drift. If the selected ARIMA processes are appropriate then we should expect residuals whose mean

does not deviate significantly from zero. We consider a couple of test frameworks for identifying structural breaks, the F test statistics and the Generalised Fluctuations test:

- F statistics (Andrews 1993; Andrews and Ploberger 1994) are designed for a specific alternative, and
- generalized fluctuation tests (Kuan and Hornik 1995) do not assume a particular pattern of deviation from the null hypothesis.

F statistics test against a single-shift alternative of unknown timing, and tests against this alternative are usually based on a sequence of F statistics for a change at time i the OLS residuals $\hat{u}(i)$ from a segmented regression, i.e., one regression for each subsample, with breakpoint i, are compared to the residuals $\hat{u}$ from the unsegmented model via

$$F_i = \frac{\hat{u}^T\hat{u} - \hat{u}(i)^T\hat{u}(i)}{\hat{u}(i)^T\hat{u}(i)\hat{u}(i)/(n-2k)} \tag{4.8}$$

These F statistics are then computed for $i = n_h, \ldots, n - n_h$, for $(n_h > k)$ and H_0 is rejected if their supremum or average or exponential functional, see Andrews and Ploberger (1994), is too large. In applications, $n_h = [n_h]$ is a trimming parameter that is set to $h = 0.15$ in our case. Bai and Perron (1998, 2003) extend this approach to F tests for 0 vs. ℓ breaks and ℓ vs. ℓ + 1 breaks respectively with arbitrary but fixed ℓ.

The generalized fluctuation test framework includes formal significance tests but its philosophy is basically that of data analysis as expounded by Tukey (1962). Essentially, the techniques are designed to bring out departures from constancy in a graphic way instead of parameterizing particular types of departure in advance and then developing formal significance tests intended to have high power against these particular alternatives.[4]. More precisely, the model is fitted to the data and an empirical process

[4]see Brown, Durbin, and Evans (1975) for details

is derived that captures the fluctuation either in residuals or in parameter estimates. Under the null hypothesis these are governed by functional central limit theorems (see Kuan and Hornik 1995) and therefore boundaries can be found that are crossed by the corresponding limiting processes with fixed probability α under the null hypothesis. Under the alternative the fluctuation in the process is in general increased. Also, the trajectory of the process often sheds light on the type of deviation from the null hypothesis such as the dating of the structural breaks. In this paper we carry out the empirical fluctuations tests leaving the F-stat tests for future research. We compute several empirical fluctuations processes based upon the residuals described in Zeileis (2005) and repeated below for information:

4.4.1 Cumulative sums of residuals - CUSUM processes

The first type of empirical fluctuations process that can be computed are (CUSUM) processes, which contain cumulative sums of standardized residuals, *Rec-CUSUM*. Brown et al. (1975) suggested to consider cumulative sums of recursive residuals:

$$W_n(t) = \frac{1}{\hat{\sigma}\sqrt{\eta}} \sum_{i=k+1}^{k+[t_\eta]} \tilde{u_i} \tag{4.9}$$

where $\eta = n - k$ is the number of recursive residuals and $[t_\eta]$ is the integer part of t_η.

Under the null hypothesis the limiting process for the empirical fluctuation process $W_n(t)$ is the Standard Brownian Motion (or Wiener Process) $W(t)$. More precisely the following functional central limit theorem (FCLT) holds:

$$W_n \Rightarrow W \tag{4.10}$$

as $n \Rightarrow \infty$, where $\Rightarrow$ denotes weak convergence of the associated probability measures.

Under the alternative, if there is just a single structural change point t_0, the recursive residuals will only have zero mean up to t_0. Hence the path of the process should be close to 0 up to t_0 and leave its mean afterwards.

Ploberger and Krämer (1992) suggested to base a structural change test on cumulative sums of the common OLS residuals. Thus, the *OLS-CUSUM* type empirical fluctuation process is defined by:

$$W_n^0(t) = \frac{1}{\hat{\sigma}\sqrt{n}} \sum_{i=1}^{[nt]} \tilde{u_i} \tag{4.11}$$

The limiting process for $W_n^0(t)$ is the standard Brownian bridge $W^0(t) = W(t) - tW(1)$. It starts in 0 at $t = 0$ and it also returns to 0 for $t = 1$. Under a single structural shift alternative the path should have a peak around t_0.

4.4.2 Moving sums of residuals - MOSUM processes

Another possibility to detect a structural change is to analyze moving sums of residuals, *Rec-MOSUM* (instead of using cumulative sums of the same residuals). The resulting empirical fluctuation process does then not contain the sum of all residuals up to a certain time t but the sum of a fixed number of residuals in a data window whose size is determined by the bandwidth parameter $h \in (0, 1)$ and which is moved over the whole sample period. Hence the Recursive MOSUM process is defined by

$$M_n(t|h) = \frac{1}{\hat{\sigma}\sqrt{\eta}} \sum_{i=k+[N_\eta t]+1}^{k+[N_\eta t]+[\eta h]} \tilde{u_i}, (0 \leq t \leq 1-h) \tag{4.12}$$

where $N_\eta = (\eta - [\eta h])/(1-h)$.

Similarly the OLS-based MOSUM process, *OLS-MOSUM* is defined by

$$M_n^0(t|h) = \frac{1}{\hat{\sigma}\sqrt{n}}\Big(\sum_{i=k+[N_n t]+1}^{k+[N_n t]+[nh]} \tilde{u_i}\Big) = W_n^0\Big(\frac{[N-nt]+[nh]}{n}\Big) - \Big(\frac{[N-nt]}{n}\Big)(0 \le t \le 1-h) \tag{4.13}$$

where $N_n = (n - [nh])/(1-h)$. As the representations 4.13 and 4.12 suggest, the limiting process for the empirical MOSUM processes are the increments of a Brownian motion or a Brownian bridge respectively. This is shown in detail in Chu et al. (1995). If again a single structural shift is assumed at t_0, then both MOSUM paths should also have a strong shift around t_0.

4.5 Fitting and forecasting results for Lee Carter and its variants

Each of the models we consider in this paper improves upon the previous one by incorporating additional patterns identified in the data and forecasting these. The key assumption in each of these models is that future patterns in mortality can be ascertained from the past patterns and indeed these do not change over time. Fitting each of the models to the data over the period from 1950-2000 the tables below present the results using the mean average percentage error (E1), the Mean absolute percentage error (E2) and the root mean square error (E3). The definitions of these are set out below for information;

The average error, E1 – this equals the average of the standardized errors i.e.

$$E1 = \frac{1}{X_1 - X_2 + 1} \sum_{x=X_1}^{X_2} \sum_{t=1}^{T} \frac{projected(m_{x,t}) - actual(m_{x,t})}{projected(m_{x,t})} \tag{4.14}$$

This is a measure of the overall bias in the projections.

The average absolute error, E2 – this equals the average of absolute value of the standardized errors i.e.

$$E2 = \frac{1}{X_1 - X_2 + 1} \sum_{x=X_1}^{X_2} \sum_{t=1}^{T} |\frac{projected(m_{x,t}) - actual(m_{x,t})}{projected(m_{x,t})}| \tag{4.15}$$

This is a measure of the magnitude of the differences between the actual and projected rates.

The standard deviation of the error, E3 – this equals the square root of the average of the squared errors,

$$E3 = \sqrt{\frac{1}{X_1 - X_2 + 1} \sum_{x=X_1}^{X_2} \sum_{t=1}^{T} (\frac{projected(m_{x,t}) - actual(m_{x,t})}{projected(m_{x,t})})^2} \tag{4.16}$$

As can be seen from the fitting results 4.1the models, excluding the Cairns Blake Dowd model which was designed for older ages, the accuracy of the fit is very good for all of the above models. Indeed the errors are in the main less that a few percentage points suggesting that each model is adequately capturing the variability present in past mortality data. However, if mortality models are to be of any use they need to adequately forecast mortality rates. Again, each of the following models have been backtested and the results of forecasting from 2001-2006 have been tested and measured against each of the three error measures. Table 4.2 shows the forecasting results when each model is backtested over the years 2001-2006.

Whilst the forecasting and fitting results look reasonable it is difficult to see from

Table 4.1: Fitting results for US, UK, Netherlands and Australia male mortality rates by single age 20-89, 1950-2000 measured on E1 E2 and E3

USA	E1	E2	E3
Lee and carter (1992)	0.4%	3.8%	0.3%
Cairns Blake Dowd (2006)	-1.6%	12.0%	3.3%
Plat (2009)	0.2%	2.8%	0.2%
Ohare and Li (2012)	0.1%	2.6%	0.1%
UK	E1	E2	E3
Lee and carter (1992)	0.5%	4.9%	0.5%
Cairns Blake Dowd (2006)	0.4%	13.8%	4.3%
Plat (2009)	0.3%	2.8%	0.2%
Ohare and Li (2012)	0.2%	2.8%	0.2%
NTH	E1	E2	E3
Lee and carter (1992)	0.0%	6.1%	0.7%
Cairns Blake Dowd (2006)	-1.9%	13.7%	4.5%
Plat (2009)	0.3%	4.0%	0.4%
Ohare and Li (2012)	0.2%	3.9%	0.3%
AUS	E1	E2	E3
Lee and carter (1992)	0.4%	5.2%	0.6%
Cairns Blake Dowd (2006)	-1.3%	16.4%	6.1%
Plat (2009)	0.7%	4.8%	0.6%
Ohare and Li (2012)	0.6%	4.6%	0.5%

Table 4.2: Forecasting results for US, UK, Netherlands and Australia male mortality rates by single age 20-89, 2001-2006 measured on E1 E2 and E3

USA	E1	E2	E3
Lee and Carter (1992)	2.8%	8.0%	1.2%
Cairns, Blake and Dowd (2006)	-3.6%	16.4%	5.6%
Plat (2009)	-3.6%	7.4%	0.8%
Ohare and Li (2012)	-0.1%	5.5%	0.5%
UK	E1	E2	E3
Lee and Carter (1992)	-1.7%	9.3%	1.2%
Cairns, Blake and Dowd (2006)	0.4%	24.0%	9.3%
Plat (2009)	3.3%	9.6%	2.2%
Ohare and Li (2012)	7.6%	9.6%	2.1%
NTH	E1	E2	E3
Lee and Carter (1992)	7.4%	10.6%	2.0%
Cairns, Blake and Dowd (2006)	4.4%	22.7%	8.5%
Plat (2009)	5.4%	10.0%	1.7%
Ohare and Li (2012)	11.1%	13.2%	3.0%
AUS	E1	E2	E3
Lee and Carter (1992)	4.8%	11.5%	2.6%
Cairns, Blake and Dowd (2006)	2.3%	31.0%	14.0%
Plat (2009)	8.9%	13.2%	6.5%
Ohare and Li (2012)	13.8%	16.3%	6.6%

the tabular results the issue of structural breaks. To demonstrate the potential problems with the above models we plot the extracted main period effects below in figure 4.1 and in the appendix figures 4.4-4.6. Note for the Lee Carter model there is only 1 period effect, κ_t, whilst for the for the Cairns Blake Dowd model there are two, κ_t^1 and κ_t^2 (we plot κ_t^1) and for the Plat (2009) and Ohare and Li (2012) models there are 3, κ_t^1, κ_t^2, and κ_t^3 (again, we plot κ_t^1) for each of the countries considered.

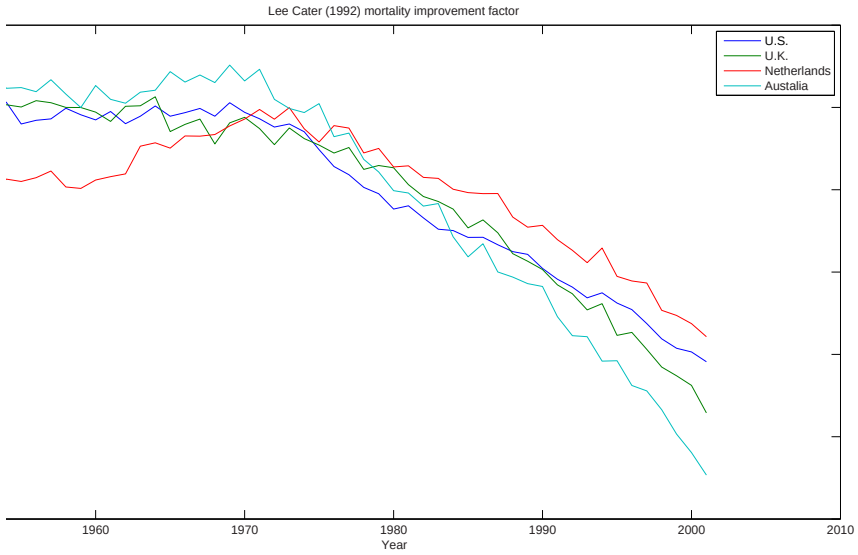

Figure 4.1: Plots of the κ_t factor for the Lee carter model for countries Australia (purple), Great Britain (orange), England and Wales (green), USA (blue), Japan (light green) and New Zealand (light blue)

As can be seen from the diagrams the trend in each of the time series appears to shift and in particular to accelerate as we move beyond the 1970's. For the Lee Carter (1992) model this is clear in the graph 4.1 where a peak has occurred around 1970 for each of the countries considers with the value of κ_t falling rapidly thereafter. Equivalent figures for the Cairns, Blake, Dowd (20060, Plat (2009) and O'Hare and Li (2012) models can be found in the appendix figures 4.4-4.6. With the Cairns, Blake, Dowd (2006) model the same comments as for the Lee Carter (1992) κ_t can be made

for the parameter κ_t^1 (albeit with a less obvious change in direction) but the direction of the κ_t^2 parameter is less clear. For the Plat (2009) and the O'Hare and Li (2012) models the direction change is clear for the κ_t^1 parameter but again, whilst a shift can be seen in κ_t^2 and κ_t^3, the shift is less dramatic. In each of these models it is the κ_t^1 parameter which drives the mortality forecast for all ages. The other time series parameters reflect the variation in the logarithmically transformed mortality rates specific to certain age groups but not to all age groups. As such, it is the forecast of κ_t^1 that is the most important to get right. If we do not forecast the direction of mortality improvement correctly then this will cause problems not least because most mortality models use a fitting period from 1950 onwards and so may underestimate this improvement rate. This could have severe implications for the pricing of annuity and life products which are sensitive to the mortality rate.

4.6 Identifying structural breaks

In the following sections we now formalise our tests for structural breaks in the time series that we have extracted from each of the models in the previous section. To clarify, we have fit each of the models to the Male mortality data of the U.S., U.K., Netherlands and Australia between the years 1950 - 2000 inclusive. We have then taken the main mortality improvement series κ_t or κ_t^1 from each of the models and fitted a random walk with drift process as the Box-Jenkins identified best ARIMA process to fit the time series. If the random walk process is indeed an appropriate time series capturing the κ_t or κ_t^1 factor then the resulting residuals should have a mean which does not deviate from zero. We test this using the empirical fluctuations test framework described by Zeileis *et al* (2003) and the function *efp* which sits within the R

package *strucchange* which is freely available[5]. Specifically we use the methods of Bai and Perron to identify structural breaks. The foundation for estimating breaks in time series regression models was given by Bai (1994) and was extended to multiple breaks by Bai (1997ab) and Bai & Perron (1998). The *breakpoints* function in the *Strucchange* package implements the algorithm described in Bai & Perron (2003) for simultaneous estimation of multiple breakpoints given we know they exist. The distribution function used for the confidence intervals for the breakpoints is given in Bai (1997b). The ideas behind this implementation are described in Zeileis *et al.* (2003).

The structural change is located and dated by plotting the cumulative sum of residuals and then comparing this with the known limiting processes. Fluctuations that fall outside these known boundaries are judged to be improbably large and hence suggest a structural change in the mean value. The known limiting process, otherwise known as the boundary is plotted along with the cumulative sum of residuals in figures 4.2 below and 4.7 to 4.9 in the appendix.

As can be seen from the above tests there are unexpectedly large fluctuations (demonstrating the presence of a structural break) for the majority of cases considered. However, the results are not as conclusive as in the cases of Li Chan and Cheung (2011) and Coelho and Nunes (2011). For the Lee Carter model there appears to be a structural break present for 3 out of the 4 countries considered with the only country for which the above test does not confirm the presence of a structural break being the U.K. In the case of the Cairns, Blake and Dowd (2006) model we have similar conclusions with again the extracted mortality improvement factor κ_t^1, for the U.K. not demonstrating any presence of a structural break under the test we have carried out. For the larger factor models of Plat (2009) and O'hare and Li (2012) the results are mixed again. In the case of Plat (2009), the fluctuations test showed up a struc-

[5]The package strucchange is available at `http://cran.r-project.org/web/packages/strucchange/index.html` and can be implemented using the statistical software package R

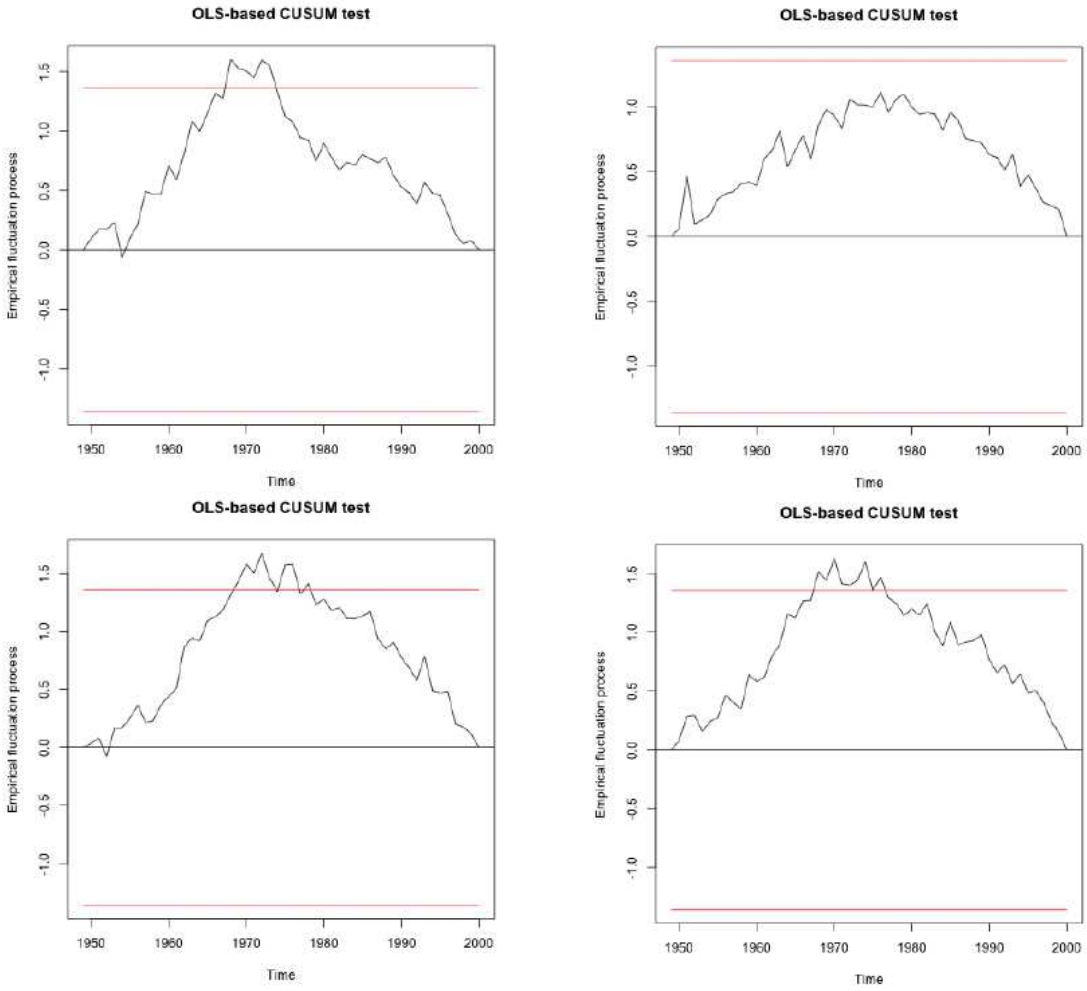

Figure 4.2: Cumulative sum of residuals test for the Lee Carter (1992) model for (from top left clockwise) US, UK, Netherlands and Australia

tural break for all four countries considered whilst the O'Hare and Li (2012) model did not show up any structural breaks for the U.S. or the U.K. This could be due to the quadratic effect parameter α_x^1 applied to the κ_t^1 in the case of the O'hare and Li (2012) model as compared with the linear α_t^1 in the Plat (2009) case. In all the models considered we were able to identify some structural breaks for some countries.

Having identified the presence of structural breaks the next step is to date these and then using this information, re-forecast the models allowing for the structural break. We date the structural break using the functions *breakpoints* within the package *strucchange*. As can be seen from the table 4.3 and the figures 4.3 below and figures 4.10 - 4.12 in the appendix the structural breaks identified occur with the period 1968 - 1979 with the vast majority occurring in the early 1970's. This confirms previous findings and would be well worth further investigation.

4.7 Empirical analysis of modelling with and without structural changes

Having identified and dated the presence of a structural break in each of the κ_t factors for our models we fit each of the models again but this time allowing for the structural break. Again we fit up to the year 2000 and forecast from 2001 to 2006 inclusive. We allow for the identified acceleration in the mortality improvement factor κ_t by refitting the model using a data set in each case which excludes data prior to the year in which the structural break occurs. Again we measure the forecasting quality using the average error (E1), the mean absolute error (E2) and the root mean square error (E3). The results are outlined in table 4.4 with and without this adjustment.

As can be seen in table 4.4 in two-thirds of the cases allowing for the structural break results in a more accurate forecast measured on any of the three measures con-

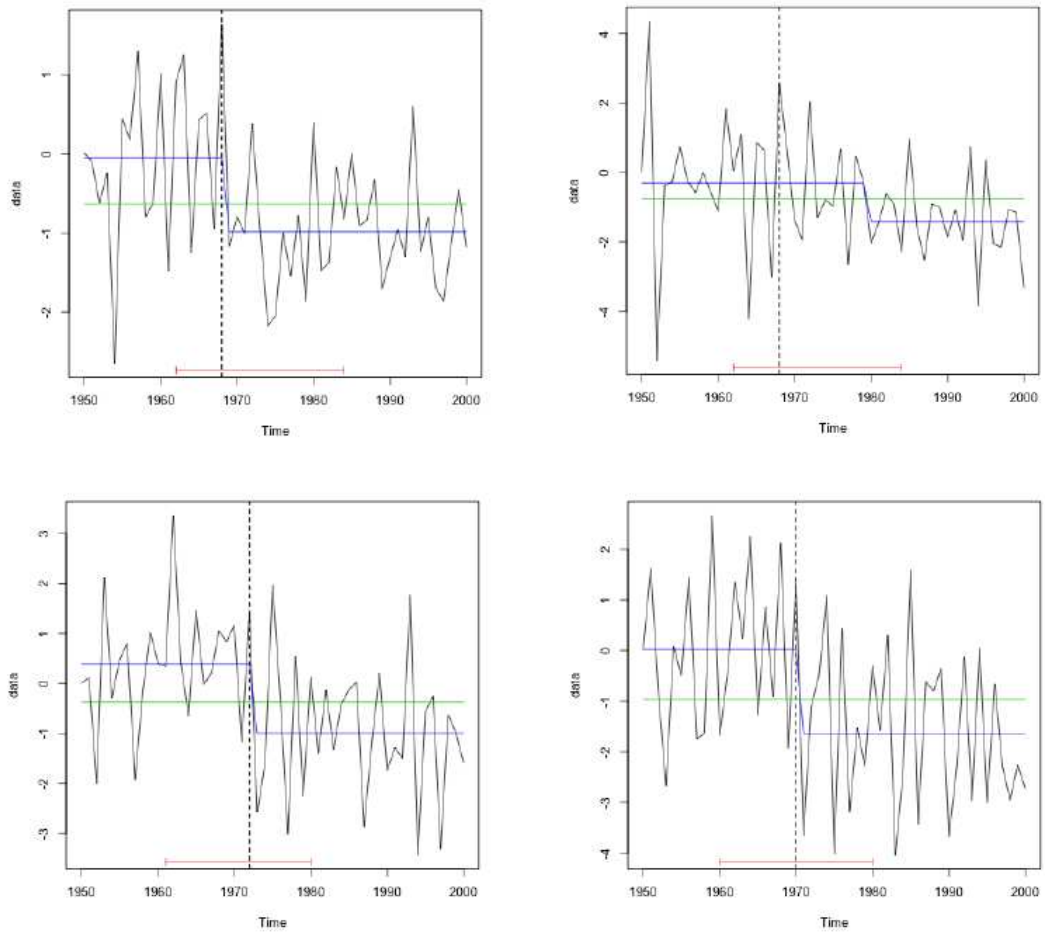

Figure 4.3: Test of the structural break for the Lee Carter (1992) model for (from top left clockwise) US, UK, Netherlands and Australia

sidered. In the case of the Netherlands, in every case allowing for the structural break improved the result. This indicates that structural changes should be allowed for if we are to accurately forecast mortality rates.

4.8 Conclusions

In this paper we have considered several of the leading extrapolative models of mortality rates and have applied the methods of Bai and Perron (2003) to test for the presence of structural breaks in the model specifications. More specifically we have fitted the models of Lee and Carter (1992), Cairns, Blake and Dowd (2006), Plat (2009) and O'Hare and Li(2012) to the data for U.S., U.K., Australia and the Netherlands. Having noted that the forecasts of mortality resulting from these models are driven by the κ_t^1 parameter, we have fitted the best ARIMA process to the extracted time series (in each case the ARIMA process used was a random walk with drift) and then tested the residuals for deviation from zero.

In each case we found that there was indeed a breakpoint visible in the residuals falling somewhere around the 1970's confirming previous demographic research. We then carried out the forecasting process again making allowance for the structural break providing the results in section 5. The results show that in nearly two-thirds of cases the model allowing for structural breaks provides a more accurate forecast measured on each of the E1, E2 and E3 measures. Whilst the findings are important in highlighting the importance of the sample period when fitting a model to mortality data they make no reference to future structural breaks. Further research could look at more recent developments in the identification of structural breaks in models. Namely, monitoring data for structural breaks as and when they occur. This would then allow for these breaks to be incorporated into mortality models more efficiently reducing any future

forecast errors.

4.9 Appendix: Additional Figures and Tables

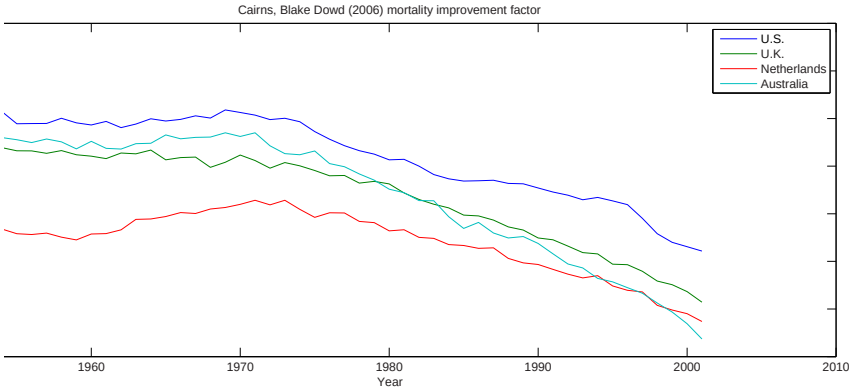

Figure 4.4: Plots of the κ_t^1 factor for the CBD 2 factor model for countries Australia (purple), Great Britain (orange), England and Wales (green), USA (blue), Japan (light green) and New Zealand (light blue)

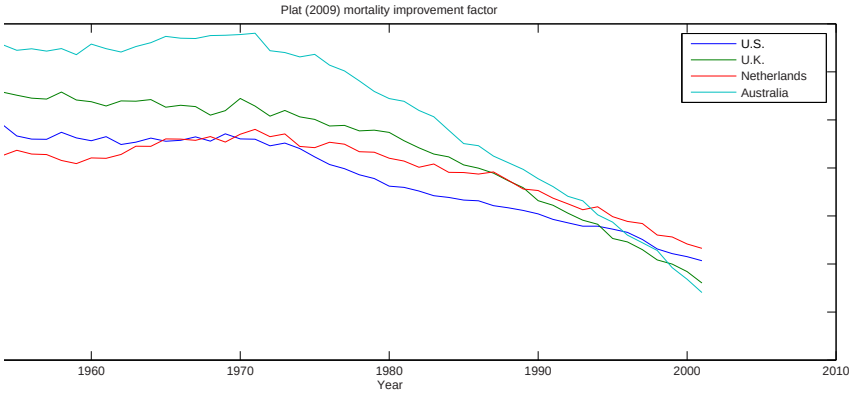

Figure 4.5: Plots of the κ_t^1 factor for the Plat(2009) model for countries Australia (purple), Great Britain (orange), England and Wales (green), USA (blue), Japan (light green) and New Zealand (light blue)

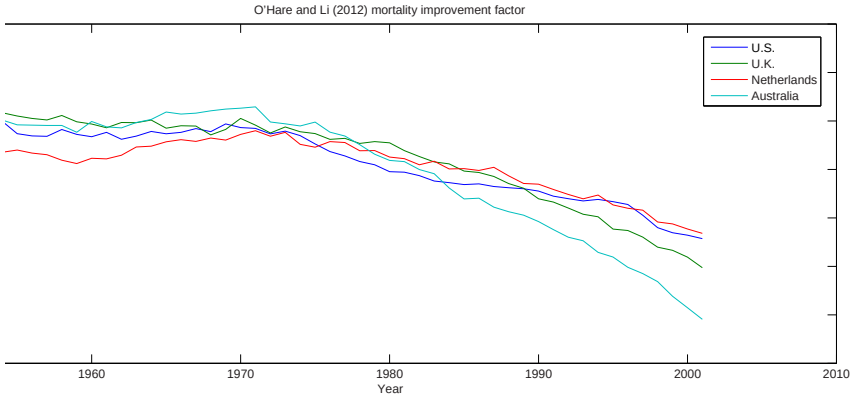

Figure 4.6: Plots of the κ_t^1 factor for the O'Hare and Li (2012) model for countries Australia (purple), Great Britain (orange), England and Wales (green), USA (blue), Japan (light green) and New Zealand (light blue)

Table 4.3: Break date results for US, UK, Netherlands and Australia using the Lee Carter (1992), Cairns, Blake Dowd (2006), Plat (2009) and O'Hare and Li (2012) models

Model and Country	Break date	test statistic	p-value
Lee Carter (1992)			
U.K.	1979	1.116	0.166
U.S.	1968	1.603	0.012
Netherlands	1972	1.673	0.007
Australia	1970	1.623	0.010
Cairns Blake Dowd (2006)			
U.S.	1979	1.248	0.089
U.S.	1972	1.582	0.013
Netherlands	1972	1.863	0.002
Australia	1970	1.822	0.003
Plat (2009)			
U.K.	1979	1.566	0.015
U.S.	1972	1.392	0.041
Netherlands	1970	1.728	0.005
Australia	1970	2.361	0.001
O'Hare and Li (2012)			
U.K.	1979	1.366	0.048
U.S.	1972	1.288	0.072
Netherlands	1972	1.648	0.009
Australia	1970	2.126	0.001

Table 4.4: Forecasting results for US, UK, Netherlands and Australia male mortality rates by single age 20-89, 2001-2006 measured on E1 E2 and E3 with and without allowance for structural breaks

	Structural break			No Structural break		
	E1	E2	E3	E1	E2	E3
			U.S.			
LC (1992)	**1.5%**	**7.3%**	**1.0%**	2.8%	8.0%	1.2%
CBD (2006)	-6.2%	**15.2%**	**5.3%**	-3.6%	16.4%	5.6%
Plat (2009)	-6.4%	9.0%	1.3%	-3.6%	7.4%	0.8%
Ohare and Li (2012)	-6.1%	8.5%	1.4%	-0.1%	5.5%	0.5%
			U.K.			
LC (1992)	3.3%	**9.0%**	1.6%	-1.7%	9.3%	1.2%
CBD (2006)	-4.8%	**20.0%**	**7.5%**	0.4%	24.0%	9.3%
Plat (2009)	**3.3%**	**9.5%**	2.5%	3.3%	9.6%	2.2%
Ohare and Li (2012)	**3.8%**	12.0%	4.1%	7.6%	9.6%	2.1%
			Netherlands			
LC (1992)	**5.5%**	**9.4%**	**1.6%**	7.4%	10.6%	2.0%
CBD (2006)	**0.8%**	**20.7%**	**7.5%**	4.4%	22.7%	8.5%
Plat (2009)	**4.3%**	**9.3%**	**1.5%**	5.4%	10.0%	1.7%
Ohare and Li (2012)	**4.1%**	**10.1%**	**1.7%**	11.1%	13.2%	3.0%
			Australia			
LC (1992)	**4.6%**	12.1%	3.2%	4.8%	11.5%	2.6%
CBD (2006)	**-1.7%**	**28.7%**	**12.6%**	2.3%	31.0%	14.0%
Plat (2009)	**6.6%**	**13.1%**	**6.7%**	8.9%	13.2%	6.5%
Ohare and Li (2012)	**11.8%**	29.6%	28.6%	13.8%	16.3%	6.7%

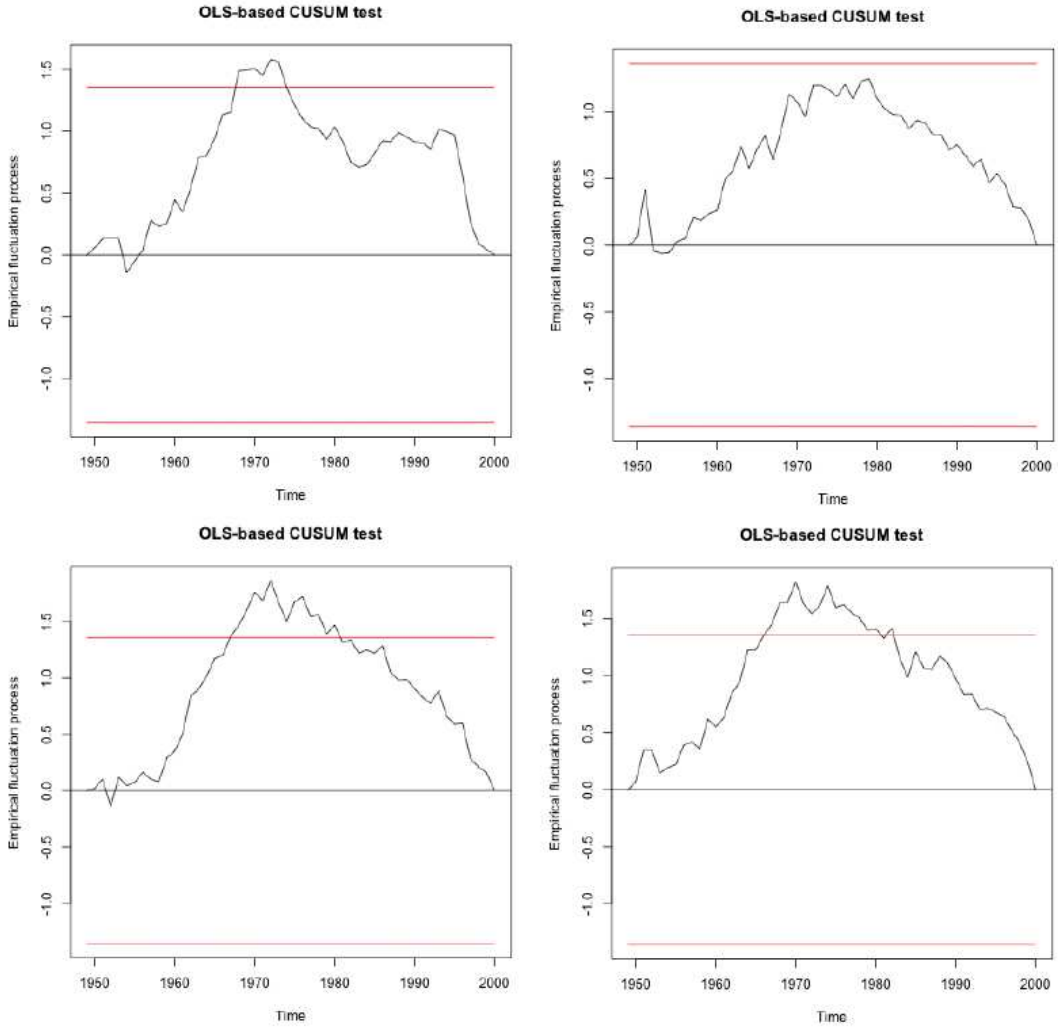

Figure 4.7: Cumulative sum of residuals test for the Cairns, Blake and Dowd (2006) model for (from top left clockwise) US, UK, Netherlands and Australia

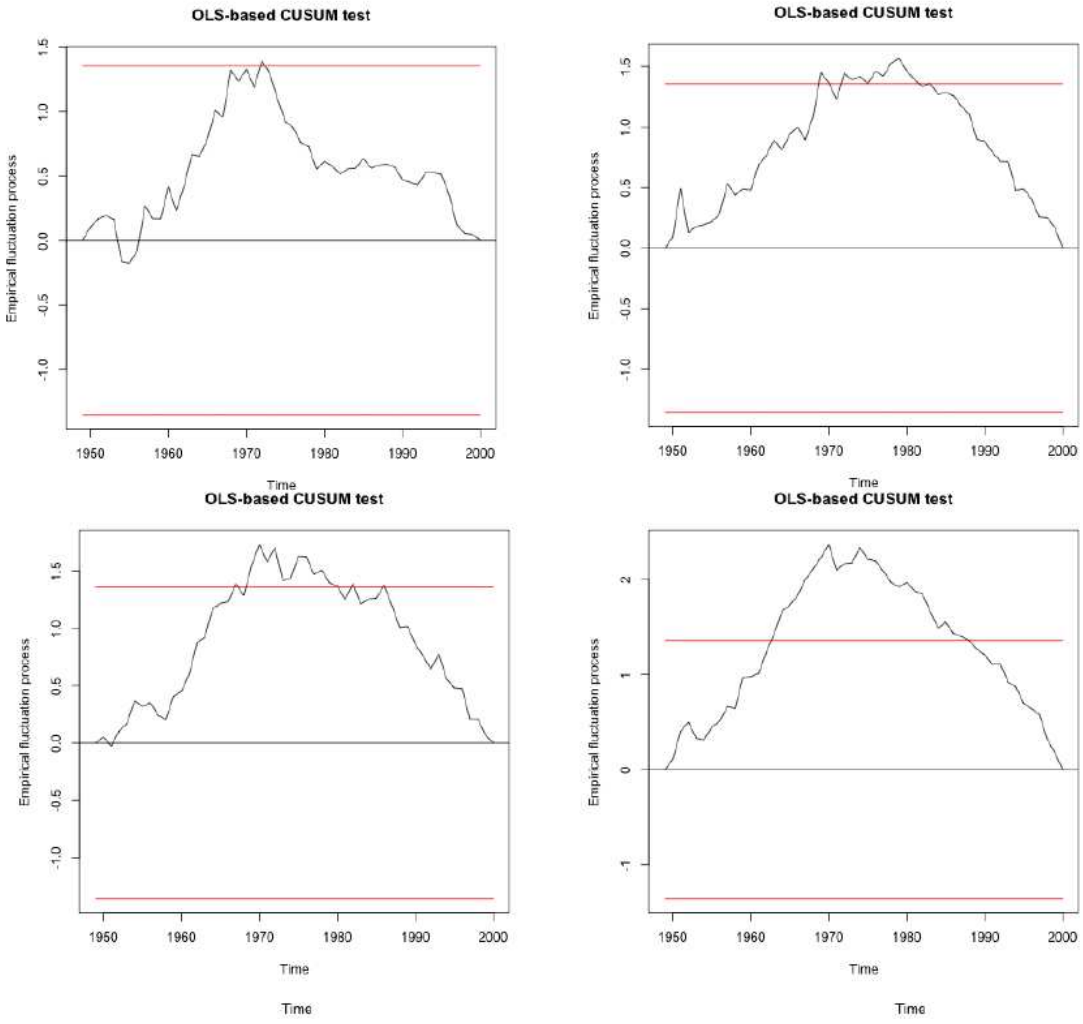

Figure 4.8: Cumulative sum of residuals test for the Plat (2009) model for (from top left clockwise) US, UK, Netherlands and Australia

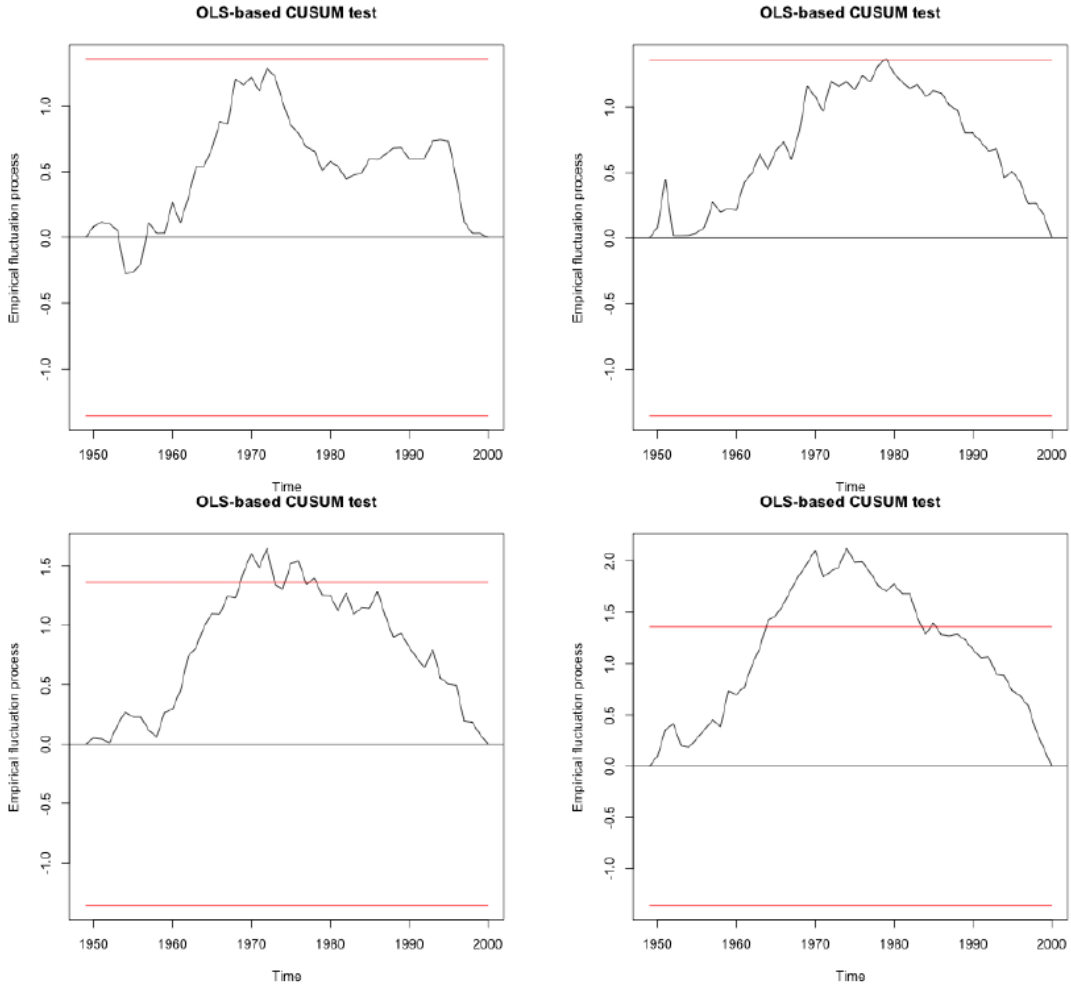

Figure 4.9: Cumulative sum of residuals test for the O'Hare and Li (2012) model for (from top left clockwise) US, UK, Netherlands and Australia

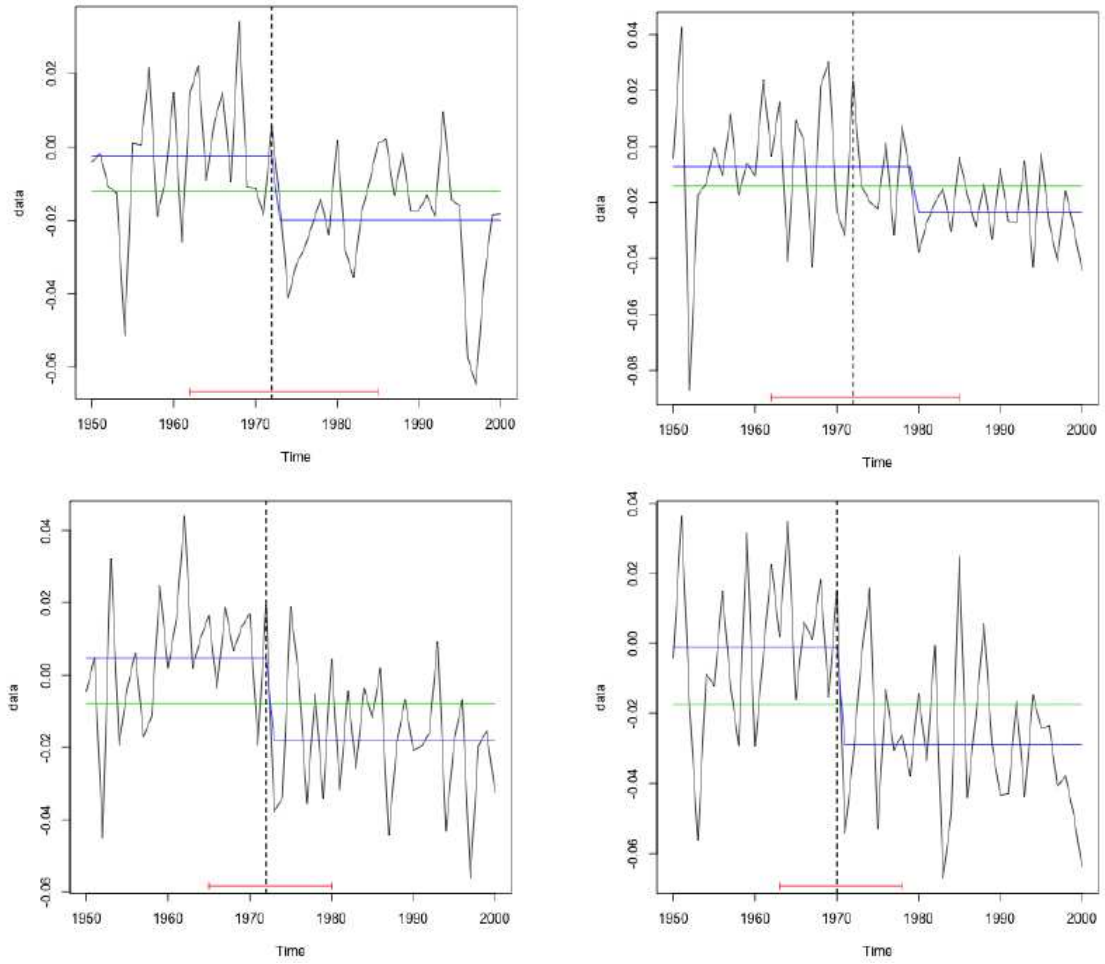

Figure 4.10: Test of the structural break for the Cairns, Blake, Dowd (2006) model for (from top left clockwise) US, UK, Netherlands and Australia

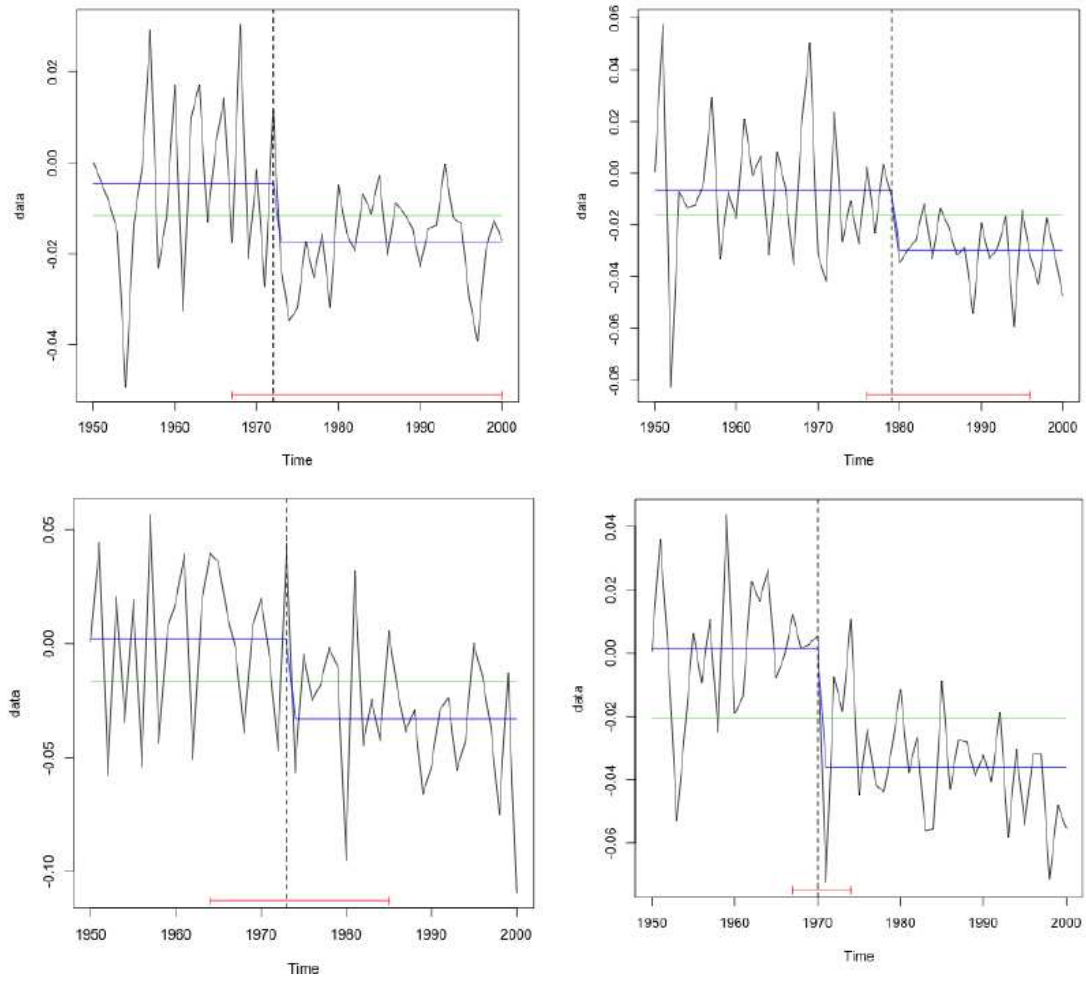

Figure 4.11: Test of the structural break for the Plat (2009) model for (from top left clockwise) US, UK, Netherlands and Australia

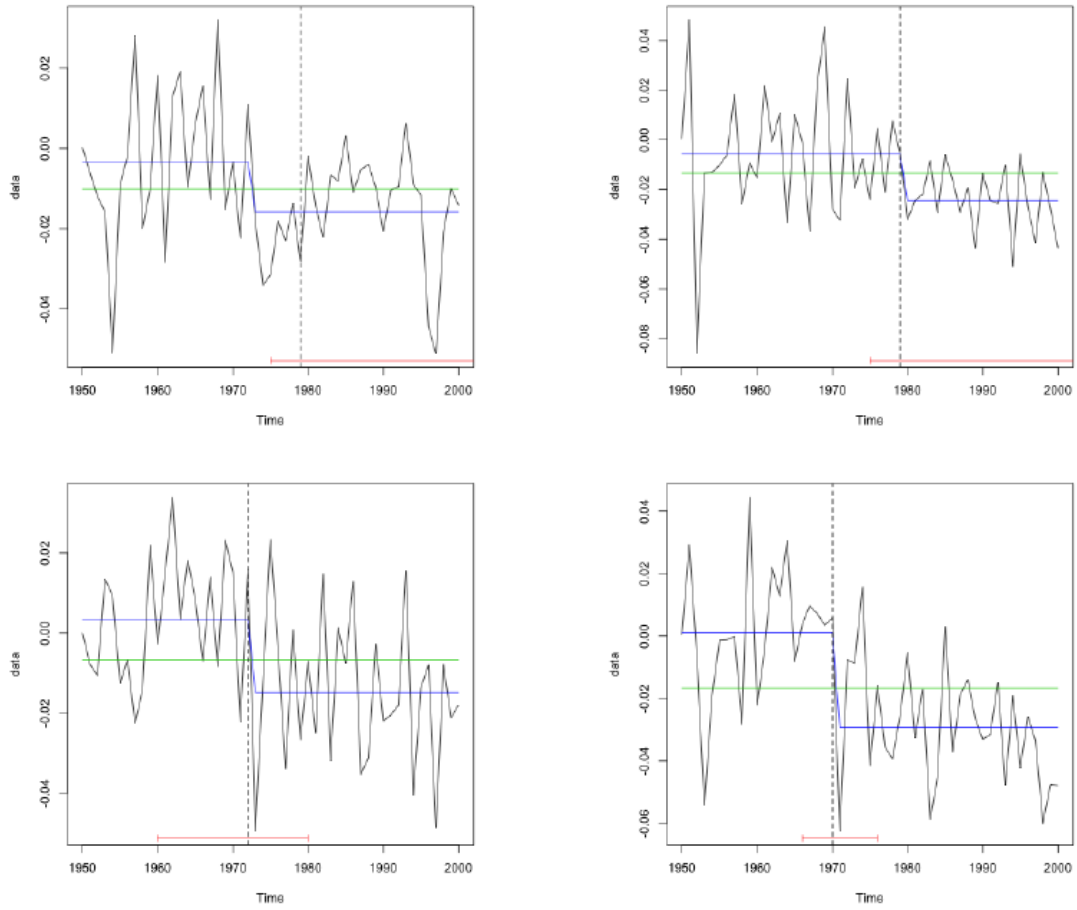

Figure 4.12: Test of the structural break for the O'hare and Li (2012) model for (from top left clockwise) US, UK, Netherlands and Australia

Chapter 5

Spatial modelling of mortality rates

5.1 Introduction

Spatial variability in mortality rates as well as the effect of socioeconomic factors have attracted increased attention recently. Life insurers issuing life and annuity products allow for significant risk factors known to affect mortality rates including age, gender, and smoking status. In life insurance adjustments are made to mortality rates used for premiums based on health status. With the increase in sales of annuity products insurers are offering impaired lives annuity rates. Postcode underwriting reflecting geographical variation in mortality is increasingly used by insurers in mainland UK.

Mortality is known to vary spatially but should we expect regions in close proximity to one another to have similar mortality characteristics, otherwise known as spatial clustering? Arguments for spatial clustering might include a lack of social mobility, or exposure to similar socioeconomic externalities. Arguments against spatial clustering may include the increased social mobility of individuals and the dispersion of public health interventions. Significant geographical variation in mortality occurs in many countries. In the United Kingdom, socioeconomic factors are implicitly allowed for

in insurance based on annuity amounts. For example individuals with larger annuity amounts have lower mortality rates. However, this proxy is not very robust and indeed Richards (2008) shows that a mortality model using geographic classifications better fits United Kingdom annuitant mortality than a model using pension amounts. In this paper we assess the geographical variation in mortality rates in Northern Ireland. We investigate mortality rates to see if the same conclusions can be drawn for Northern Ireland data and we find a more mixed set of results. In particular, we find that allowing for spatial frailties does not improve the model fit dramatically and indeed the additional parameters necessary lead to a poor fit when measured using an information criterion. It is common in general insurance for companies to collate significant amounts of data (risk factor data) in order to appropriately price general insurance products (car, home, contents, insurances etc.) however in the life insurance industry this is less common. Life insurance companies need to assess mortality heterogeneity more accurately for pricing and risk management of insurance policies if they are to avoid adverse selection issues for example. With the collection of measures of deprivation this should be more possible. We conclude in this paper that the information available from socioeconomic profiling is valuable in pricing longevity risk where even within a locality there is significant heterogeneity.

The paper is laid out as follows. In the next section we give a summary of the methods that we use to model our mortality rates in terms of socioeconomic covariates (risk factors) with and without an allowance for spatial frailties. In section 3 we summaries the data used in the analysis, looking at each of the socioeconomic factors in turn exploring its variation across the region. In section 4 we summarise the results of this study, we begin with the basic logistic regression model then carry out a general to specific analysis. Finally we introduce two extensions to the simplified logistic model and discuss the findings. Section 5 concludes.

5.2 Literature review and Models of mortality

Spatial models have been developed and applied to modelling house prices, crime levels and diseases amongst many others[1]. Rosen (1974) models house prices using spatial covariates including environmental attributes and geographical characteristics. Waller *et al.* (2007) models geographic variation in alcohol distribution and violent crime in Houston. Kazembe (2007) examined spatial clustering of malaria risks in northern Malawi. Geodemographic modelling, the spatial modelling of demographic data, is used in a range of applications. Commercial applications include customer profiling for product marketing and development. Grubesic (2004) applies geodemographic models to assess broadband access. Richards (2008) uses geodemographic profiles based on postcodes to analyse life insurance and pension scheme mortality. Tuljapurkar and Boe (1998) outline mortality differentials by sex, education and socio-economic variables. Richards and Jones (2004) discuss the impact of socio-economic status on mortality rates in the UK. For Northern Ireland, there is limited formal modelling and analysis of mortality variation by geographical location using spatial models and limited analysis of variation of mortality according to socio-economic risk factors.

Mortality can be modeled using survival (time to death) data or aggregate death rate data. If data on individual characteristics including death dates are available then hazard rates can be estimated using proportional hazards models to quantify the effect of covariates[2]. For aggregate data on deaths and exposures, the effect of covariates on death rates can be estimated using logistic regressions. Both approaches to modelling can be modified to include spatial variation. Frailty models are used for heterogeneity in mortality rates to account for unobserved covariates (Vaupel *et al.*, 1979). Frailty

[1] For a review of the various applications see Sherris and Tang (2010)

[2] A proportional hazards model including covariates for different geographic regions are discussed in Sherris and Tang (2010)

models can also be used to capture spatial variation and unobserved heterogeneity. Banerjee and Carlin (2003) provide details of spatial frailty models in the hierarchical Bayes model. As the data we have in this study is at an aggregate level we will be using a logistic regression model to fit the data with frailty modified versions to investigate spatial clustering.

5.2.1 Spatial Logistic Modelling

With aggregate level death data the logistic regression model[3] can be used. The time to death or "event time" data t_{ij} is replaced with an indicator:

$$Y_{ij} = \begin{cases} 1 & \text{if survived} \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

and $\rho_{ij} = Pr(Y_{ij} = 1)$, has a logistic form:

$$\text{logit}(\rho_{ij}) = \beta^T \mathbf{x_{ij}}, \tag{5.2}$$

Independent frailties can be included in such a model by including the element W_i as below:

$$\text{logit}(\rho_{ij}) = \beta^T \mathbf{x_{ij}} + \mathbf{W_i}, \tag{5.3}$$

where W_i is the frailty term for region i, β are the parameters and $\mathbf{x_{ij}}$ are the individual-specific covariates for the jth subject in region i, and where simple i.i.d. specifications

[3] Ingram and Kleinman (1989) and Doksum and Gasko (1990) show that the results for the β parameters can be quite similar in the two different models when the probability of death is small or where there is no censoring. However, since the proportional hazards model is based on more information than the logistic regression model, Banerjee *et al.* (2003) note that the proportional hazards model should be more powerful in detecting significant covariate effects.

including the gamma distribution, the log-normal distribution and the normal distribution, can be assumed for W_i (McGilchrist and Aisbett, 1991; Wienke, 2003).

We can further investigate whether spatial clustering present by adjusting the assumption of independent frailties. With independent frailties the models do not account for spatial clustering, but these effects can be modeled using continuous geostatistical models or discrete lattice models. The geostatistical approach uses the exact geographical location of a region (Cressie, 1993). Frailties W are indexed continuously throughout a geographical region D.[4] In the case of Northern Ireland, where we have aggregated death data for distinct regions we will be using a lattice modelling approach to investigate spatial clustering.

5.2.2 Lattice Modelling

When W is defined only for discrete regions such that the regions form a partition of the geographical study space D, then this is the lattice model. Banerjee and Carlin (2002) use the conditionally autoregressive model (CAR) for the prior distribution:

$$\mathbf{W}|\lambda \sim \mathbf{CAR}(\lambda), \tag{5.4}$$

introduced in Besag *et al.* (1991). Bernardinelli and Montomoli (1992) refer to the most common form of this prior having the following joint distribution:

$$\mathbf{W}|\lambda \propto \lambda^{\mathbf{1/2}}\mathbf{exp}\Big[\frac{\lambda}{\mathbf{2}}\sum_{\mathbf{i=1}}^{\mathbf{I}}\sum_{\mathbf{j=1}}^{\mathbf{n_i}}\alpha_{\mathbf{ij}}\mathbf{W_i}\big(\mathbf{W_i} - \frac{\mathbf{1}}{\sum_{\mathbf{j=1}}^{\mathbf{n_i}}\alpha_{\mathbf{ij}}}\sum_{\mathbf{j=1}}^{\mathbf{I}}\alpha_{\mathbf{ij}}\mathbf{W_j}\big)\Big], \tag{5.5}$$

where α_{ij} represents the weights between region i and region j. With spatial correlation, higher weights should be assigned to regions in closer proximity to each other.

[4]For a geostatistical model a prior distribution, given observations W_i for known locations i, $i = 1 \ldots I$, is used for the unobserved frailty values at other target locations.

For example,

$$\alpha_{ij} = \begin{cases} 1 & \text{i adj j,} \\ 0 & \text{otherwise} \end{cases} \tag{5.6}$$

in which case Eq. (5.5) simplifies to:

$$\mathbf{W}|\lambda \propto \lambda^{\mathbf{1/2}}\mathbf{exp}\Big[-\frac{\lambda}{\mathbf{2}}\sum_{\mathbf{i=1}}^{\mathbf{I}}\mathbf{m_i}\mathbf{W_i}(\mathbf{W_i}-\hat{\mathbf{W}}_\mathbf{j})\Big] \tag{5.7}$$

where $\hat{\mathbf{W}}_\mathbf{i}$ is the average of the frailties $W_{j\neq i}$ that are adjacent to region i, and m_i is the number of the adjacent pairs. This then gives,

$$\mathbf{W}|\lambda \propto \mathbf{N}\Big(\hat{\mathbf{W}}_\mathbf{i}, \frac{\mathbf{1}}{\lambda\mathbf{m_i}}\Big) \tag{5.8}$$

The CAR model displaces each individual region-effect estimates towards the local "mean effect" $\hat{W}_i$. Banerjee *et al.* (2003) note that the lattice model is computationally simpler compared to the geostatistical approach.

Given that mortality rates evolve through time (Wilkinson *et al.*, 2000) a further model could be proposed with both dependence in space and through time. It would require data relating to deaths by geographical region across time and is not being proposed here because of a lack of available data. For completion, such a model is referred to as a spatial-temporal model. If t_{ijk} denotes the time to death for the j^{th} subject in region i in the year k with $i = 1 \dots I$, $k = 1 \dots K$ and $j = 1 \dots n_{ik}$, $\mathbf{x_{ijk}}$ denotes the vector of covariates, W_{ik} the spatial-temporal frailties corresponding to the i^{th} region in the k^{th} year, then the proportional hazards model becomes:

$$h(t_{ijk}, \mathbf{x_{ijk}}) = \mathbf{h_0}(\mathbf{t_{ijk}})\mathbf{exp}(\beta^{\mathbf{T}}\mathbf{x_{ijk}} + \mathbf{W_{ik}}), \tag{5.9}$$

where h_0 is the baseline hazard. Assuming a lattice structure and a CAR(λ_k) model (see Besag *et al.*, 1991) the prior distribution is:

$$\lambda_k^{1/2} exp\Big[-\frac{\lambda_k}{2}\sum_{iadjj}(W_{ik}-W_{jk})^2\Big] \propto \lambda_k^{1/2} exp\Big[-\frac{\lambda_k}{2}\sum_{i=1}^{I} m_i W_{ik}(W_{ik}-\hat{W}_{jk})^2\Big], \tag{5.10}$$

where $iadjj$ denotes that region i and region j are adjacent to each other, and $\hat{W}_{ik}$ is the average of the frailties W_{jk} adjacent to region i for the k^{th} year and m_i represents the number of these adjacencies. The conditional distribution of W_{ik} becomes:

$$W_{ik}|W_{(j\neq i)k} \propto N\big(\hat{W}_{ik}, \frac{1}{\lambda_k m_i}\big). \tag{5.11}$$

5.2.3 Assessing Model Choice

When we fit each of these models we have to determine which model provides the best fit. We use an information measure, the Deviance Information Criterion (DIC), as we wish to ensure that when we add additional explanatory variables, spatial or socioeconomic that we are improving the model fit significantly. The DIC is an extension of the Akaike Information Criterion (AIC), is commonly used to compare the performance of different models (Spiegelhalter *et al.*, 2002). It is readily calculated using Markov Chain Monte Carlo (MCMC) methods (Banerjee and Carlin, 2003). The DIC is defined as:

$$DIC = \bar{D} + p_D, \tag{5.12}$$

with closeness of fit to the data measured by $\bar{D} = E_{\theta|y}[D]$ and the effective number of parameters measured by p_D. p_D is defined as,

$$p_D = E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) = \bar{D} - D(\bar{\theta}), \tag{5.13}$$

which is the deviance of the posterior mean subtracted from the posterior mean at the deviance. The deviance statistic is:

$$D(\theta) = -2 \ln f(y|\theta) + 2 \ln h(y), \tag{5.14}$$

where $f(y|\theta)$ is the likelihood, y the data vector, θ the parameter vector, and $h(y)$ a standardising function of the data alone. It does not have any impact on model selection. Small values of $\bar{D}$ represent a good fit and small values of p_D indicate a more parsimonious model. Smaller values of DICs are preferred. DICs are only used to compare models.

5.3 Data

The analysis in this paper is based on data from 2008 provided by the Northern Ireland Statistical Research Agency NISRA and is divided geographically using Super Output Areas (SOA) of which there are 890 in Northern Ireland. SOA's are a set of geographies developed in the UK after the 2001 census[5].

[5]In Northern Ireland SOA's typically contain 2000 lives but range from 1300 to 2800. They were designed to improve reporting of small area statistics and to ensure that each area was of similar size, unlike electoral wards which varied widely in size. The data is taken from the 2010 deprivation study.http:www.nisra.gov.uk/deprivation/nimdm_2010.htm

Table 5.1: Super Output Areas in Northern Ireland split by county

County	Number of Super Output Areas
Antrim	174
Armagh	69
Belfast	150
Derry	122
Down	288
Fermanagh	25
Tyrone	62

5.3.1 Geographical Classification for Northern Ireland data

The socioeconomic and demographic factors used in this study include measures of employment levels, education, crime and violence, income, proximity to services, environment and health deprivation. Separating Belfast from the counties the SOA's separate into the 6 counties and Belfast (See table 5.1). No SOAs were omitted in the analysis as all had sufficiently large populations ranging from 997 to $3,667$[6].

5.3.2 Deaths Data

Death data and mid population estimates are available for each of the SOA's in 2008 for males and females individually from the NISRA website[7]. Mortality rates were determined by dividing the total number of deaths in each SOA by the annual mid-year population estimates for each SOA.

The figures 5.1 and 5.2 show the normalised mortality rates for males and females in 2008 split by super output area. Higher mortality rates occur in the more urban areas of Northern Ireland, Belfast, North Down, Lisburn. Lower mortality rates occur in the more rural parts of Northern Ireland, for example Fermanagh and Tyrone. However, within urban areas the mortality can still be seen to experience some degree

[6] A table of the Super output areas along with their names and SOA codes is available on request

[7] see `http://www.nisra.gov.uk/`

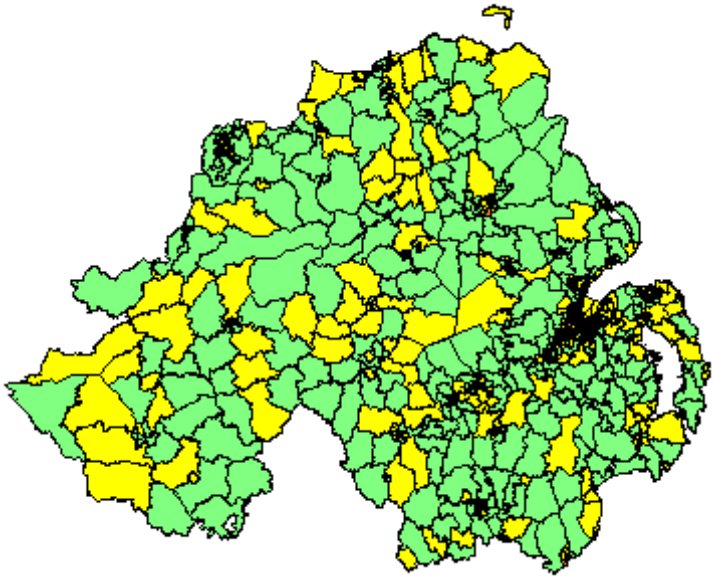

Figure 5.1: The normalised mortality rates for males in 2008 split by super output area

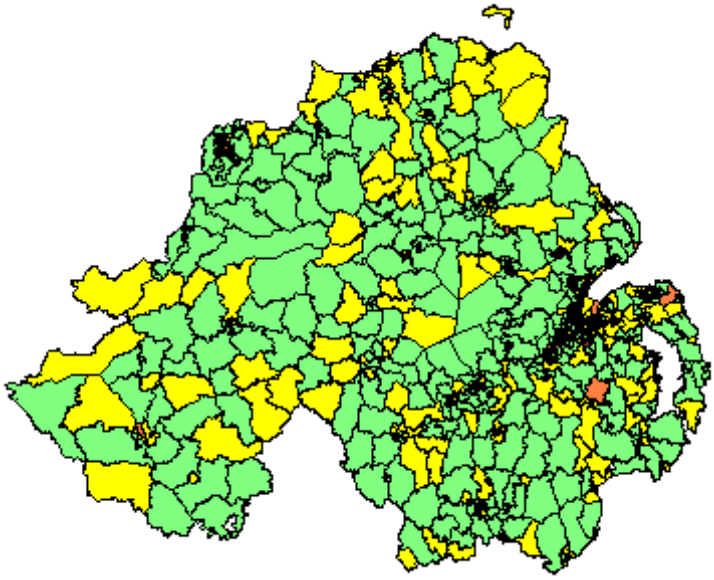

Figure 5.2: The normalised mortality rates for females in 2008 split by super output area

of variability, figure 5.3 shows the mortality within the belfast area. Similar maps of Northern Ireland for 1999 and 2003 show that the geographical variation in Northern Irish mortality rates changed little over the period 1999-2008.

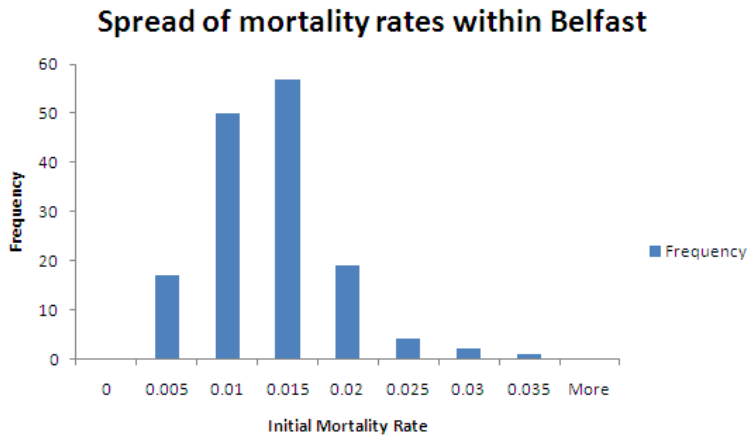

Figure 5.3: Histogram of the variation of the mortality across the Belfast region

5.3.3 Deprivation Data

The data series' obtained were selected to reflect the major factors expected to affect mortality. A detailed summary of the data upon which the deprivation factors were based can be found on the NIRSA website[8].

For interest, and specific to the Northern Ireland context, we define a measure of political polarisation which we calculate using the proportion of catholics in each region as the data item and converting this into a measure ranging from 0 to 1 reflecting the concentration of one particular political orientation over another. The focus here is on mortality rates in the two distinct types of area in Northern Ireland. "Polarised" areas are regions where there is a predominance of the population from one particular religious or political focus. The opposite to these sorts of areas are what are known

[8] http://www.nisra.gov.uk/deprivation/archive/Updateof2005Measures/NIMDM_2010_Indicator_Summary.pdf

as "mixed" areas where there is no particular concentration. We include this in the analysis to see if there is any explanatory power for mortality in this variable. We use the percentage of catholic believers in each SOA[9] to measure political polarisation. If we denote α_A as the proportion of catholics in region A then we calculate a factor to reflect the concentration of one political orientation in an area as:

$$2 * max(\alpha_A, 1 - \alpha_A) - 1. \tag{5.15}$$

The formulation of this factor is such that areas with large densities of catholics will measure equally with areas of the same density of non-catholics. The aim of the factor being to capture areas of high polarisation (either catholic or non-catholic). Table 5.2 summarizes the definitions of the covariates used for analysis.

Table 5.2: Deprivation covariates

Symbol	Covariate	Detail
Inc_i	Income	Index of income deprivation
Emp_i	Employment	Index of employment deprivation
Edu_i	Education	Index of education deprivation
Env_i	Environment	Index of environmental deprivation
Hea_i	Health	Index of health deprivation
$Prox_i$	Proximity to services	index of proximity to services deprivation
$Crime_i$	Crime	Index of crime deprivation
Age_i	Age	Proportion of the population over age 60
PP_i	Political polarisation	Based on the weighting of catholics / non-catholics in any region

We plot the socioeconomic standardised factors from the 2010 deprivation study in figure 5.4 to show the distribution of socioeconomic characteristics in Northern Ireland. Each covariate was standardized by subtracting its mean and then divided by its standard deviation. In Appendix A we show the same distributions but in a geographical fashion with thematic maps of the deprivation factors by super output area. In the plots, darker blue indicates greater deprivation and lighter blue indicates

[9] `http://www.ninis.nisra.gov.uk/mapxtreme/viewdata/Census/CensusKS07b.xls`

lower deprivation in the SOA. Appendix B shows the distribution of age profiles across the SOA's of Northern Ireland. We have divided this into younger populations (people under the age of 18), older populations (people over the age of 60/65) and the overall population. Finally, in Appendix C we have summarised the political polarisation of each SOA.

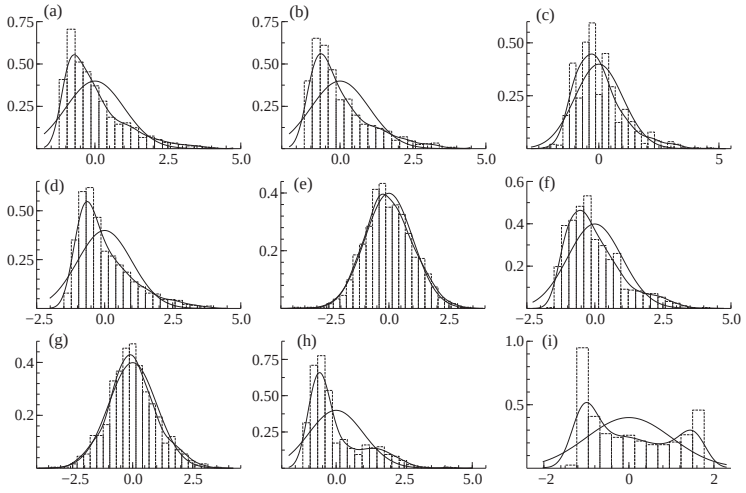

Figure 5.4: Distribution of deprivation measures (a) Crime, (b) Education, (c) Employment, (d) Environment, (e) Healthcare, (f) Income, (g) Age, (h) Proximity to services, and (i) Political polarisation .

These plots show how all covariates (deprivation factors and age and political orientation) are spatially correlated, with similar measures of demographic and economic characteristics between nearby SOAs. High levels of Employment, Income, Education and Health deprivation exist in the North Western areas of Northern Ireland (Fermanagh and Derry) and in the South Down and Armagh areas of Northern Ireland. Deprivation of the living environment is high in the North Antrim and Fermanagh areas (these areas are significantly rural and have poor transportation links) and is also high in some inner city areas of Belfast and Lisburn and Newtownards. Crime and Disorder is more significant in the cities but also shows high levels in some pockets of rural areas such as Tyrone, Armagh and Antrim. One significant factor to note for

Northern Ireland is that nearly all areas (all areas except the Belfast, Lisburn, Newtownards, Bangor quadrangle) suffer from a lack of facilities.

5.4 Empirical Analysis

The data, split by super output area, is aggregate death data, d_i, and exposure data, e_i, for each super output area $i = 1, \ldots, 890$ over the year 2008. The variation in this data will be heavily masked by the fact that the age profile of each super output area[10] will be different. Since we are investigating the impact of socio-economic measures on the variation in mortality rates we first need to standardise the data for age. We do this by running a simple regression model on the natural logarithm of the mortality rate with only constant and age as the covariates.

$$\ln \frac{d_i}{e_i} = \alpha + \beta * Age_i + \epsilon_i \tag{5.16}$$

The results of this model for males and females are given in table 5.3. From the results we can see that the constant explains a significant amount of the log mortality rates but this is to be expected since, whilst there is variation in the rates they all hover around a similar level. Having identified the constant and age parameters we recalculate log transformed age-standardised mortality rates using the residuals specific to each area. The remainder of the analysis will focus on the age standardised mortality rates.

[10] Age profile data was taken from `http://www.ninis.nisra.gov.uk/mapxtreme/viewdata/Census/CensusKS02.xls`

Table 5.3: Results of Simple GLM for Males 2008 and Females 2008 to control for age and constant

	Coefficient	Std.Error	t-value	t-prob	Part R^2
Males					
Intercept	-4.9342	0.0191	-258	0	0.9868
age	0.3387	0.0191	17.7	0	0.2609
Females					
Intercept	-5.0132	0.0237	-211	0	0.9805
age	0.4463	0.0237	18.8	0	0.285

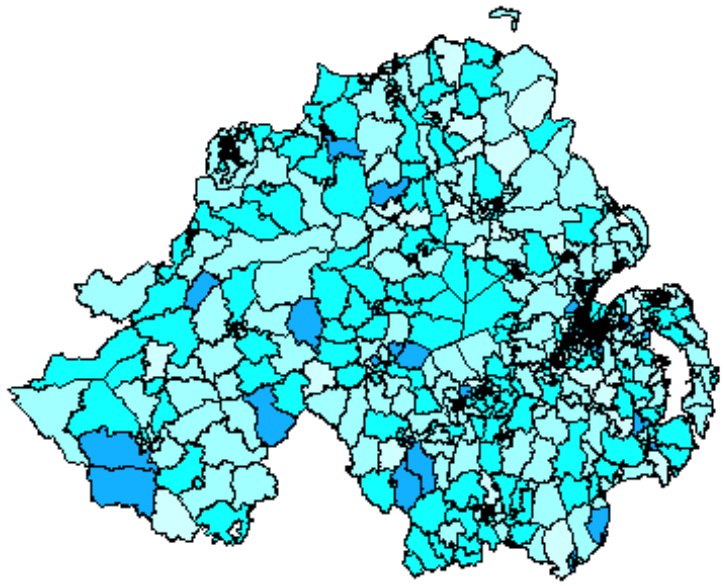

Figure 5.5: Male age standardised mortality rates

5.4.1 Simple Regression Model

In this subsection we show the results of fitting a simple linear model to the age standardised mortality rates derived from the data in Northern Ireland. We then demonstrate how this model can be extended to allow for spatial frailties.

The modelling uses a hierarchical Bayes method. A prior distribution is assumed for each of the parameters which we then combine with the likelihood of the data given the parameters to give us the posterior distribution for the parameters given the data. Parameters are estimated using Markov Chain Monte Carlo methods. The Condition-

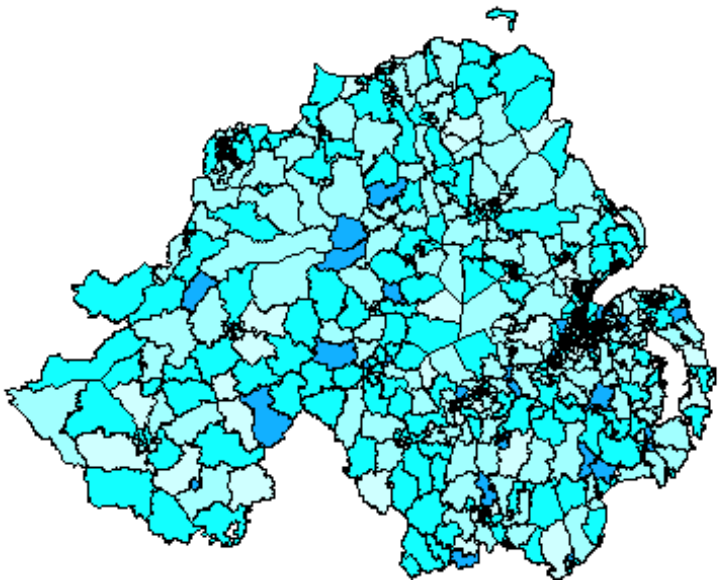

Figure 5.6: Female age standardised mortality rates

ally Autoregressive (CAR) model for spatial variation is used.

The generalized linear model of mortality rate m_i for region i with a logit link function has the following likelihood:

$$L(m_i;\beta,\mathbf{x}) \propto \prod_{i=1}^{I}\Big(\frac{exp(\sum_{j=1}^{n}\beta_j x_{ij})}{1+exp(\sum_{j=1}^{n}\beta_j x_{ij})}\Big), \tag{5.17}$$

where n is the number of covariates (socioeconomic or demographic factors), x_i is the vector of covariates respectively for each super output area i, $i = 1 \ldots I$. The posterior distribution is:

$$p(\beta|\mathbf{x}) \propto L(m_i;\beta,\mathbf{x})p(\beta), \tag{5.18}$$

where the first term on the right represents the logistic likelihood, and the second is the prior distribution for the parameters. A vague uniform prior distribution is assumed with small mean and large variance because of a lack of prior knowledge about the

parameters (Banerjee *et al.*, 2003). This prior for β is used in all the models.

The results of fitting the standardised mortality rates to a generalised linear logistic regression model using all the standardised socioeconomic factors for Males 2008 and Females 2008 are shown in table 5.4. We provide the parameter estimates and p-values of the covariates for the 2008 data. However, we also find Employment, Healthcare and Environment deprivation are significant. Proximity to local services was significant at the 10% level for females but not for the male data. The political polarisation (PP) variable does not show any significance on its own but proves to be significant through income, education and employment for males and through employment and income for females. Improved levels of Employment, Income, Education and Proximity to services lead to lower mortality rates.

To demonstrate the confidence we have for the parameter estimates for each covariate table 5.5 provide the 2.5%, 50% and 97.5% posterior percentiles for each of the predictors and interaction terms for the generalised linear model for both Males and Females.

The quality of the fit using this model is measured using the DIC measure which pits quality of fit against parsimony. Using this measure we have a fit given in table 5.6

This measure of fit quality is of little use on its own. One of the hypotheses in this paper is that in the Northern Ireland context, introducing spatial frailties does not improve the fit. Before addressing this we simplify the existing model. To simplify the model we carry out a general to specific modelling analysis for both the Male and Female data to eliminate those variables that are not significant in explaining the variation we see in the age standardised mortality rates. We then extend the resulting models adding spatial and non-spatial frailties to explain any residuals.

Table 5.4: Results of Simple linear regression for Males 2008 and Females 2008

MALE	Coefficient	Std.Error	t-value	t-prob	Part R^2
Constant	-4.9879	0.0286	-174.000	0.0000	0.9720
Crime	-0.0474	0.0377	-1.260	0.2090	0.0018
Education	-0.0937	0.0575	-1.630	0.1034	0.0030
Employment	-0.1008	0.0803	-1.250	0.2099	0.0018
Environment	0.1040	0.0358	2.910	0.0038	0.0095
Healthcare	0.3765	0.0649	5.800	0.0000	0.0369
Income	-0.0412	0.1045	-0.394	0.6937	0.0002
Proximity	0.0173	0.0288	0.601	0.5481	0.0004
PP	-0.0050	0.0312	-0.162	0.8711	0.0000
PP*Inc	-0.2305	0.0830	-2.780	0.0056	0.0087
PP*Edu	0.0507	0.0508	0.998	0.3186	0.0011
PP*Emp	0.1660	0.0671	2.480	0.0135	0.0069
FEMALE	Coefficient	Std.Error	t-value	t-prob	Part R^2
Constant	-4.9284	0.0219	-225.000	0.0000	0.9829
Crime	-0.0008	0.0289	-0.027	0.9788	0.0000
Education	-0.0286	0.0440	-0.650	0.5157	0.0005
Employment	-0.1381	0.0615	-2.240	0.0251	0.0057
Environment	0.04781	0.0274	1.740	0.0814	0.0035
Healthcare	0.3369	0.0497	6.770	0.0000	0.0497
Income	0.0466	0.0801	0.582	0.5608	0.0004
Proximity	0.0287	0.0221	1.300	0.1944	0.0019
PP	-0.0249	0.0239	-1.040	0.2975	0.0012
PP*Inc	-0.0878	0.0636	-1.380	0.1674	0.0022
PP*Edu	0.0018	0.0389	0.045	0.9641	0.0000
PP*Emp	0.0790	0.0514	1.540	0.1243	0.0027

5.4.2 General to specific modelling of covariates in Northern Ireland

In the General to Specific method the specification of the general model from which reductions are made is crucial (Hendry, 2000 page 482) because a poorly specified general model stands little chance of leading to a good final specific model. With a cross sectional study we evaluated the congruency of the general model by applying several mis-specification tests. These same tests are also applied at every stage of the reduction process. We follow the approach of Krolzig and Hendry (2001) and carry out

Table 5.5: Posterior percentiles for covariates and interaction terms for Male and Female data using the GLM model

	Males			Females		
Coefficient	2.50%	median	97.50%	2.50%	median	97.50%
Intercept	-4.8680	-4.8440	-4.8190	-4.860	-4.834	-4.807
Employment	-0.2475	-0.1823	-0.0913	-0.146	-0.065	0.015
Income	-0.0569	0.0244	0.1107	-0.130	-0.033	0.067
Healthcare	0.2488	0.3190	0.3736	0.193	0.275	0.336
Education	-0.0966	-0.0428	0.0109	-0.042	0.030	0.079
Environment	-0.0138	0.0231	0.0595	-0.005	0.032	0.065
Crime	-0.0611	-0.0161	0.0221	-0.055	-0.010	0.027
Proximity	-0.0391	-0.0111	0.0182	-0.022	0.008	0.036
PP	-0.0351	-0.0092	0.0183	-0.050	-0.023	0.005
PP*Emp	0.0495	0.1266	0.1904	-0.089	-0.012	0.059
PP*Inc	-0.1868	-0.1090	-0.0186	0.002	0.086	0.182
PP*Edu	-0.0265	0.0306	0.0836	-0.088	-0.038	0.025

Table 5.6: Goodness of fit for Males 2008 and Females 2008 generalised linear model

	$\bar{D}$	p_D	DIC
Males			
GLM	5266.00	12.43	5278.43
Females			
GLM	4493.71	12.83	4506.54

the following tests on the general model: (1) Two F-tests for parameter constancy for breakpoints at the sample mid-point and 90th percentile; and (2) Doornik and Hansens (1994) χ^2 test for normality of the error terms.

The figures 5.7 and 5.8below show scatter plots of the age standardised mortality rates against the remaining depravation measures. They show that there may still be some non-normal variation across the measures.

Carrying out the GETS analysis with the age standardised mortality variation rates across Northern Ireland we find, in the case of male data, that the variables that remain in the specific model are environment, healthcare, and PP through income and employment. For the females we are left with employment, healthcare and proximity.

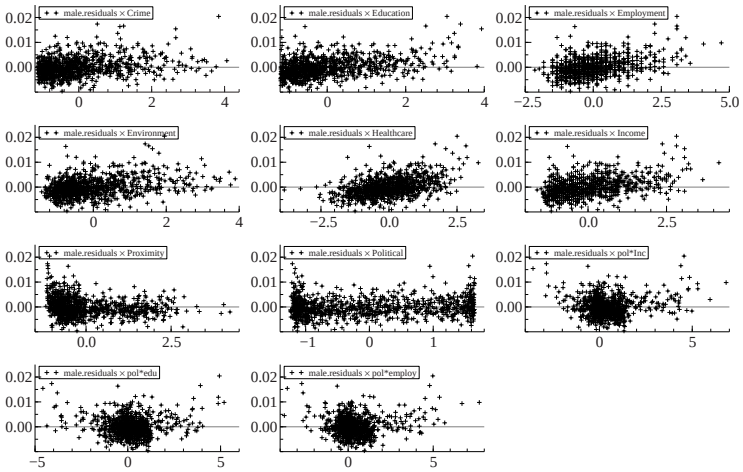

Figure 5.7: Scatter plots of the male age standardised mortality rates against deprivation measures (a) Crime, (b) Education, (c) Employment, (d) Environment, (e) Healthcare, (f) Income, (g) Proximity to services, and (h) Political polarisation (i) PP * Employment, (j) PP * Income and (k) PP * Education.

We note from this analysis that PP does explain some of the variation we see in male data but not in female data. Potential explanations for this could relate to the specific problems that the Northern Ireland region has had over the past number of years and the greater impact that this would have had on the male population. In this study we do not attempt to explain but rather just identify the significant covariates. We conclude from the GETS analysis that the covariates to use in our simplified model for the standardised males mortality rates are: Healthcare, Environment and the interaction terms PP by Income and PP by employment. For the females we will use Healthcare, Employment and Environment.

The next step is to fit our linear models to the data but only allowing for the specific factors identified by the GETS analysis. The results of the fitting process using only the specified covariates are given in table 5.7 having identified the simpler model using the general to specific modelling approach we now look at the issue of frailties and consider several adaptations to the linear regression model encompassing independent

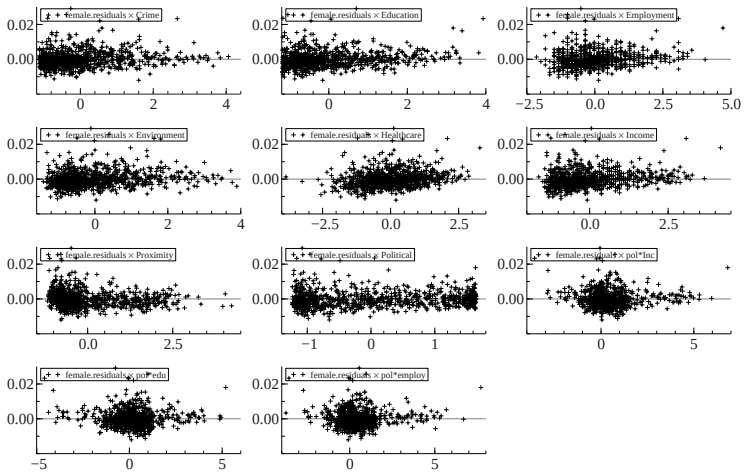

Figure 5.8: Scatter plots of the female age standardised mortality rates against deprivation measures (a) Crime, (b) Education, (c) Employment, (d) Environment, (e) Healthcare, (f) Income, (g) Proximity to services, and (h) Political polarisation (i) PP * Employment, (j) PP * Income and (k) PP * Education.

and spatial frailties.

Table 5.7: Results of Simple linear regression for Males 2008 and Females 2008

MALE	Coefficient	Std.Error	t-value	t-prob	Part R^2
Intercept	-4.9918	0.0243	-205.000	0.0000	0.9795
Health	0.2672	0.0395	6.770	0.0000	0.0493
Environment	0.0783	0.0300	2.610	0.0093	0.0076
PP*Emp	0.1332	0.0639	2.080	0.0375	0.0049
PP*Inc	-0.1739	0.0637	-2.730	0.0065	0.0084
FEMALE	Coefficient	Std.Error	t-value	t-prob	Part R^2
Intercept	-4.9342	0.0173	-285.000	0.0000	0.9892
Employment	-0.1197	0.0430	-2.780	0.0055	0.0087
Health	0.3208	0.0441	7.270	0.0000	0.0563
Environment	0.0435	0.0210	2.070	0.0388	0.0048

The confidence we have for the parameter estimates for each covariate table 5.8 provide the 2.5%, 50% and 97.5% posterior percentiles for each of the predictors and interaction terms for the specific linear model for both Males and Females.

Table 5.8: Posterior percentiles for covariates and interaction terms for Male and Female data using the GLM model

Males			
Coefficient	2.50%	median	97.50%
Intercept	-4.852	-4.830	-4.807
Health	0.114	0.144	0.175
Environment	-0.018	0.009	0.038
PP*Emp	0.024	0.098	0.163
PP*Inc	-0.148	-0.080	-0.006
Females			
Coefficient	2.50%	median	97.50%
Intercept	-4.847	-4.822	-4.797
Employment	-0.112	-0.053	0.003
Health	0.196	0.255	0.319
Environment	0.000	0.030	0.056

5.4.3 Introducing frailties

In the previous section we have identified the main socioeconomic factors that explain the variability of mortality rates across Northern Ireland. In this next section we extend this standard regression model to include a frailty aspect picking up the geographical location of each region. We then look at the fitting quality again using a DIC measure. When we add the frailties the likelihood function becomes:

$$L(m_i;\beta,\mathbf{x}) \propto \prod_{i=1}^{I}\Big(\frac{exp(\sum_{j=1}^{n}\beta_j x_{ij}+W_i)}{1+exp(\sum_{j=1}^{n}\beta_j x_{ij}+W_i)}\Big), \tag{5.19}$$

where W_i is the frailty for super output area i, which captures any remaining effects not explained by the covariates. Under this non-spatial frailties setting, the frailties are assumed to be identical and independently distributed with the following distribution:

$$W_i \sim N(0,\sigma^2). \tag{5.20}$$

Eq. (5.20) assumes no spatial dependence since frailties in one SOA are indepen-

dent of frailties in another. The hierarchical Bayes model is:

$$p(\beta, \mathbf{W}, \sigma^2|\mathbf{x}) \propto L(m_i; \beta, \mathbf{x})p(\mathbf{W}|\sigma^2)p(\beta)p(\sigma^2), \tag{5.21}$$

where the likelihood is given by Eq. (5.19). As in Banerjee *et al.* (2003), a Gamma (0.001, 0.001) prior distribution is used for $\tau = 1/\sigma^2$ with mean 1 and variance 1000. A flat Uniform prior was adopted for β.

To allow for spatial clustering (i.e., adjacent super output areas showing similar mortality characteristics), we allow for spatial correlations between nearby SOAs through a Conditional Auto-regression specification. In this specification an adjacency matrix is defined to capture the geographical variations in mortality. The adjacency matrix is defined by assigning the ij^{th} entry a value of 1 if the super output area i is adjacent to super output area j and 0 otherwise. The hierarchical Bayes model becomes:

$$p(\beta, \mathbf{W}, \lambda|\mathbf{x}) \propto L(m_i; \beta, \mathbf{x})p(W|\lambda)p(\beta)p(\lambda), \tag{5.22}$$

where the prior $W|\lambda$ is given by:

$$\lambda_k^{1/2} exp\Big[-\frac{\lambda_k}{2}\sum_{iadjj}(W_{ik} - W_{jk})^2\Big] \propto \lambda_k^{1/2} exp\Big[-\frac{\lambda_k}{2}\sum_{i=1}^{I} m_i W_{ik}(W_{ik} - \hat{W}_{jk})^2\Big], \tag{5.23}$$

where $iadjj$ denotes that Super output area i and Super output area j are adjacent to each other, $\hat{W}_i$ is the average of the frailties W_j , adjacent to SOA i and m_i represents the number of these adjacent regions (Bernardinelli and Montomoli, 1992). A

consequence of the above prior is that:

$$W_i|W_{j\neq i} \propto N\big(\hat{W}_i, \frac{1}{\lambda m_i}\big) \tag{5.24}$$

Repeating the fitting analysis now using the independent frailties model and the spatial clustering model the parameter estimates along with confidence intervals are shown in tables 5.9 and 5.10.

Table 5.9: Posterior percentiles for covariates for Male and Females 2008 data with the independent frailties model

Males			
Coefficient	2.50%	median	97.50%
Intercept	-4.936	-4.897	-4.821
Health	0.103	0.143	0.188
Environment	-0.019	0.019	0.060
PP*Emp	0.001	0.104	0.193
PP*Inc	-0.186	-0.095	0.009
Female			
Coefficient	2.50%	median	97.50%
Intercept	-4.864	-4.835	-4.803
Employment	-0.128	-0.059	0.000
Health	0.195	0.259	0.334
Environment	0.000	0.031	0.062

Looking at the deviance measure now for the independent frailties and the spatial clustering models alongside the standard generalised model we have the results shown in table 5.11. From this we can see that there is a minor benefit to be had by allowing for non-spatial frailties however, after adding spatial frailties we do not improve the model fit significantly. This suggested that our proposition that mortality rates may have some form of spatial clustering is clearly not the case in Northern Ireland.

We test the residuals[11] for normality for each of the three models above and note

[11]We follow the approach of Dowd *et al.* (2010) and test the age standardised mortality residuals for males and females using the three models proposed. The tests used aim to identify whether the mortality

Table 5.10: Posterior percentiles for covariates for Male and Females 2008 data with the spatial frailties model

Males			
Coefficient	2.50%	median	97.50%
Intercept	-4.937	-4.896	-4.842
Health	0.132	0.196	0.265
Environment	-0.090	-0.032	0.025
PP*Emp	-0.062	0.038	0.147
PP*Inc	-0.153	-0.043	0.055
Female			
Coefficient	2.50%	median	97.50%
Intercept	-4.859	-4.830	-4.800
Employment	-0.116	-0.045	0.025
Health	0.190	0.253	0.331
Environment	-0.026	0.013	0.047

Table 5.11: Goodness of fit for Males 2008 and Females 2008 generalised linear model with spatial and non-spatial frailties

	$\bar{D}$	p_D	DIC
Males			
GLM	5303.21	5.056	5308.27
IID	4378.52	460.203	4838.72
SPA	5590.56	1641.24	7231.8
Females			
GLM	4501.95	4.248	4506.2
IID	4326.85	150.187	4477.04
SPA	5492.28	1185.74	6678.02

that the results are in line with those of Dowd et al (2011) who carried out similar tests on a range of stochastic mortality models.

residuals are consistent with i.i.d. N(0,1) as assumed under the null hypothesis. The tests involves three types of tests: (1) t-test of mean prediction; (2) Variance ratio (VR) test (see Cochrane, 1988; and Lo and MacKinley, 1988 and 1989); (3) Jarque and Bera test of Normality based on the skewness and kurtosis predictions. A statistically significant result for any of these tests - which we take to be any test which produces a p-value of less than 1% - indicates inconsistency with i.i.d. N(0,1).The results are available on request.

5.4.4 Can geographical information replace socioeconomic data

We have demonstrated that spatial frailties do not add any fitting value when we have sufficient socioeconomic data suggesting that spatial clustering of mortality rates may not be present in the Northern Ireland region. However, we might expect that spatial models will be more useful when we have little other data upon which to fit our mortality rates. To test this we consider the situation where we do not have significant amounts of socioeconomic data available to us. We assume the only covariate we have is an aggregate age parameter for each SOA and we fit a simple regression model and some extensions of it to the data for males and females in 2008 to the non-age standardised mortality rates. We chose this covariate to assess the impact of adding spatial structure as it is the most readily available piece of information we might have and the one which is likely to have the most significance when trying to explain raw mortality rates. Three models were fitted to the data, firstly a simple generalised linear model (GLM) with no allowance for spatial dependence (the no frailties model), secondly a linear model allowing for independent, identically distributed frailties (non-spatial frailties model) and finally a linear spatial frailties model which allows for spatial dependence which employs a CAR specification to capture the spatial dependence between adjacent SOAs.

The models are again fitted using a Markov Chain Monte Carlo algorithm under a basins hierarchical framework. A flat Uniform (-10000,10000) prior was adopted for the parameter and a Gamma (0.001,0.001) was chosen for the frailties W_i in the non-spatial frailties model. For the spatial frailties model the smoothness parameter λ of the CAR specification was given a Gamma(0.001,0.001) prior distribution. Table 5.12 compares the three models using the Deviance Information Criterion (DIC), which is the sum of the expected deviance and the number of effective parameters p_D.

Comparisons of DIC values show that the model with i.i.d. non-spatial frailties

Table 5.12: Goodness of fit for Males 2008 and Females 2008 spatial and non-spatial modelling

	Males		Females	
Model	p_D	DIC	p_D	DIC
Non-Spatial No frailties Model	1.94	5340.09	2.09	4772.49
Non-Spatial Frailties Model	452.19	4989.52	138.04	4743.69
Spatial Frailties Model	1698.49	7358.10	1259.65	6675.01

shows a mild improvement over the no-frailties model, despite the increase in the number of effective parameters. However, the inclusion of spatial frailties does not improve the fit further. In fact the DIC for the spatial frailties is poorer due to the additional number of parameters.

5.5 Conclusions

This paper has considered geographical variation of mortality and the effect of socioeconomic explanatory factors using Northern Ireland data. Raw, logarithmically transformed mortality rates where firstly standardised for age distribution differences across the super output areas of Northern Ireland to generate a series of age-standardised mortality rates for males and females. Regression models were then used to estimate the impact of covariates extracted from the Northern Ireland Statistical Research Agency (NISRA) on the age standardised rates. A General to specific analysis was carried out to reduce and simplify the models which identified some differences between the males and females in our study, in terms of the covariates that proved to be significant in the final specific model. In the case of males we see that age standardised mortality rates can best be explained by the variation in deprivation measures; healthcare, environment and political polarisation (through income and education). This would suggest that the combined effect of living in a densely populated catholic (or protes-

tant area) with poor education and income levels has an impact on mortality rates for males. In the case of females we see that employment, healthcare and environment are important. In both cases of males and females the quality of the immediate environment (heating, double glazing etc.) has come through as being a significant factor in explaining mortality rate variation.

To assess the ability of geographical location to explain mortality rates and the proposition that age standardised mortality rates may show some geographical clustering, models including independent frailties and spatial frailties were fitted to the data. When we allowed for independent frailties we found some improvement in the fit measured on a deviance information criteria despite the additional parameterisation. However when we fit a spatial structure to the data the information criteria was not improved. We conclude that there is little evidence of spatial clustering in mortality rates in Northern Ireland.

Finally we considered the case of frailties when we have little or no socio-economic data available. In this case we seen again that with the inclusion of independent frailties the fit can be improved using a DIC measure and in this case the improvement is more dramatic which is to be expected since there is a larger unexplained element to the mortality rates. This supports the view that there is a place for geographical location data when we have limited other socio-economic data.

We carried out residual tests for normality on each of the frailty and non-frailty models for the males and females using an approach of Dowd et al (2011) and find that the models perform adequately when compared to the results found in that paper.

The results of this paper show that there is a place in mortality modelling for allowing for underlying socioeconomic characteristics in the modelling process. Particularly when forecasting mortality it would appear important that we forecast the underlying socioeconomic profile and through this identify the direction in which mortality rates

are moving. The inability of frailties to improve the fitting quality could be due to some correlation between the socioeconomic variables included in this study and the frailties themselves. We acknowledge this and suggest that this would be a possible route for further investigation.

5.6 Appendix: Additional Figures and Tables

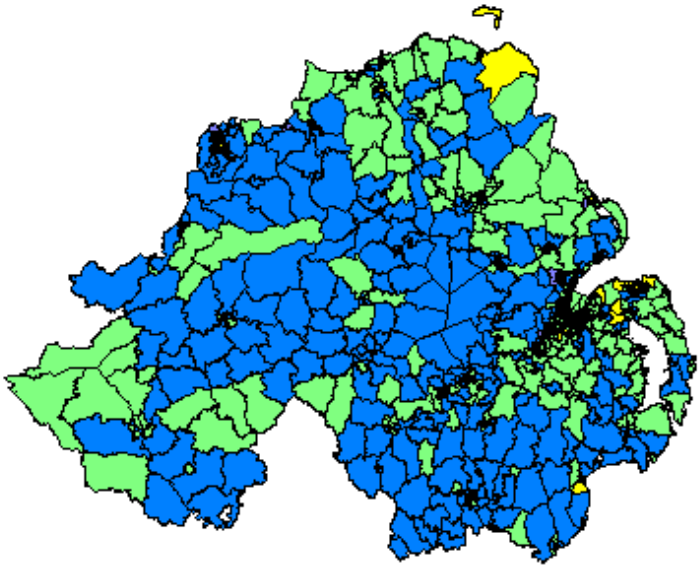

Figure 5.9: Spatial distribution of age across Northern Ireland SOA's

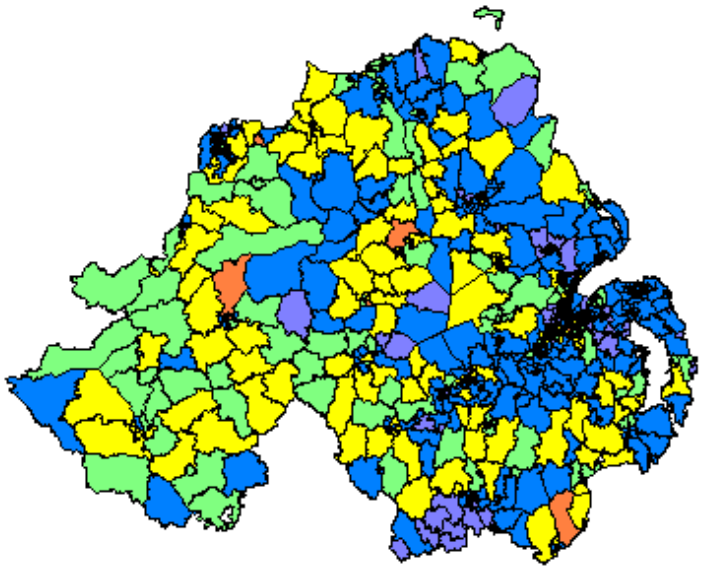

Figure 5.10: Spatial distribution of political density across Northern Ireland SOA's

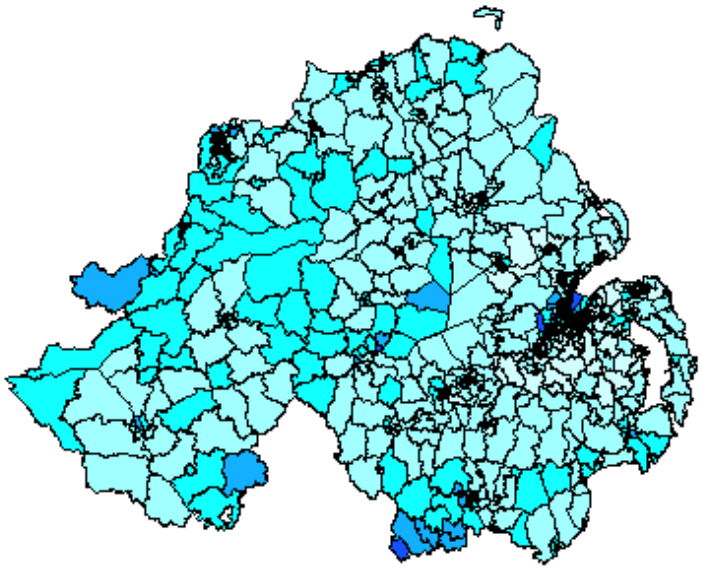

Figure 5.11: Income deprivation across Northern Ireland SOA's

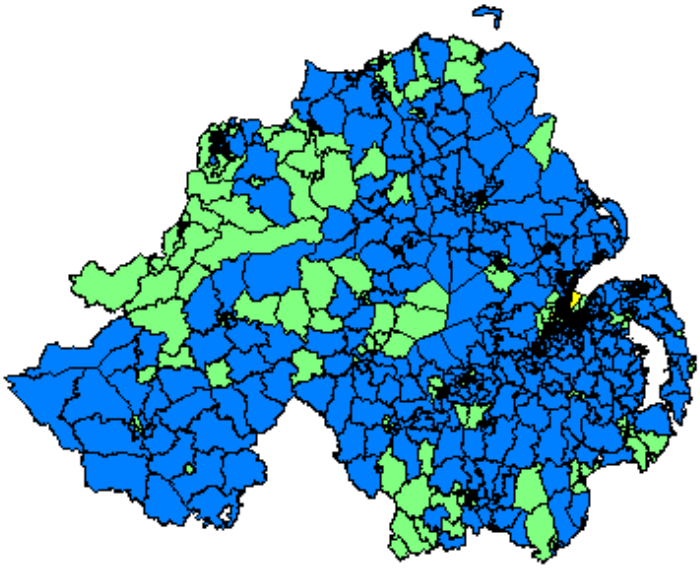

Figure 5.12: Employment deprivation across Northern Ireland SOA's

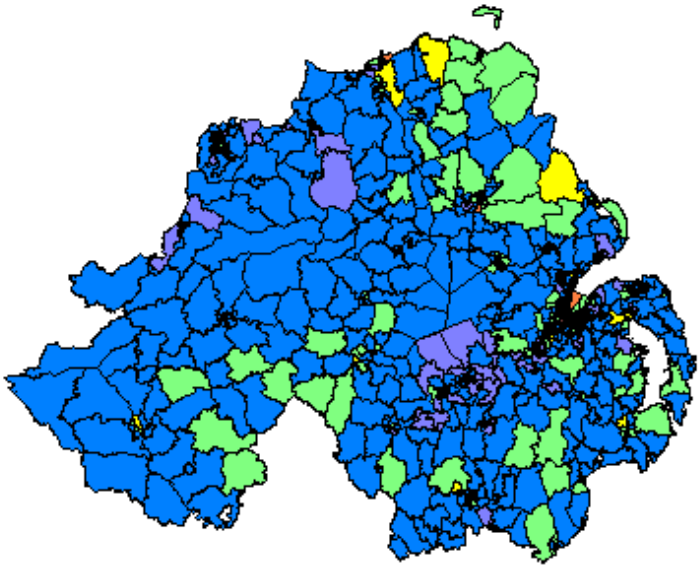

Figure 5.13: Environment deprivation across Northern Ireland SOA's

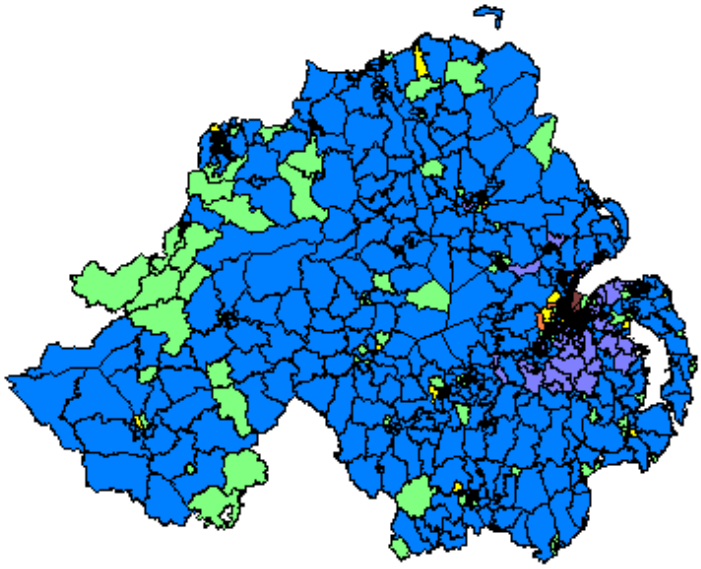

Figure 5.14: Education deprivation across Northern Ireland SOA's

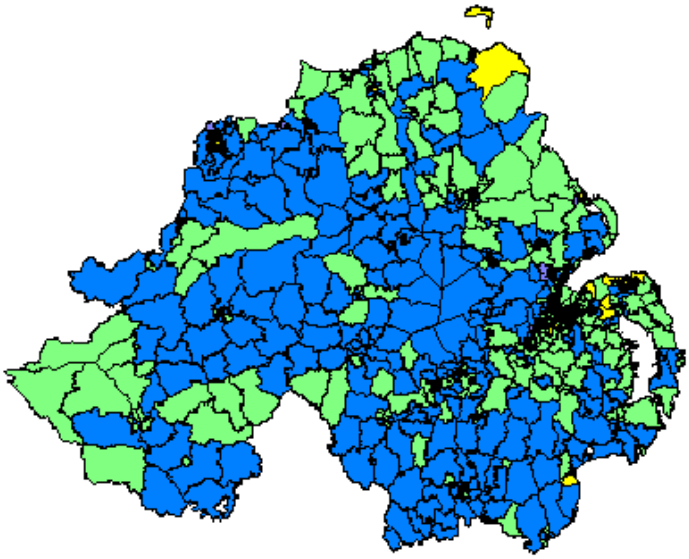

Figure 5.15: Crime deprivation across Northern Ireland SOA's

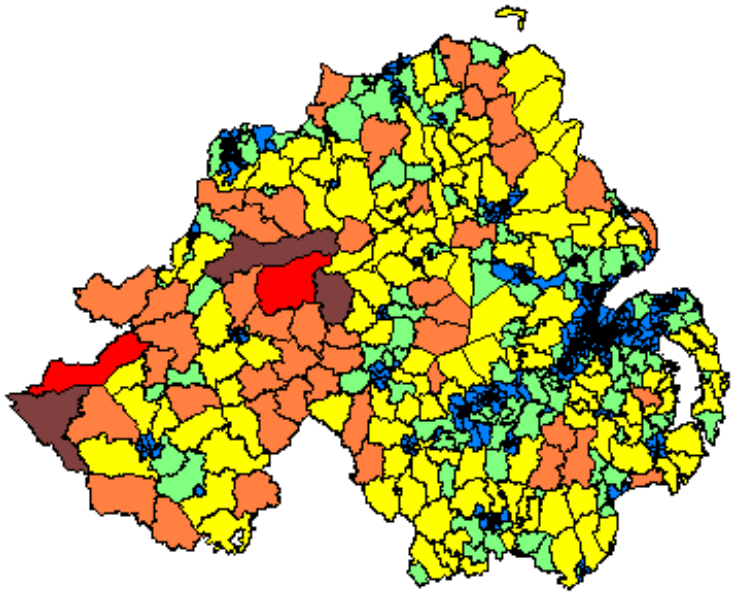

Figure 5.16: Proximity deprivation across Northern Ireland SOA's

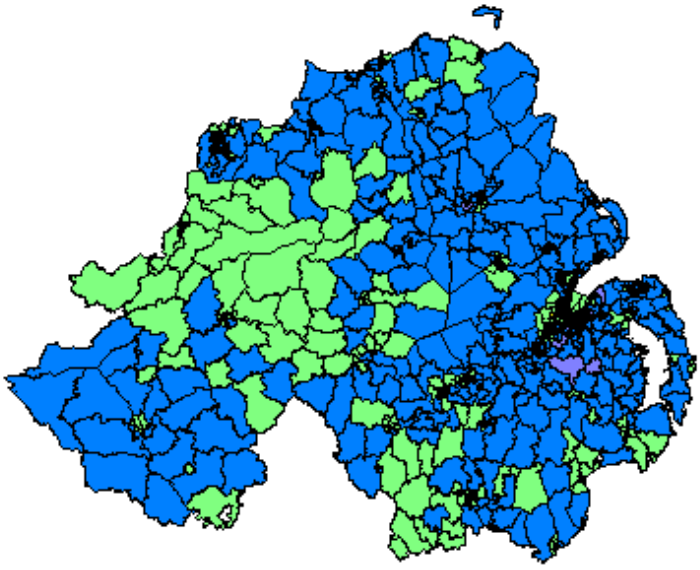

Figure 5.17: Proximity deprivation across Northern Ireland SOA's

Chapter 6

Forecasting death rates using exogenous variables

6.1 Introduction

Stochastic mortality models exploit patterns of common variation in deaths data across ages over time. We argue in this paper that taking account of trends in the factors explaining mortality decline such as income, health expenditure and lifestyle leads to improved forecasts. Lee and Carter (1992) model provided the seminal approach using a one factor time series approach. Subsequent innovations include modelling the cohort effect (Renshaw & Haberman, 2003, 2006; Currie, 2006), adding a second period effect (Cairns et al., 2006 hereafter CBD), widening of the model to fit ages 20-89 (Plat, 2009). Girosi and King (2008) use Bayesian methods to smooth over time, age and country and this approach is extended further by King and Soneji (2011) to incorporate lagged exogenous variables in a Bayesian hierarchical model of mortality rates. These approaches, which make use of the regularities found in the age and time profile of mortality data have been the most successful methods to date but fail to ex-

plain the drivers of mortality improvements and assume that trends seen in the past will be continued into the future (Booth and Tickle, 2008). The work presented in this paper can be divided into three parts. In the first part we use a principal components approach to identify the factor structure of the mortality data for the U.S., U.K. and Japan. The second stage takes the latent factors and explains these factors by observed, exogenous factors (GDP, health expenditure, smoking levels, alcohol consumption etc) using appropriate statistical techniques and using stopping rules to prevent the model become over-parameterised. Finally, having identified the most appropriate exogenous determinants we forecast the exogenous variables using ARIMA techniques and build the forecasted exogenous variables into a model using the King and Soneji (2011) approach. Whereas King and Soneji (2011) used lagged exogenous variables to explain mortality and so avoid the need to forecast these variables, we will forecast the exogenous variables separately. There are many possible explanations for recent changes in mortality rates. The health production function approach where health is proxied by mortality provides a framework for understanding the determinants of mortality. Auster et al. (1969) used the following health production model:

$$M_i = c_i + \alpha Z_i + \beta X_i + \gamma HC_i + \delta E_i + u_i \tag{6.1}$$

where M_i are logged (standardised) mortality rates by US state, Z_i socio-economic status (income, education), X_i lifestyle inputs (alcohol, tobacco), HC_i are healthcare inputs (drugs, doctors, hospital capital stock), E_i captures environmental variables (urbanization, industrialization) and u_i is a random element. Higher incomes allow people to spend more on health inputs. As average incomes rise, people can purchase more non-healthcare inputs that benefit health such as better housing, more nutritious

food and gym membership. Where healthcare coverage must be privately paid for (US and partly in Japan), higher incomes also allow people to spend more on better doctors and better hospital care.[1] The choices that individuals make in relation to their health also affect mortality. Lifestyle factors such as smoking (Leon, 2011; Thornton et al. 2002), obesity (Cutler et al., 2009) and alcohol consumption (Miller and Frech, 2000) are all recognised as significant risk factors. In studying secular trends in mortality, the role of advances in medical technology must be also considered. Cutler and Meara (2004) attributed much of the decline in US adult mortality in the second half of the twentieth century to cardiovascular disease treatment (new drugs, new surgical procedures and specialised equipment). Other factors considered are economic instability (Bethune, 1997; Iverson et al, 1998), environmental air pollution (Schwarz and Dockery, 1992), pharmaceutical expenditure (Miller and Frech, 2000) and crime (Thornton et al., 2002). The remainder of this paper is laid out as follows. In section 2 we discuss the data that has been used in this study. The methodology is discussed in section 3. The results are presented and analysed in section 4. Section 5 concludes.

6.2 Data

Data on possible determinants of health were taken from OECD Health data 2009. Data availability dictated the candidate variables chosen: Alcohol consumption (for those aged 15+), Tobacco consumption (15+), Total fat intake, Fruit and vegetable consumption, Gross domestic product per capita (in 2000 prices) and total expenditure on health per capita (in 2000 prices). Definitions and descriptive statistics are given in Table 1. The health expenditure time series begins in 1970 while Japanese tobacco

[1] Although higher incomes also permit increased consumption of goods injurious to health such as alcohol and tobacco. In addition, Ruhm (2004) argues that there are less motor vehicle accidents and people adopt healthier lifestyles in economic downturns.

consumption is only available from 1968 onwards. Therefore, data over the period 1970-2006 were only considered. GDP and Health expenditure have been both logged in the statistical analysis. Other variables were excluded. The obesity time series are short and patchy and to some extent this information is captured by the food measures included. Data on pharmaceutical expenditure and medical technology capital stock (CT and PET scanners, MRI units, radiation therapy etc) are insufficient and are captured crudely by aggregate health expenditure. Air quality emissions data (SOx,NOx and CO) are inadequate.

Not all determinants of mortality are contemporaneous. Barker (1992) provided evidence that insults to foetal health had life-long consequences based on an analysis of the risk factors for cardiovascular diseases found in adults who were born at the time of the WW2 Dutch famine. The short time series data considered in our study precludes the inclusion of variables of large lag length. Several authors indicate that these effects may be relatively minor. Murphy (2010) argues that exposure to a health shock has two opposing consequences: selection (excess mortality in the relevant period perhaps leading to the survival of a more robust cohort than average) and scarring (a weakened cohort more susceptible to illness going forward) and that the resultant effect is ambiguous. In a study of twins, Herskind et al. (1996) found no evidence that family environment had an impact on longevity whereas current environmental influences were influential. Similarly, Cutler et al. (2006) indicate cardiovascular risk factors experienced in adulthood are much more significant on mortality than early life exposures. The mortality data used is taken from the Human Mortality Database collated by the Department of Demography at the University of California, Berkeley and the Max Planck Institute for Demographic Research in Rostock, Germany. Death rates, a ratio of the death count by single age and year divided by an estimate of the exposure-to-risk in the same interval, for males in the US, UK and Japan over the pe-

riod 1970-2006 were selected. Models were estimated over the period 1970-2000 and mortality rates for the remaining 6 years, 2001-2006, were retained for comparison with forecasts. These countries were chosen so as to provide a variety of results. On the one hand, being non-tropical countries with developed health care systems similar factors should determine trends in mortality. On the other hand, they are distinct in terms of culture, diet, and the importance of private versus public provision in health care which should generate distinct results. Due to the exponential nature of mortality rates we model the logarithmically transformed mortality rates. We carry out the analysis on two age ranges to test the robustness of our conclusions.

6.3 Methodology

Many of the approaches to mortality modelling used in the actuarial literature are based on a principal component analysis (PCA) of time series of mortality data by single age. The Lee-Carter model is a one –PC model and other multifactorial derivatives of this model add further cohort terms or additional factors to capture younger or older age mortality. Yang et al. (2010) building on previous PCA studies of mortality (Bell, 1997; Hyndman and Ullah, 2005) considers a two-PC model.

An econometric literature on factor analysis is well-developed. Factor analysis has been used extensively in economic forecasting, modelling business cycles and analyzing contagion effects of economic crises. In order to put an economic interpretation on latent factors extracted in these cases, Bai and Ng (2006) developed a statistical test for large cross section (N) and large time dimension (T) datasets to test the adequacy of observed variables as proxies for the unobserved factors. These tests take into account that latent factors are not known but must be estimated.

Assuming that a set of N variables,m_{it} can be described by a weighted linear com-

bination of r (smaller than N) factors,F_t, we can apply Factor Analysis to the datasets. This statistical technique accounts for the maximum amount of data variance with a small number of factors while best reproducing the observed correlations between the variables.

$$m_{it} = \lambda_i' F_t + e_{it} \tag{6.2}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. In classical factor analysis the error terms e_{it} are presumed to be independent across i and t. In approximate factor analysis this condition is relaxed.

Using principal components as estimates for the factors, the matrix of factor estimates $\tilde{F} = (\tilde{F_1}, \ldots, \tilde{F_T})'$ is given by the r eigenvectors associated with the largest eigenvalues of the matrix $\frac{XX'}{NT}$. The factor loadings $\Lambda = (\lambda_1, \ldots, \lambda_N)'$ are given as $\Lambda = \frac{X'\tilde{F}}{T}$. In order to determine r, we use the stopping rule for principal component analysis of the approximate factor model developed by Bai and Ng (2002). A number of variants of the information criteria are given with the most popular statistic being:

$$IC_p(r) = log\tilde{\sigma}^2(r) + r\frac{N+T}{NT}\ln[min(N,T)] \tag{6.3}$$

where $\tilde{\sigma}^2(k) = \frac{1}{NT}\sum_{i=1}^{N}\tilde{e_{it}^2}$ and the tilda ( ) indicates estimation by PCA. The number of factors r for which this criterion is minimised gives the estimated number of factors $\tilde{r}$.

Given a matrix G_t of m observed variables, we want to know if they are a linear combination of the r latent variables F_t. Tests have been developed for testing each variable of G_t singly and for testing G_t as a group. Considering the single tests, each variable of G_t may be an exact factor i.e. $G_{jt} = \delta_j' F_t \forall t$ or an approximate factor $G_{jt} = \delta_j' F_t + \epsilon_{jt} \forall t$. Let $\hat{\delta_j}$ be the least squares estimate of δ_j . Two tests have been

developed for the exact case. Letting $\hat{G_{jt}} = \hat{\delta_j'}\hat{F_t}$ and $\tau_t(j) = \frac{\hat{G_{jt}} - G_{jt}}{(var(\hat{G_{jt}})^{1/2})}$, we count the proportion of the time series for which $\hat{G_{jt}}$ deviates from G_{jt} by more than ϕ_α, the α percent critical value of the limiting distribution of $\tau_t(j)$.

This gives the statistic

$$A(j) = \frac{1}{\tau}\sum_{1}^{\tau} 1(\hat{\tau}_t(j) > \phi_\alpha) \tag{6.4}$$

We also test how far $\hat{G_{jt}}$ is from G_{jt} using the statistic

$$M(j) = max_{1 \leq t \leq T} \mid \hat{\tau}_t(j) \mid \tag{6.5}$$

This is a more stringent test as it demands that $\hat{G_{jt}}$ be close to G_{jt} at every point in time. Here, e_{it} must be serially uncorrelated for the limiting distribution of $\tau_t(j)$ to be asymptotically normal.

In the approximate case, we use two goodness of fit statistics:

1. the noise to signal ratio $NS(j) = \frac{\hat{var}(\hat{\epsilon(j)})}{\hat{var}(\hat{G(j)})}$
2. the coefficient of determination $R^2(j) = \frac{\hat{var}\hat{G}(j)}{\hat{var}(G(j))}$

Testing the group G_t as a set, the canonical correlations between G_t and F_t are considered. The first canonical correlation, ρ_1, is the largest correlation that can be found for linear combinations of G_t and F_t. The second canonical correlation, ρ_2, is the largest correlation that can be found from linear combinations of G_t and F_t uncorrelated with those giving the first canonical correlation, and so on. Having to estimate F_t has no effect on the sampling distribution of the canonical correlations. For $k = 1, \ldots, min[m, r]$ and $(F_t', G_t')\prime$ identically independently normally distributed,

$$(\rho_k^{2-}, \rho_k^{2+}) = (\tilde{\rho}_k^2 - 2\phi_\alpha \frac{\tilde{\rho}_k(1-\tilde{\rho}_k^2)}{\sqrt{T}}, \tilde{\rho}_k^2 + 2\phi_\alpha \frac{\tilde{\rho}_k(1-\tilde{\rho}_k)}{\sqrt{T}}) \tag{6.6}$$

where $\tilde{\rho}_k$ is the kth canonical correlation between G_t and $\tilde{F}_t$. If all the m variables in G_t are exact factors then the canonical correlations will all be unity. If the m variables are linearly dependent then the number of non-zero canonical correlations will be less than m. Any single variables in G_t may be found to be exact or approximate factors from the single tests but may be a linear combination of other observed variables as indicated by the group tests.

Having identified the most appropriate exogenous factors to build into our model of mortality we take the models of Girosi and King (2008) and its extension allowing for exogenous variables (King and Soneji, 2011) as a starting point to build our epidemiologically informed model of mortality. Girosi and King (2008) developed a method of modelling mortality rates across ages, years and countries which uses a Bayesian hierarchical approach to information pooling. Their objective in doing this was to make use of beliefs that data across neighbouring ages, years or countries should show similar characteristics. For example, we might expect that the mortality rate experienced by a 20 year old in a given year should be similar to that experienced by the 21 year old or the 19 year old in the same year. Similarly, the mortality rate in say 2000, for a given age should be similar to the mortality rate for that same age in 1999 or in 2001. The hierarchical approach allows the smoothing of mortality rates for a single country across ages and time and so produces realistic forecasts of mortality that do not break norms in terms of age and time going forward (for example, mortality rates increasing with age and improving in time). Considering the logarithmically transformed mortality rate during year t for life aged x as $m_{x,t}$ they set out the following model specification:

$$m_{x,t} \sim N\big(\mu_{x,t}, \frac{\sigma_x^2}{b_{x,t}}\big) \tag{6.7}$$

$$\mu_{x,t} = Z_{x,t}\beta_{x,t}$$

This specification only differs from a standard linear regression model in the $b_{x,t}$ weighting that is applied to the variance and in the approach to defining the parameters β_x and σ_x^2. The specification above provides the basic building block of the Bayesian hierarchical approach in which we now interpret the coefficients β_x and standard deviations σ_x^2 as random variables with their own prior distributions. The prior on the coefficients β_x which depends on its own "hyper-parameter" θ is denoted $P(\beta|\theta)$ with prior on the hyper-parameter $P(\theta)$. The prior for the variance random variable $\sigma$is denoted $P(\sigma)$. The functional form of the priors is chosen to be tractable and diffuse so as not to influence the results with a gamma or inverse gamma density function being used.

The prior for the coefficient β is chosen to reflect the "similarity" belief across cross sections. This is formalised by introducing a density function for the prior defined as:

$$P(\beta|\theta) \propto exp\big(-\frac{1}{2}H^{\beta}[\beta,\theta]\big) \tag{6.8}$$

where

$$H^{\beta}[\beta,\theta] \equiv \frac{1}{2}\sum s_{i,j} \parallel \beta_i - \beta_j \parallel_{\theta}^2 \tag{6.9}$$

where the notation $\parallel \beta_i - \beta_j \parallel_{\theta}^2$ denotes a weighted Euclidean norm and where the symmetric matrix s is called the adjacency matrix.

It's entries reflect the "proximity" of cross section i to cross section j and hence the weight put on the relationship between the coefficients of cross section i and cross section j. Using this approach the fitted model shows forecasts that are smooth in the age and time dimension and that do not violate the smoothness beliefs across age and time that "may" be violated by using multiple regression methods.

Linear regression provides a useful framework for including potentially informa-

tive covariates, either a 'cohort effect' (e.g., a cohort's earlier smoking patterns) or a 'period effect'. Further by doing this within their model they also incorporate the empirical regularities of smoothing by age and time imposed in this set-up. The method used to develop their model with exogenous covariates was to identify links between mortality rates and lagged covariates, specifically smoking habits and obesity. They argue against using contemporaneous relationships in favour of lagged relationships and from the literature determined the optimal lag period to be 25 years in the case of smoking. They also argue that the additional forecasting step required to project the exogenous variables would lead to additional uncertainty in the model. We argue in this paper that this is not the case where the most appropriate covariates are first identified by objective statistical criteria. Furthermore, although it may be appropriate to use current data to determine future mortality in the case of smoking rates this approach does not facilitate the inclusion of more contemporaneous variables such as current GDP, health expenditure, alcohol consumption or diet. Therefore we attempt to forecast these variables here while acknowledging better forecasts could be obtained using more adequately specified structural models or more sophisticated statistical techniques.

In our model we forecast the identified exogenous variables using ARIMA methods and, taking the resulting forecasts, we build a model of mortality with exogenous variables using the King and Soneji approach.

6.4 Results

6.4.1 Identifying exogenous factors

The pattern of mortality change over the last sixty years can be seen in figures 6.1-6.3 for the UK, US and Japan. Although not labelled, it can be inferred that mortality

rates at younger ages are near the bottom of the graphs and older age mortality rates at the top. A secular decline in mortality at all ages can be seen in each country. The most noticeable differences between the graphs are at the younger ages. UK death rates for those aged 20-40 are subject to more noisy fluctuations than those at other ages. Figure 6.2 shows US mortality rates for these ages increasing gradually around 1970, decreasing over the next decade and increasing again around 1990. Younger male mortality rates have decreased in Japan much more quickly than mortality rates at older ages especially over the period 1950-1980.

We first of all try to get a sense of the latent factor structure of the mortality data for males in each country over the fitting period 1970-2000. This shorter time period was chosen due to data constraints imposed by the availability of suitable exogenous variables. The number of factors is first determined and these factors are analysed to check their association with younger or older age mortality variation.

Applying the stopping rule (equation 6.3) we find a similar factor structure for each of the three countries: the estimated number of factors is $\hat{r} = 2$ for the UK and captures 86% of the variation in the data while $\hat{r} = 4$ for the US and Japan capturing 98% of the data variation in both cases. A Lee-Carter model with one factor or other early derivatives of this model would be therefore inadequate to capture all the common variation in the US and Japanese data while more recent multifactorial models such as Plat (2009) or Cairns et al. (2006) would provide a factor structure of a more suitable dimension for model fitting.

Before associating the factors extracted from the data with real-world trends, the *communality* (the percentage of the variation explained) at each age is estimated and graphed in figure 6.4. In US and Japan male mortality data, the four principal components extracted explain almost all the variation in the data at every age. In the UK data, older age mortality is also almost completely explained by the two principal compo-

nents. However younger age mortality variation is only partially explained by common factors and is subject to factors peculiar to particular ages or very small groups of ages (as was mentioned above in relation to figure 6.2). Causes of death among younger males in the UK are different at different ages with cancers and circulatory diseases accounting for a large proportion of deaths among 35-39 year-olds while among 20-24 year-olds external causes such as transport accidents, suicides and violence are more significant.[2]

The common factors extracted for each country tend to be associated with particular ages. From the factor loadings graphed in figures 6.5 and 6.6, we see that UK and US factors are either associated with younger or older age mortality. For the UK (figure 6.5) male mortality over 45 years of age is explained by factor 1 while the other factor explains younger age mortality. From figure 6.6, it can be seen factors 1 and 3 explain US male mortality at older ages while the other two factors explain younger age mortality. This would indicate that we need at least two types of exogenous factors to explain the variation in mortality rates: perhaps lifestyle-related factors to explain younger age mortality (e.g. alcohol consumption) and factors related to health treatment improvements to explain older age mortality. Japanese mortality rates behave differently with most factors not particularly associated with any particular ages except perhaps factor 4 on which middle age mortality ages are more heavily loaded. This would indicate that we require exogenous factors associated with mortality improvements at every age (e.g. income) or alternatively a large set of exogenous factors which together explain each principal component extracted.

The proposed exogenous factors are graphed for each country in figures 6.8-6.13. Alcohol consumption in the US and Japan has peaked and declined while it continues to increase in the UK. The US decline although not exactly concurrent with declines

[2] See for example mortality data for 2000 in Office for National Statistics (UK), Mortality Statistics, Series DH2 No. 27. Table 3

in younger age mortality observed in figure 5 suggest a possible explanation for trends there. Even at 2000 levels, UK annual consumption at 10.4 litres of pure alcohol per capita was moderate compared to other countries in the OECD - Luxembourgers consumed 15.4 litres per head in the same year - although the more harmful pattern of binge drinking is more common among young adults in the UK that in many countries (Kuntsche et al., 2004).

Smoking decline (figure 6.9), economic growth (figure 6.12) and steep increases in health expenditure (figure 6.13) may explain declines in mortality at all ages for all countries. Diets have been improving in the UK with less fat and more fruit and vegetables being consumed. The opposite is true in Japan with fruit and vegetable consumption declining slightly and fat intake increasing dramatically although it was still second lowest in the OECD in 2000. Good diet behaviours have been negated by bad ones in the US with both more fat and fruit and vegetables being consumed.

Each of the proposed exogenous determinants of mortality are, in turn, compared to the principal components extracted from the data using the statistical tests outlined in 6.4 and 6.5. Results for the UK are given in Table 2. None of the proposed exogenous determinants is an exact factor using the A(j) statistic which should be 5% if factor j is an exact factor. Health expenditure per capita comes closest to being a linear combination of the extracted principal components with A(j) = 0.484. The A(j) statistic allows the relationship between the exogenous determinants of mortality and the latent factors not to hold at some points in time. The M(j) statistic is a stronger test and requires that at every point in time the relationship must hold within a small degree of error. Using a 5% significance level with T=50 the critical value is 3.28. Not surprisingly the test rejects all of the proposed factors. Bai and Ng (2006) note that if anything both tests are underpowered so we can safely conclude that none of the proposed factors are exact factors. Allowing for a close relationship between pro-

posed factors and latent factors as opposed to an exact relationship is more realistic where variables are measured with error, the statistical indicators do not reflect the underlying construct accurately, the relationship might not be exactly linear or the relationship might be moderated by other factors. The goodness of fit statistic $R^2(j)$ and the noise to signal ratio (NS(j)) indicate how far the proxies are from the true factors. Bai and Ng suggest that if NS(j) > 0.5 and/or R^2(j) < 0.95 then errors in the linear relationship between the proposed factors and the latent factors are non-negligible and the proposed factors are not strong proxies for the latent factors. According to these measures, Health expenditure per capita and GDP per capita are particularly strong proxies and Total fat intake is a particularly poor proxy. Of course, numerous studies have found cointegration between national income and health expenditure (Freeman, 2003; Westerlund, 2007; Moscone and Tosetti, 2010) and using the two variables may provide little extra information than simply using one.

The squared canonical correlations are given in the final column. The first value indicates that there is a linear combination of the proposed proxies and a linear combination of the two latent factors that are highly correlated. The second value indicates that any linear combinations orthogonal to those already found are much less correlated (the confidence interval for the second canonical correlation almost includes zero). There is therefore only one well-defined relation between the two sets and the set of six proxy factors as a whole does not span the latent factor space. As both GDP per capita and health expenditure per capita are both individually highly associated with the latent factors either variable can be used to improve forecasting models.

The results for the US are given in Table 6.4. None of the proposed exogenous determinants is an exact factor A(j) - alcohol consumption followed by health expenditure per capita are closest to being exact factors according to both the A(j) and M(j) statistics. The importance of alcohol consumption in the US contrasts with its rela-

tive unimportance in explaining variations in the UK mortality data. This bears out the correlation noted above between variations in US younger age morality and alcohol consumption. Allowing for some deviation between the observed variables and the latent factors (columns 4 and 5) and using the rule from before, Health expenditure, Tobacco consumption, Alcohol consumption and GDP are all strong proxies for the four latent factors. However, the set of six proposed factors does not span the latent factor space as there are only two well-defined relations. The squared canonical correlations between the latent factors and a set of just the two variables Alcohol consumption and Health expenditure per capita are 0.997 and 0.934 suggesting that little is gained by adding the extra four variables. For the purposes of forecasting, these two variables - one a lifestyle variable and the other a medical care variable - appear to be strongly associated with mortality trends and are sufficiently orthogonal to provide distinct forecasting power.

In the case of Japan (Table 6.6), none of the proposed exogenous determinants is an exact factor. Almost all the variables except tobacco consumption and fruit and vegetable consumption are strong proxies for the four latent factors. This finding is in keeping with figure 6.7 where a more complicated latent factor structure than for the other two countries was observed. Nevertheless, the set of factors considered does not encompass the latent factor structure. There are two well-defined relations between proxies and latent factors although the third canonical correlation is large. As health expenditure and GDP cointegrate, the three variables - alcohol consumption, fat intake and health expenditure - provide an appropriate basis for forecasting models. This set has squared canonical correlations of 0.995, 0.851 and 0.416 with the latent factors which when compared to column 6 of Table 6.6 indicates some information is lost by focusing on this smaller subset.

Although we have actual values of all potential covariates for 2001-06 we decided

that using predictions of these values would make a fairer assessment of any gains in forecastibility. Forecasts were made using the Box-Jenkins methodology. The usual approach of deciding on the number of auto-regressive (AR) and moving average parameters (MA) using information criteria gives poor forecasts. The inclusion of moving average terms generally tends to worsen forecasts. We presume that a researcher would realise this in practice by testing their model. Therefore, the Schwarz information criteria were used to decide between models with various AR terms. As GDP is generally found to be non-stationary (e.g. Westerlund, 2007), this variable was first-differenced and consequently health expenditure per capita also. In summary, the ARIMA(p,d,q) models were UK GDP (2,1,0), US health expenditure (1,1,0), US Alcohol (3,0,0), Japan health expenditure (0,1,0), Japan alcohol (1,0,0) and Japan fat (1,0,0). Assessing prediction errors post-hoc, these models did not necessarily provide the best forecasts but reflect the level of uncertainty encountered in practice. This approach is not dissimilar to forecasting the common factor with ARIMA in Lee and Carter (1992) and making mortality forecasts conditional on these forecasts.

Taking the predicted exogenous factors we apply the King and Soneji (2011) approach to forecast mortality rates. We do this using the YourCast software developed by Girosi and King[3]. For U.K. we use GDP as the only exogenous variable to be built into the model while for the U.S. heath expenditure and alcohol consumption are used. Finally for Japan we use health expenditure, alcohol consumption and fat intake. We present the fitting and forecasting results of our model in tables 5-10, for the U.S., U.K., and Japan. We also present the results of fitting and forecasting the mortality data using the models of Lee Carter(1992), CBD(2006) and the Girosi and King (2008) model with no exogenous variables for comparison. For a given mortality rate at time t and for age x ($m_{x,t}$), we measure the fitting and forecasting quality using

[3]For more details on the YourCast software used in this study and developed by King and Soneji go to http://www.gking.harvard.edu/yourcast.

the three measures E1, E2 and E3 outlined below taking the standardized error to be

$$Error_{x,t} = \frac{projected(m_{x,t}) - actual(m_{x,t})}{projected(m_{x,t})}$$

(i). The average error, E1 – this equals the average of the standardized errors i.e.

$$E1 = \frac{1}{X_1 - X_2 + 1} \sum_{x=X_1}^{X_2} \sum_{t=1}^{T} Error_{x,t} \tag{6.10}$$

This is a measure of the overall bias in the projections.

(ii). The average absolute error, E2 – this equals the average of absolute value of the standardized errors i.e.

$$E2 = \frac{1}{X_1 - X_2 + 1} \sum_{x=X_1}^{X_2} \sum_{t=1}^{T} |Error_{x,t}| \tag{6.11}$$

This is a measure of the magnitude of the differences between the actual and projected rates.

(iii). The standard deviation of the error, E3 – this equals the square root of the average of the squared errors,

$$E3 = \sqrt{\frac{1}{X_1 - X_2 + 1} \sum_{x=X_1}^{X_2} \sum_{t=1}^{T} (Error_{x,t})^2} \tag{6.12}$$

This is the root mean square of the standardized errors.

The results for the full 20-89 age range are presented in tables 6.8-6.10 (forecasting results) and 6.11-6.13 (fitting results). We first comment on the comparison of atheoretical models ignoring the use of exogenous variables and then secondly on the value added by incorporating epidemiological information in the King and Soneji model focusing on the forecasting results.

6.4.2 Forecasting results

The most important point to note is that our technique, using forecast exogenous variables in a structured model (labelled King and Soneji in Tables 6.8-6.10) performs best across all the countries in the study and using all three measures of fit quality. The improvement over the atheoretical models of Lee Carter and Girosi and King is of the order of 1-2% for the U.K. and the U.S irrespective of error measure and 4-4.5% for Japan. The CBD model performs particularly badly over the 20-89 age range but this was already noted by Cairns et al (2006) where the model is only fitted to older age ranges.

The three measures E1, E2 and E3 are measuring differing aspects of the forecasting and fitting quality. E1 measures the level of bias seen in the forecast whereas E2 and E3 measure the level of spread seen in the forecast with E3 only differing from E2 in the significance placed on outliers. When looking at the forecasting results on E1, the level of bias is seen to be lower in the King and Soneji model than in the atheoretical models. The difference when compared with Lee Carter in the case of Japan is particularly marked at 3.8% (Table 6.10). In some cases the measures for E1 and E2 are the same suggesting that the fitted mortality rates for some models and some countries are overestimated for every age and every year. This indicates that true mortality rates are improving at a faster rate than the atheoretical models can accommodate and suggests that perhaps using a logarithmically transformed mortality rate is no longer sufficient to linearize the data before applying these types of models.

In most cases the Girosi and King method outperforms the Lee Carter model. Indeed for the Japanese case, the Girosi and King model results in poorer forecasts when compared with the Lee Carter model (E1-E3 in Table 6.10). This may be due to the more complex factor structure of the Japanese data and the associated cost in terms of fit quality when applying the smoothing method employed by the Girosi and King

methods.

Table 6.10 which shows results for Japan would indicate that the improvement in forecasting quality appears to be more pronounced where we have been able to identify more exogenous variables to explain the latent factor structure. The statistical analysis identified three exogenous variables to explain the latent factor structure and the improvement in forecasting over the next best model (Lee Carter) is 3.9% on E2 and 4.3% on E3. In the case of the U.K. (Table 6.8) where the statistical analysis identified only one exogenous variable to explain the latent factor structure the improvement in forecasting over the next best is more modest at 1.1% on E2 and 0.9% on E3. With more data and a bigger spanning set of exogenous variables we may be able to repeat the superior performance shown in the Japanese case for the U.K. and U.S. as well as other countries not considered here.

Looking at figures 6.14-6.16, where we exclude the CBD model due to its poor performance, we can see the fitted (from 1970-2000) and forecast (2001-2006) mortality rates for each of the models considered in this paper along with the actual mortality rates. The first point to note is that we can see a high level of volatility in the actual mortality rates (the solid bold lines) for 20 and 40 year olds. The profile for 60 and 80 year olds is less volatile perhaps due to larger numbers of deaths. In each case the Lee Carter model (the solid line with "x" symbols) appears to overestimate the mortality rate and this is particularly pronounced in the case of Japan (see Figure 16a for example). The Girosi and King model appears to underestimate mortality rates for most countries particularly where there has been an upward trend in the most recent mortality experience (Figure 15a). This can be attributed to the smoothing mechanism of the Girosi and King specification. Incorporating exogenous variables in our model rectifies some of this underestimation. A particularly good example of our approach improving upon the Girosi and King approach is for younger age mortality where the

inclusion of exogenous variables substantially corrects for underpredictions leading to very accurate forecasts (Figures 14b and 15a). This is possibly due to the fact younger age mortality shows less common variation with mortality at older ages and is more responsive to trends in lifestyle behaviours. A more detailed analysis of the performance of the King and Soneji model at each individual age and year (not shown) shows it gives the most accurate forecasts for the middle-ages, 30-60 years of age, for the U.K. and this is consistent across all measures of fitting quality. For the U.S. the story is different, where here the forecasting performance is superior mainly over the older ages. In Japan where we had the largest set of exogenous determinants, performance is superior across many young and many older ages. This would indicate that the larger the set of exogenous variables identified the broader the age range over which forecasts are better.

6.4.3 Fitting results

Turning to the fitting results (Tables 6.11-6.13) our model is generally superior across all measures of fit quality but not across all countries. The exceptions are the UK where the fit is poorer on all error measures and U.S. where fit is worse on measure E1. Looking at the fitting quality of the Girosi and King approach against the Lee Carter model and the CBD model, it can be seen that the Girosi and King model shows less bias (shown by the lower value of E1) for all three countries considered in the study. However, the Girosi and King model shows a poorer fit quality for all three countries measured on the E2 and E3 measures (e.g. 0.4% worse on E2 and 0.2% on E3 for the UK). This can be explained by the smoothing mechanism in the Girosi and King model which improved smoothness at the cost of a more accurate fit. Finally, when we include forecast exogenous variables the fit is improved. For example, comparing the King and Soneji model with Girosi and King for Japan there is much less bias (0.07% compared

with 0.24%) and forecasts show less variation around the actual realised mortality rates (the E2 measured is reduced by 2.1% and E3 is reduced by 2.6%). The improvement in fit would be expected for a model with a greater number of explanatory variables. However an overparameterised model will not necessarily give better forecasts. That we have found improved forecasts by incorporating additional exogenous variables indicates that this approach has merit.

6.4.4 Robustness

As mentioned above in section 4.1, figures 6.5-6.7 indicate that in all three countries mortality among those 40 and over share a common latent factor while mortality rates among those under 40 share common drivers distinct from those at older ages. It therefore would seem appropriate to consider mortality at older ages separately to see if forecast can be further improved in this age range. The drawback of using a narrower age range is that common factors are estimated on fewer variables and the results of Bai and Ng (2002 and 2006) are only true asymptotically.

Conclusions from the latent factor analysis on ages 40-89 are very similar to the analysis based on the age range 20-89 (not shown). For the UK as before, none of the proposed exogenous determinants is an exact factor and income and health expenditure are again identified as suitable proxies for the true factors. The squared canonical correlations indicate only one well-defined relation between the two sets and either GDP per capita or health expenditure per capita can be used to improve forecasting models. For the US, health expenditure per capita is found to be an exact factor and alcohol, tobacco and income are strong proxies. There are two-well defined relations between the proxies and the latent factor space although the third canonical correlation is larger than before at 0.577. The squared canonical correlations between the latent factors and just the two variables Alcohol consumption and Health expenditure per

capita are 0.998 and 0.882 suggesting that little is gained by adding the extra four variables. In the case of Japan, the stopping rule indicates the factor structure has one more factor than before i.e. five factors. None of the proposed exogenous determinants is an exact factor and the same set of strong proxies as before is identified: Alcohol, Fat, GDP and Health Expenditure. Canonical correlations indicate two well-defined relations with the third bigger than before at 0.801 (compared to 0.786 in Table 6.6). The three variables alcohol consumption, fat intake and health expenditure therefore provide an appropriate basis for forecasting models.

The forecasting performance of the King and Soneji model is still generally superior to the atheoretical models considered except in the US where Girosi and King is best on all error measures (Tables 6.14-6.16). This may be due to the econometric limitations of applying techniques developed for datasets of large number of variables to narrow age ranges. Nevertheless for the UK and particularly Japan, the inclusion of exogenous factors improves forecasts on this age range. Results for the CBD model are dramatically improved but still inferior to all other models for the US and Japan.

6.5 Conclusion

In this paper we have taken an econometric approach recently developed in macroeconomics and finance to develop links between the latent factor structure of mortality data and observable factors. We have employed stopping rules to manage the number of factors and so avoid overfitting models. We explain the latent factors by comparing with exogenous factors deemed plausible by health economics and epidemiological literature. The techniques employed in this paper are novel to this literature although well known as an econometric method. The incorporation of the identified exogenous factors into mortality models are seen to robustly improve forecasts.

We focus on data from the U.S., U.K. and Japan and note that in each of the cases that a differing number of exogenous variables seem to explain the mortality variation in each case. In the case of U.K. we identify one exogenous economic factor; for the U.S., 2 factors are identified - one related to medical care and the other to lifestyle and finally for Japan, we identify 3 factors - two related to lifestyle and the other to medical care. We look at both 20-89 and 40-89 age ranges and note that the same exogenous variables are identified in each case as being significant. We forecast these exogenous factors using ARIMA methods and apply a model which incorporates these exogenous factors using the methods of King and Soneji. Using our approach forecasts can be dramatically improved over the standard atheoretical models.

Further work will involve expanding the set of the exogenous variables that could be useful in explaining mortality variation ; using age-specific variables instead of aggregate measures ; considering the inclusion of lagged variables where appropriate and considering the ability of the model over a longer term forecasting period. Also, the application of the semi-parametric estimation approach of Connor et al. (2011) in this context would allow us to unite the identification of exogenous variables and estimation of the latent factor model in one step.

6.6 Appendix: Additional Figures and Tables

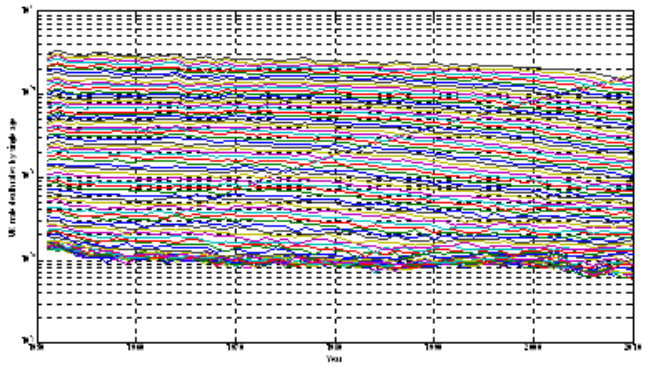

Figure 6.1: UK male crude mortality rates, 1950-2009 (log scale)

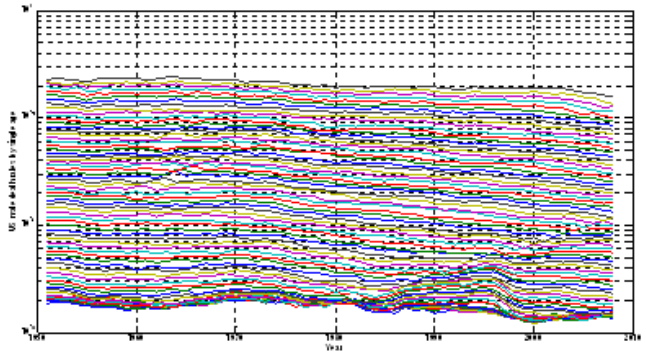

Figure 6.2: US male crude mortality rates, 1950-2009 (log scale)

Table 6.1: Descriptive statistics, 1970-2000 (in order UK, US and Japan)

	Mean	Standard Deviation	Definition
Alcohol	(UK) 9.3 (US) 9.5 (Japan) 7.8	0.6 0.8 1.0	Annual consumption of pure alcohol in litres, per person, aged 15 years and over
Tobacco	2349 2645 3227	516 667 147	Annual consumption of tobacco items (eg cigarettes, cigars) in grams per person aged 15 years and over
Fat	138.9 133.5 73.1	3.0 10.0 8.9	Total fat (grams per capita per day)
Fruit & Veg	151.8 219.7 173.1	14.3 17.6 8.8	All fruit and vegetable consumption (except wine) in kilos per capita
GDP	11896 25420 2,994,819	2275 4712 707,685	Gross domestic product per capita in national currency units at 2000 price levels
Health exp	712 2787 191,957	222 1098 63,732	Total health expenditure (private and public) per capita in national currency units at 2000 price levels

Table 6.2: Testing the factors in UK male crude mortality rates by single age 20-89, 1970-2000

G_j	A(j)	M(j)	R^2(j)	NS(j)	$\hat{\rho}(k)^2$
Alcohol	0.839	49.14	0.546	0.832	0.991
			(0.310, 0.782)		(0.987, 0.997)
Tobacco	0.871	24.62	0.809	0.236	0.323
			(0.688, 0.930)		(0.052, 0.594)
Fat	0.645	26.43	0.313	2.195	-
			(0.042, 0.584)		
Fruit & Veg	0.871	28.02	0.815	0.227	-
			(0.698, 0.933)		
GDP	0.645	10.43	0.967	0.035	-
			(0.944, 0.990)		
Health exp	0.484	14.80	0.970	0.031	-
			(0.949, 0.991)		

Table 6.3: A(j) is the frequency that $|\hat{\tau}_t(j)|$ exceeds the 5% asymptotic critical value. M(j) is the value of the test. R^2 is defined in 6.5, NS(j) defined in 6.4 and $\hat{\rho}(k)^2$ is the vector of canonical correlations G_t with respect to F_t.

Table 6.4: Testing the factors in US crude mortality rates by single age 20-89, 1970-2000

j	A(j)	M(j)	R^2(j)	NS(j)	$\hat{\rho}(k)^2$
Alcohol	0.323	5.94	0.976	0.024	0.999
			(0.960, 0.993)		(0.998, 1.000)
Tobacco	0.710	7.53	0.991	0.009	0.951
			(0.984, 0.997)		(0.918, 0.985)
Fat	0.806	25.82	0.911	0.098	0.366
			(0.850, 0.971)		(0.096, 0.636)
Fruit & Veg	0.935	44.99	0.878	0.139	0.130
			(0.798, 0.959)		(-0.091, 0.352)
GDP	0.806	14.56	0.975	0.025	-
			(0.958, 0.992)		
Health exp	0.419	5.20	0.997	0.003	-
			(0.994, 0.999)		

Table 6.5: A(j) is the frequency that $|\hat{\tau}_t(j)|$ exceeds the 5% asymptotic critical value. M(j) is the value of the test. R^2 is defined in 6.5, NS(j) defined in 6.4 and $\hat{\rho}(k)^2$ is the vector of canonical correlations G_t with respect to F_t.

Table 6.6: Testing the factors in Japan crude mortality rates by single age 20-89, 1970-2000

j	A(j)	M(j)	R^2(j)	NS(j)	$\hat{\rho}(k)^2$
Alcohol	0.484	9.494	0.978	0.023	0.995
			(0.962, 0.993)		(0.992, 0.999)
Tobacco	0.484	6.893	0.804	0.244	0.938
			(0.680, 0.928)		(0.895, 0.980)
Fat	0.581	8.961	0.992	0.008	0.786
			(0.986, 0.998)		(0.652, 0.919)
Fruit & Veg	0.839	39.868	0.795	0.258	0.310
			(0.666, 0.924)		(0.039, 0.580)
GDP	0.484	8.338	0.993	0.007	-
			(0.988, 0.998)		
Health exp	0.548	7.130	0.992	0.008	-
			(0.987, 0.998)		

Table 6.7: A(j) is the frequency that $|\hat{\tau}_t(j)|$ exceeds the 5% asymptotic critical value. M(j) is the value of the test. R^2 is defined in 6.5, NS(j) defined in 6.4 and $\hat{\rho}(k)^2$ is the vector of canonical correlations G_t with respect to F_t.

Table 6.8: Forecasting results for the U.K. for ages 20-89 and years 2001-2006

Model	E1	E2	E3
Lee Carter	10.60%	10.65%	13.92%
Girosi and King	10.35%	10.36%	12.41%
CBD	18.85%	20.91%	27.78%
King and Soneji	9.25%	9.25%	11.55%

Table 6.9: Forecasting results for the U.S. for ages 20-89 and years 2001-2006

Model	E1	E2	E3
Lee Carter	8.27%	8.33%	11.16%
Girosi and King	7.46%	7.70%	10.22%
CBD	13.35%	15.33%	23.15%
King and Soneji	6.66%	7.02%	8.84%

Table 6.10: Forecasting results for Japan for ages 20-89 and years 2001-2006

Model	E1	E2	E3
Lee Carter	8.58%	8.58%	10.26%
Girosi and King	8.79%	8.79%	11.50%
CBD	14.31%	16.19%	23.41%
King and Soneji	4.73%	4.73%	6.01%

Table 6.11: Fitting results for the U.K. for ages 20-89 and years 1970-2000

Model	E1	E2	E3
Lee Carter	0.25%	3.50%	4.90%
Girosi and King	0.14%	3.90%	5.13%
CBD	-0.08%	14.42%	21.86%
King and Soneji	0.12%	3.76%	4.94%

Table 6.12: Fitting results for the U.S. for ages 20-89 and years 1970-2000

Model	E1	E2	E3
Lee Carter	0.33%	3.83%	5.78%
Girosi and King	0.04%	4.12%	5.99%
CBD	-2.26%	13.50%	19.62%
King and Soneji	-0.09%	2.55%	3.40%

Table 6.13: Fitting results for Japan for ages 20-89 and years 1970-2000

Model	E1	E2	E3
Lee Carter	0.39%	3.86%	5.05%
Girosi and King	0.24%	4.99%	6.46%
CBD	-6.39%	13.31%	20.68%
King and Soneji	0.07%	2.94%	3.85%

Table 6.14: Forecasting results for the U.K. for ages 40-89 and years 1970-2000

Model	E1	E2	E3
Lee Carter	10.96%	10.96%	13.95%
Girosi and King	10.67%	10.67%	12.24%
CBD	10.49%	10.49%	13.17%
King and Soneji	9.20%	9.20%	10.38%

Table 6.15: Forecasting results for the U.S. for ages 40-89 and years 1970-2000

Model	E1	E2	E3
Lee Carter	6.39%	6.39%	7.70%
Girosi and King	5.40%	5.40%	6.97%
CBD	9.08%	9.08%	10.71%
King and Soneji	5.72%	5.72%	7.52%

Table 6.16: Forecasting results for the Japan for ages 40-89 and years 1970-2000

Model	E1	E2	E3
Lee Carter	8.29%	8.29%	9.71%
Girosi and King	6.09%	6.09%	8.05%
CBD	9.45%	9.45%	11.03%
King and Soneji	4.84%	4.84%	6.21%

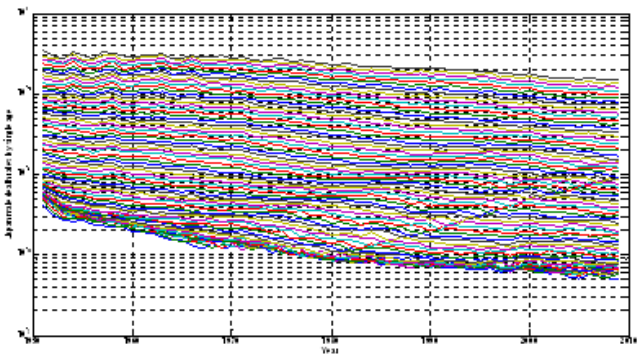

Figure 6.3: Japan male crude mortality rates, 1950-2009 (log scale)

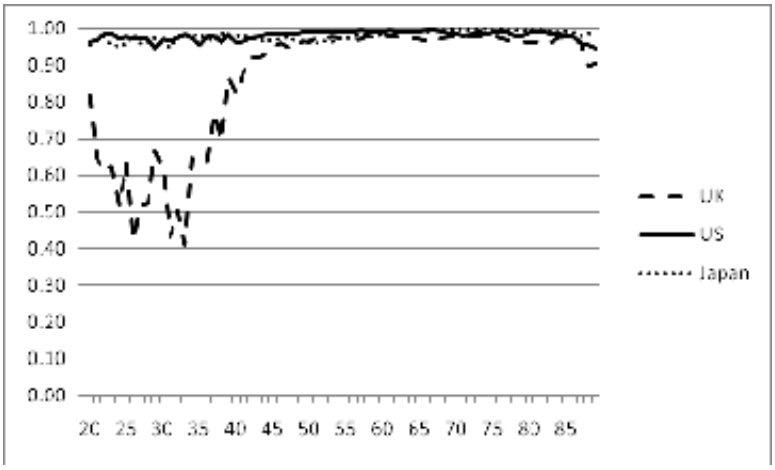

Figure 6.4: Proportion of variance explained by principal component extraction (communality) by age for male crude mortality rates, 1970-2000

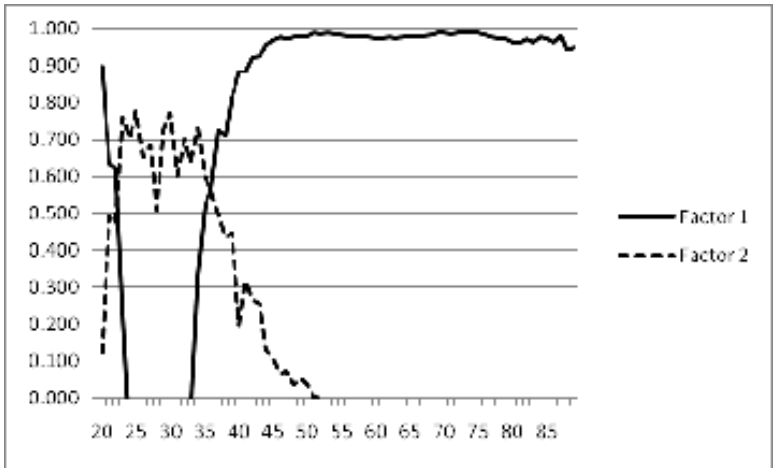

Figure 6.5: Rotated factor loadings by age for UK male crude mortality rates, 1970-2000.

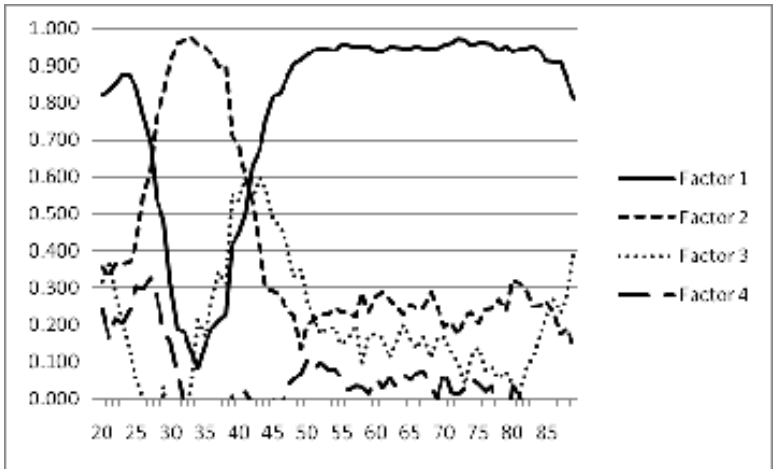

Figure 6.6: Rotated factor loadings by age for US male crude mortality rates, 1970-2000

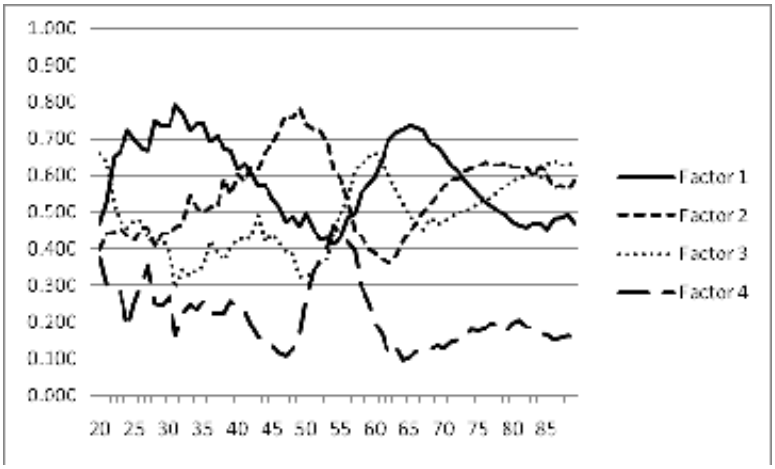

Figure 6.7: Rotated factor loadings by age for Japan male crude mortality rates, 1970-2000.

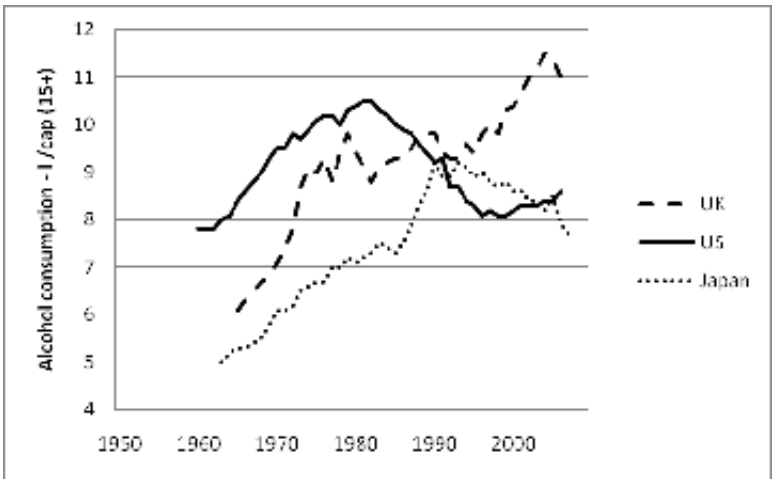

Figure 6.8: Alcohol consumption - Liters per capita (15+)

Figure 6.9: Tobacco consumption - Grammes per capita (15+)

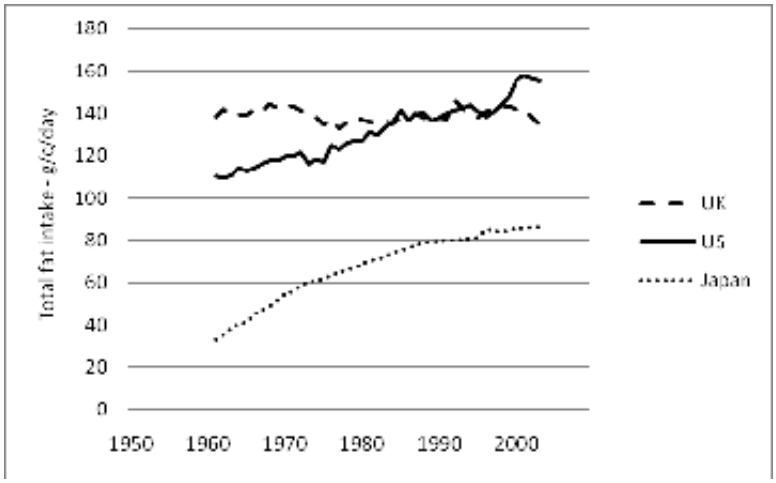

Figure 6.10: Total fat intake - grammes per capita per day

Figure 6.11: Fruit and Vegetables consumption - kilos per capita

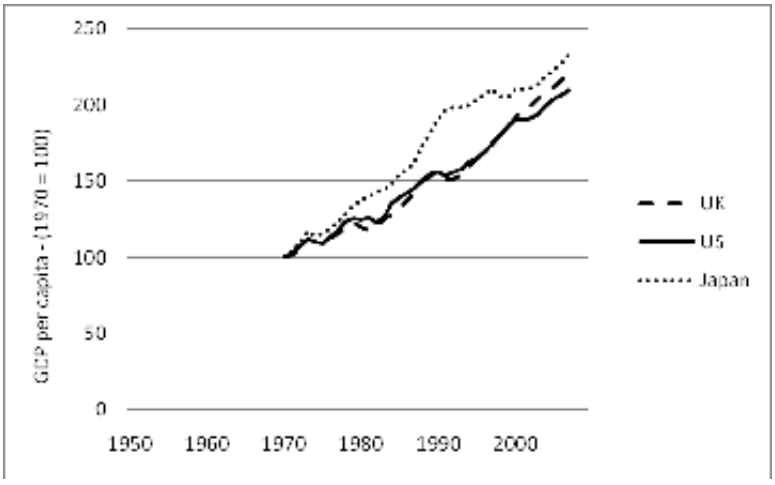

Figure 6.12: Gross domestic product per capita at constant prices (1970 = 100)

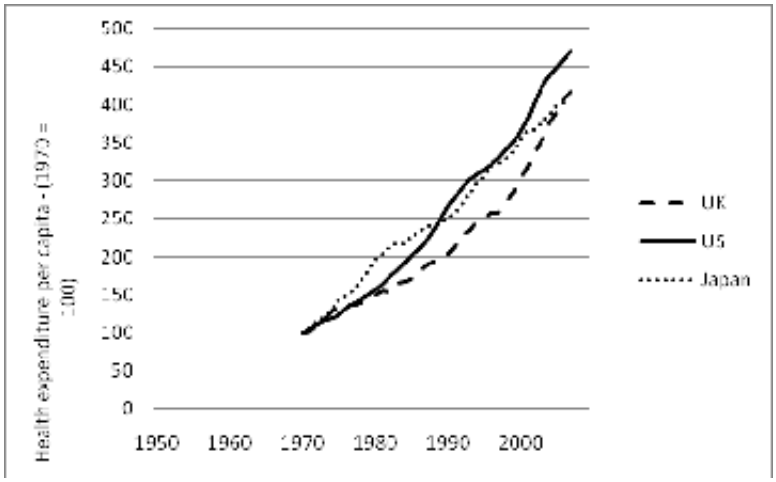

Figure 6.13: Total expenditure on health per capita at constant prices (1970=100)

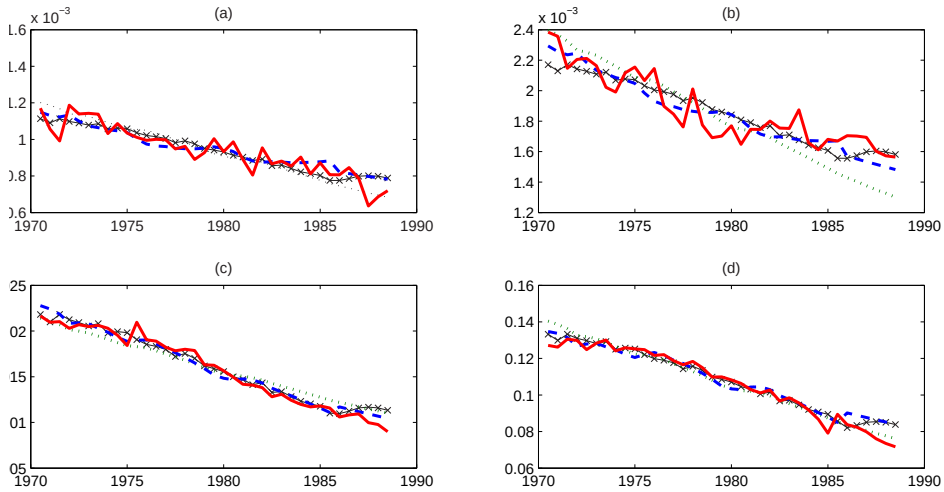

Figure 6.14: U.K. mortality rates fitted between 1970-2000, and forecast from 2001-2006 for the Lee Carter,(black with "x"'s), Girosi and King (green), King and Soneji (blue) models and actual mortality rates 1970-2006 (red) for males aged (a) 20, (b) 40, (c) 60 and (d) 80

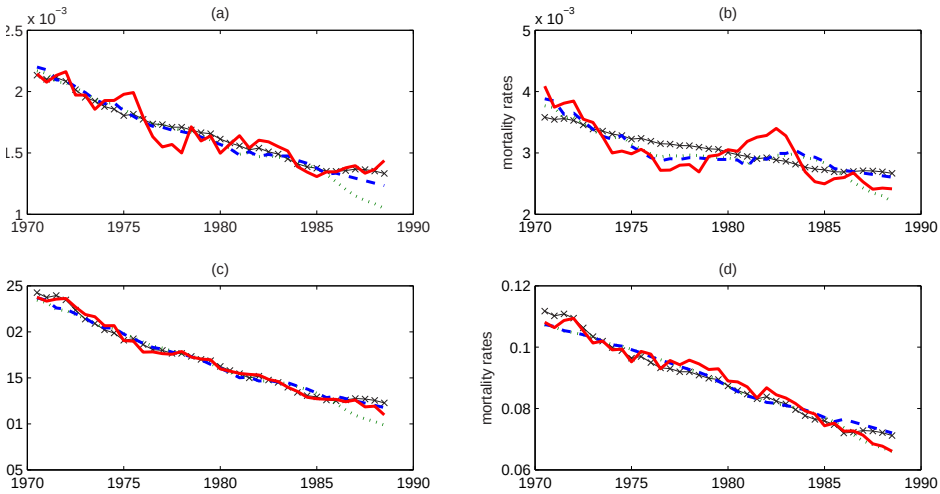

Figure 6.15: U.S. mortality rates fitted between 1970-2000, and forecast from 2001-2006 for the Lee Carter,(black with "x"'s), Girosi and King (green), King and Soneji (blue) models and actual mortality rates 1970-2006 (red) for males aged (a) 20, (b) 40, (c) 60 and (d) 80

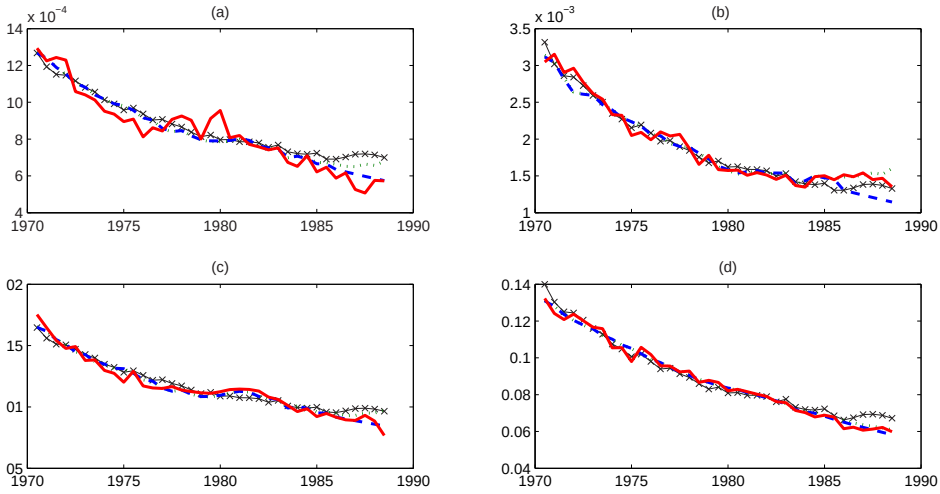

Figure 6.16: Japanese mortality rates fitted between 1970-2000, and forecast from 2001-2006 for the Lee Carter,(black with "x"'s), Girosi and King (green), King and Soneji (blue) models and actual mortality rates 1970-2006 (red) for males aged (a) 20, (b) 40, (c) 60 and (d) 80

Chapter 7

Conclusions

The field of mortality risk and longevity risk and in particular the accurate forecasting and financial management of such risks has become a topic of great interest to academics, actuaries and financial professionals. As individual life expectancies continue to improve and the era of low equity returns and low interest rates persists the current mechanisms for providing adequate coverage for individuals in their later years are coming under strain. In addition the development of financial hedging products has enabled many financial risks to be laid off and has exposed longevity risk as arguably the most significant un-hedged risk in the developed world. The new visibility of longevity risk and the constraint capacity of insurance firms and pension providers to accommodate it has led to a need and desire to create innovative ways to lay this risk off to new parties, namely the capital markets. This has resulted in the area of longevity becoming a key new growth area for the capital markets who are designing the products to be able to isolate, transfer and manage this risk.

Essential to the desire to create mechanisms to transfer longevity risk is the need to accurately forecast mortality rates. This will ensure that new products are priced appropriately and that a transparent market with willing sellers and buyers of the risk

can emerge. Current research in mortality modelling does not prioritise the forecasting of mortality rates, instead it focuses on providing models that give a best fit to the data and on providing adequate short term forecasts. It also focuses on statistical, extrapolative time-series methods rather than engaging with the socio-economic and epidemiological factors that may be causing mortality improvements.

In this thesis I contribute to the existing literature on modelling mortality rates with several papers in the broad areas of extrapolative modelling and socio-economic and epidemiologically informed modelling of mortality rates.

In chapter 2 of the thesis I identify and addressed a limitation of the Plat (2009) model and previous stochastic mortality models. This limitation is in the inability of existing models to adequately fit mortality rates at the lower ages due to the non-linear dynamics at the lower ages, the so called *lifestyle* mortality profile. I have designed a new 4 factor extrapolative model which is demonstrated to have better fit and forecasting ability on a wider age range (5-89). The results of the analysis exposed the weakness of previous models when trying to fit to non-linear features of the data and showed that a more non-linear flexibility was needed to capture the mortality profile, particularly at lower ages.

In chapter 3 of the thesis I look to the macro-economic forecasting research and develop an alternative approach to modeling and forecasting mortality rates using a method of Forni*et al.* The method is shown to give superior forecasts to many of the standard stochastic mortality models that form the basis of longevity hedging products. In particular the longer term forecasts were far superior. By focusing on the forecasting quality of the model using out of sample comparisons we can conclude that existing multifactorial models give poor forecasting performance and even with the simplest of specifications the dynamic model outperformed the existing models of mortality.

In chapter 4 of the thesis the issue of structural breaks in mortality trends is con-

sidered. Focusing on several of the leading extrapolative models of mortality rates and applying the methods of Bai and Perron (2003) I test for the presence of structural breaks in the model specifications. By testing the model residuals in each case I found that there was indeed a breakpoint visible in the residuals falling somewhere in the 1970's reaffirming previous demographic research. I then carry out the forecasting process again making allowance for the structural breaks identified with the results showing that in nearly two-thirds of cases the model allowing for structural breaks provided a more accurate forecast of mortality. These findings are significant in highlighting the importance of the sample period when fitting a model to mortality data.

In chapter 5 of the thesis the geographical variation of mortality is considered using socioeconomic explanatory factors in Northern Ireland data. Using a general to specific analysis we identified some differences between the males and females in our study, in terms of the covariates that proved to be significant in explaining the variation in mortality rates. In the case of males we identified that age standardised mortality rates could best be explained by the variation in healthcare, environment and political polarisation (through income and education). This suggests that the combined effect of living in a densely populated catholic (or protestant area) with poor education and income levels has an impact on mortality rates for males. In the case of females we identified that employment, healthcare and environment are important factors in explaining mortality rates. In both cases of males and females the quality of the immediate environment (heating, double glazing etc.) appeared to be a significant factor in explaining mortality rate variation.

To assess the ability of geographical location to explain mortality rates and the proposition that age standardised mortality rates may show some geographical clustering, models including independent frailties and spatial frailties were fitted to the

data. When we allowed for independent frailties we found some improvement in the fit measured on a deviance information criteria despite the additional parameterisation. However when we fit a spatial structure to the data the information criteria was not improved. We conclude that there is little evidence of spatial clustering in mortality rates in Northern Ireland.

Finally we considered the case of frailties when we have little or no socio-economic data available. In this case we seen again that with the inclusion of independent frailties the fit can be improved using a DIC measure and in this case the improvement is more dramatic which is to be expected since there is a larger unexplained element to the mortality rates. This supports the view that there is a place for geographical location data when we have limited other socio-economic data. The results showed that there is a place in mortality modeling for socio-economic characteristics.

Finally, in chapter 6 of the thesis we took an econometric approach recently developed in macroeconomics and finance to develop links between the latent factor structure of mortality data and observable (socio-economic and epidemiological) factors. We employed stopping rules to manage the number of factors and so avoided overfitting models. We explained the latent factors by comparing with exogenous factors deemed plausible by health economics and epidemiological literature. The techniques employed in the paper were novel to the mortality literature although well known as an econometric method. The incorporation of the identified exogenous factors into mortality models were shown to robustly improve forecasts.

We focused on data from the U.S., U.K. and Japan and noted that in each of the cases a differing number of exogenous variables seemed to explain the mortality variation. In the case of U.K. we identified one exogenous economic factor; for the U.S., 2 factors were identified - one related to medical care and the other to lifestyle and finally for Japan, we identified 3 factors - two related to lifestyle and the other to med-

ical care. We looked at both 20-89 and 40-89 age ranges as a robustness check and found that the same exogenous variables were identified in each case as being significant. We forecasted these exogenous factors using ARIMA methods and applied a model which incorporated these exogenous factors using the methods of King and Soneji. Using this approach forecasts were shown to be dramatically improved over the standard atheoretical models.

Chapter 8

Practical impact of findings

The research carried out in this thesis deals with the issue of accurately forecasting mortality rates. This is of enormous importance to the actuarial profession which was founded on the basis of mathematically pricing mortality risk through annuities. The first mortality tables were created using historic data and assuming that future lifetimes would continue to experience the same mortality risk as previous lives had.

The actuarial profession has published reports, working papers and updated mortality tables since the early 1970's through the Continuous Mortality Investigation (C.M.I.). These have been based on pension and insurance data gathered from its members and were designed to give actuaries a base from which to make their own judgements regarding the mortality experience to apply when advising their own particular clients. The tables created were deterministic in nature with the actuary having the ability to *age rate* the tables to better fit the particular client experience. In more recent years the C.M.I. have created projection models (CMI2009, CMI2010 and CMI2011)[1] which enable the actuary to create their own mortality forecasts by ad-

[1]These can be found in working papers 41, 49 and 55 or at `http://www.actuaries.org.uk/research-and-resources/pages/continuous-mortality-investigation-mortality-projections`.

justing some of the key parameters in the model. The structure of the model is a blend between short term current and immediate past mortality improvement rates and the expert judgement of the long term mortality improvement rate. In this way the model allows the actuary to retain some autonomy with regards to their beliefs about the long term rate of improvement in mortality that the actuary might expect. The model currently does not provide the facility to input any factors relating to socio-economic or epidemiological aspects and more importantly it does not have the facility to provide fan charts of mortality projection uncertainty.

The mortality market is currently in a phase of rapid change as, on the one hand, more parties work to innovate around the management of mortality and longevity risk, and on the other defined benefit pension schemes close and move their liabilities to life insurance companies amongst others. This is leading to a need to develop more accurate longer forecast mortality models as parties begin to trade on the price of mortality itself. The closure of many defined benefit schemes also provides an opportunity to improve the efficiency with which pension scheme valuations are carried out. The reduced data administration requirements with closed schemes should allow firms to better automate many of the processes involved in scheme valuation.

The chapters presented in this thesis contribute to the field by providing evidence of the importance of socio-economic and epidemiological variables and by providing a framework for the incorporation of them in a projection model of mortality. All the models presented in the thesis have the ability to provide forecasts which incorporate levels of uncertainty which can then also be communicated to the client. The methods used are more innovative than current C.M.I. research by incorporating explanatory variables and by demonstrating superior longer term forecasts, an important consideration for mortality risk.

The models developed in this thesis are currently being applied to a stochastic pen-

sion valuation tool that is being developed jointly between myself and a local actuarial employer to be able to provide an integrated data management and stochastic valuation platform which has the potential to redefine the pension scheme valuation landscape. The tool when developed will provide immediate, stochastic simulation results of pension scheme liabilities using stochastic mortality modelling combined with stochastic economic scenario generation to provide a distribution of funding levels with associated confidence from an online platform. this is one example of the application of the work in this thesis.

Bibliography

[1] Andrews D.W.K. (1993), Tests for parameter instability and structural change with unknown change point, *Econometrica*, 61, 821-856.

[2] Andrews D.W.K., Ploberger W. (1994), Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica*, 62, 1383-1414.

[3] Auster, R., Leveson, I. & Sarachek, D. (1969), The Production of Health, an Exploratory Study, *The Journal of Human Resources* vol. 4, no. 4, pp. pp. 411-436.

[4] Bai, J. & Ng, S. (2006), Evaluating latent and observed factors in macroeconomics and finance, *Journal of Econometrics* vol. 131, no. 1-2, pp. 507-537.

[5] Bai, J. & Ng, S. (2002), Determining the Number of Factors in Approximate Factor Models, *Econometrica* vol. 70, no. 1, pp. 191-221.

[6] Bai, J. & Ng, S. (2007), Determining the Number of Primitive Shocks in Factor Models, *Journal of Business and Economic Statistics* vol. 25, no. 1, pp. 52-60.

[7] Bai, J. (1994), Least Squares Estimation of a Shift in Linear Processes, *Journal of Time Series Analysis*, 15, 453-472.

[8] Bai, J. (1997a), Estimating Multiple Breaks One at a Time, *Econometric Theory*, 13, 315-352.

[9] Bai, J. (1997b), Estimation of a Change Point in Multiple Regression Models, *Review of Economics and Statistics*, 79, 551-563.

[10] Bai, J., Perron, P., (1998), Estimating and Testing Linear Models With Multiple Structural Changes, *Econometrica*, 66, 47-78.

[11] Bai, J., Perron, P., (2003), Computation and Analysis of Multiple Structural Change Models, *Journal of Applied Econometrics*, 18, 1-22.

[12] Banerjee, S., and Carlin, B. P. (2002) Spatial Semiparametric Proportional Hazards Models for Analyzing Infant Mortality Rates in Minnesota Counties, *Case Studies in Bayesian Statistics*, Volume VI, New York: Springer.

[13] Banerjee, S., and Carlin, B. P. (2003) Semiparametric Spatio-Temporal Frailty Modeling, *Environmetrics*, 14, 523-535.

[14] Banerjee, S., Wall, M. M., and Carlin, B. P. (2003) Frailty Modeling for Spatially Correlated Survival Data, with Application to Infant Mortality in Minnesota. *Biostatistics*, 4, 123-142.

[15] Barker, D.J. (1992), Fetal and infant origins of adult disease *British Medical Journal*.

[16] Barhoumi, K., Darn, O. and Ferrara, L. (2010), Are disaggregate data useful for factor analysis in forecasting French GDP?. *Journal of Forecasting*, 29, 132144

[17] Barrieu,P., Bensusan,H., El karoui, N., Hillairet, C., Loisel, S., Ravanelli, C., and Salhi, Y., (2010) Understanding, modelling and managing longevity risk: key issues and main challenges, *Scandinavian Actuarial Journal*, Forthcoming

[18] Bernardinelli, L., and Montomoli, C. (1992) Empirical Bayes versus Fully Bayesian Analysis of Geographical Variation in Disease Risk. *Statistics in Medicine*, 11, 983-1007.

[19] Besag, J., York, J. C., and Mollie, A. (1991) Bayesian Image Restoration, with two Applications in Spatial Statistics. *Annals of the Institute of Statistical Mathematics*, 43, 1-20.

[20] Blake, D. and Burrow, W. (2001), Survivor Bonds: Helping to Hedge Mortality Risk, *Journal of Risk and Insurance*, vol. 68, no. 2, pp. 339-348.

[21] Blake, D., Courbage, C., MacMinn, R., and Sherris, M., (2011), Longevity Risk and Capital Markets: The 2010-2011 Update, *The Geneva Papers* no. 36, pp 489-500.

[22] Bell, W. (1997), Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates, *Journal of Official Statistics*, vol. 13, pp. 279-303.

[23] Bernanke, B.S., Boivin, J. & Eliasz, P. (2005), Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach, *The Quarterly Journal of Economics*, vol. 120, no. 1, pp. 387-422.

[24] Bethune, A. (1997), Unemployment and mortality, *Health inequalities*, eds. F. Drever & M. Whitehead, HMSO, London.

[25] Breeden, D.T., Gibbons, M.R. & Litzenberger, R.H. (1989), Empirical Test of the Consumption-Oriented CAPM, *The Journal of Finance*, vol. 44, no. 2, pp. 231-262.

[26] Breitung, J. & Eickmeier, S. (2006), Dynamic factor models, *Allgemeines Statistisches Archiv* vol. 90, no. 1, pp. 27-42.

[27] Booth, H., & Tickle, L., (2008), Mortality modeling and forecasting: A review of methods, *The Australian Demographic & Social Research Institute.*

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

[29] Brouhns, N., Denuit, M., and Vermunt, J.K., (2002). A Poisson log-bilinear approach to the construction of projected lifetables. *Insurance: Mathematics and Economics* vol. 31 no. 3, pp. 373-393.

[30] Brown R.L., Durbin J., Evans J.M. (1975), Techniques for testing constancy of regression relationships over time, *Journal of the Royal Statistical Society, Series B*, 37, 149-163.

[31] Cairns, A.J.G., (2000), A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics* 27(3), 313-330.

[32] Cairns, A.J.G., Blake, D., and Dowd, K., (2006), A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance* 73, 687-718.

[33] 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), 1-35.

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

[35] Carter, L. R., & Prskawetz., A. (2001) Examining Structural Shifts in Mortality Using the Lee-Carter Method. *Working Paper WP 2001-007*, Max Planck Institute for Demographic Research

[36] Chamberlain, G. and Rothschild, M. (1983), Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets, *Econometrica*, vol. 51, no. 5, pp. 1281-1304.

[37] Chu C.S., Hornik K., Kuan C.M. (1995), MOSUM tests for parameter constancy, *Biometrika*, 82, 603-617.

[38] Clayton, D. & Schifflers, E., (1987). Models for temporal variation in cancer rates. II: Age-period-cohort models. *Statistics in Medicine* 6, 469- 481.

[39] CMI, (1990). *Continuous Mortality Investigation Reports,1990*. Institute of Actuaries and Faculty of Actuaries.

[40] CMI, (1999). *Continuous Mortality Investigation Reports,1999*. Institute of Actuaries and Faculty of Actuaries.

[41] CMI, (2002). *Continuous Mortality Investigation Reports,2002*. Institute of Actuaries and Faculty of Actuaries.

[42] CMI, (2005). *Continuous Mortality Investigation Reports,2005*. Institute of Actuaries and Faculty of Actuaries.

[43] CMI, (2006). Stochastic projection methodologies: Further progress and P-Spline model features, example results and implications. *Working Paper 20*.

[44] CMI, (2007). Stochastic projection methodologies: Lee- Carter model features,example results and implications. *Working Paper 25*.

[45] Connor, G.,Hagmann, M. & Linton, O. (2011), Efficient Semiparametric Estimation of the Fama-French Model and Extensions, forthcoming in *Econometrica*.

[46] Cochrane, J. H. (1988) How big is the random walk in GNP? *Journal of Political Economy*, 96, 893-920.

[47] Coelho, E. and 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, 713736

[48] Coughlan, G.D., Epstein, D., Ong, A., Sinha, A., Balevich, I., Hevia Portocarrera, J., Gingrich, E.,Khalaf-Allah, M., and Joseph, P., (2007), LifeMetrics: A toolkit for measuring and managing longevity and mortality risks. *Technical Document (JPMorgan, London, 13 March). Available at www.lifemetrics.com.*

[49] Cressie, N. A. (1993) Statistics for Spatial Data. New York: Wiley.

[50] Currie, I.D., Durban, M., and Eilers, P.H.C., (2004). Smoothing and forecasting mortality rates. *Statistical Modelling* 4, 279-298.

[51] Currie, I.D., (2006) Smoothing and forecasting mortality rates with P-splines.*Presentation to the Institute of Actuaries. Available at: http://www.ma.hw.ac.uk/ iain/research.talks.html.*

[52] Currie, I.D., (2011) Modelling and forecasting the mortality of the very old. *ASTIN Bulletin*, 41, 419-427

[53] Cutler, D.,M., Glaeser, E.L. & Rosen, A.B. (2009), Is the U.S. Population Behaving Healthier? *Social Security Policy in a Changing Environment, eds. J.R.*

Brown, J.B. Liebman & D.A. Wise, National Bureau of Economic Research, University of Chicago Press pp. 423-442.

[54] Cutler, D.,M. & Meara, E. (2000), The Technology of Birth: Is It Worth It?, *Forum for Health Economics & Policy*. Vol. 3, no. (Frontiers in Health Policy Research), Article 3.

[55] Cutler, D.M. & Meara, E. (2004), Changes in the Age Distribution of Mortality over the Twentieth Century. *Perspectives on the Economics of Aging* University of Chicago Press, , pp. 333-366.

[56] Cutler, D., Deaton, A. & Lleras-Muney, A. (2006), The Determinants of Mortality, *The Journal of Economic Perspectives* vol. 20, no. 3, pp. 97-120.

[57] D'Agostino, Antonello & Giannone, Domenico, (2007), Comparing Alternative Predictors Based on Large-Panel Factor Models, *CEPR Discussion Papers 6564, C.E.P.R. Discussion Papers*.

[58] Delwarde, A., Denuit, M., and Eilers, P., (2007), Smoothing the Lee & Carter and Poisson log-bilinear models for mortality forecasting: A penalized log-likelihood approach. *Statistical Modelling* 7, 29-48.

[59] De Jong, P., and Tickle, L., (2006), Extending the Lee & Carter model of mortality projection. *Mathematical Population Studies* 13, 1-18.

[60] Dickson, D.C., Hardy, M.R., Waters, H.R., (2009), Actuarial Mathematics for Life Contingent Risks, *Cambridge University Press*.

[61] Doksum, K. A., and Gasko, M. (1990) On a Correspondence Between Models in Binary Regression Analysis and in Survival Analysis. *International Statistical Review*, 58, 243-252.

[62] Doornik, J. A. and Hansen, H. (1994) An omnibus test for univariate and multivariate normality. *Working Paper, Nuffield College, Oxford*

[63] Dowd, K., Blake, D., Cairns, A.J.G., Dawson, P., (2006), Survivor Swaps, *Journal of Risk and Insurance*, vol. 73, no. 1, pp. 117.

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

[65] Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M., (2010), Backtesting stochastic mortality models: An ex-post evaluation of multi-period-ahead density forecasts, *North American Actuarial Journal* (14)3, 281-298.

[66] Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M., (2010a). Evaluating the Goodness of fit of stochastic mortality models *Insurance Mathematics and Economics* 47, 255-265.

[67] Eberly, L. E., and Carlin, B. P. (2000) Identifiability and Convergence Issues for Markov Chain Monte Carlo Fitting of Spatial Models. *Statistics in Medicine*, 19, 2279-2294.

[68] Ecker, M. D., and Gelfand, A. E. (1999) Bayesian Modeling and Inference for Geometrically Anisotropic Spatial Data. *Mathematical Geology*, 31, 67-83.

[69] Fupuy-Wong, C., Haberman, S., (2004), Projecting Mortality Trends: Recent Developments in the UK and the US, *North American Actuarial Journal*, 8(2), p.56-83

[70] Freeman, D.G. (2003), Is health care a necessity or a luxury? Pooled estimates of income elasticity from US state-level data, *Applied Economics* vol. 35, no. 5, pp. 495-502.

[71] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005), The Generalized Dynamic Factor Model, *Journal of the American Statistical Association*, vol. 100, no. 471, pp. 830-840.

[72] Geweke, J. (1977), The Dynamic Factor Analysis of Economic Time Series, in *Latent Variables in Socio-Economic Models, eds. D. J. Aigner and A. S. Goldberger, Amsterdam: North-Holland*, pp. 365-383.

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

[74] Girosi, F., and King, G., (2005). *A reassessment of the Lee-Carter mortality forecasting method*, Working Paper, Harvard University.

[75] Governemnt Actuary's Department (1995). National population projections 1992-based. H.M.S.O., London.

[76] Governemnt Actuary's Department (2001). National population projections: review of methodology for projecting mortality. London.

[77] Governemnt Actuary's Department (2002). National population projections 2000-based. H.M.S.O., London.

[78] Grubesic, T. H. (2004) The Geodemographic Correlates of Broadband Access and Availability in the United States. *Telematics and Informatics*, 21, 335-358.

[79] Hendry, D. F. and H.-M. Krolzig (2001). Automatic Econometric Model Selection using PcGets. London: Timberlake Consultants Pres

[80] Hendry, D. F. (2000) Epilogue: the success of general-to-specific model selection, in D.F. Hendry, Econometrics: Alchemy or Science? (New Edition). *Oxford. Oxford University Press*

[81] Hári, N., Waegenaere, A., Melenberg, B. and Nijman, T., (2008). Estimating the term structure of mortality. *Insurance: Mathematics and Economics* 42, 492-504.

[82] Hamilton, J.D., (1994), Times Series Analysis, *Cambridge University Press.*

[83] Harris, D., Harvey, D. I., Leybourne, S. J. and Taylor, A. M. R. (2009) Testing for a unit-root in the presence of a possible break in trend. *Econmetric Theory*, 25, 15451588.

[84] Harvey, D. I., Leybourne, S. J. and Taylor, A. M. R. (2009) Simple, robust and powerful tests of the changing trend hypothesis. *Econmetric Theory*, 25, 9951029.

[85] Hauser, R.M., and Weir,D.R., (2011). Recent Developments in Longitudinal Studies of Aging. *Demography*.

[86] Heligman. L., & Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107(1, No 434), 49-80.

[87] Hendry, D.F. and Neale, A. J. (1991) A Monte Carlo study of the effects of structural changes on tests for unit-roots. *Economic Structural Change, Analysis, and Forecasting (eds P. Hackl and A. H. Westlund)*, pp. 95119. New York: Springer.

[88] Hendry, D.F. (2000) Epilogue: the success of general-to-specific model selection, in D.F. Hendry, *Econometrics: Alchemy or Science? (New Edition). Oxford: Oxford University Press*.

[89] Herskind, A., McGue, M., Holm, N., Srensen, T., Harvald, B. & Vaupel, J. (1996), The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900, *Human genetics*, vol. 97, no. 3, pp. 319-323.

[90] Human Mortality Database (HMD), (2004). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research, Rostock (Germany). http://www.mortality.org or http://www.humanmortality.de.

[91] Hyndman, R.J. & Ullah, S. (2005), Robust forecasting of mortality and fertility rates: A functional data approach *(working paper), Department of Economics and Business Statistics, Monash University, Melbourne.*

[92] Iversen, L., Andersen, O., Andersen, P.K., Christoffersen, K. & Keiding, N. (1987), Unemployment and mortality in Denmark, 1970-80, *British Medical Journal (Clinical research ed.)*, vol. 295, no. 6603, pp. 879-884.

[93] Ingram, D. D., and Kleinman, J. C. (1989) Empirical Comparisons of Proportional Hazards and Logistic Regression Models. *Statistics in Medicine*, 8, 525-538.

[94] J., T. (2002), Estimating a health production function for the US: some new evidence, *Applied Economics* vol. 34, no. 1, pp. 59-62.

[95] Jolliffe, I. T. (1986), Principal component analysis, *New York: Springer, University of Geneva*.

[96] Kannisto, V., Lauristen, J., Thatcher, A.R., & Vaupel. J.W., (1994) Reduction in Mortality at Advanced Ages: Several Decades of Evidence from 27 Countries. *Population Development Review* 20: 793810.

[97] Kuan C.M., & Hornik K. (1995), The generalized fluctuation test: A unifying view, *Econometric Reviews*, 14, 135 - 161.

[98] Kazembe, L. N. (2007) Spatial Modelling and Risk Factors of Malaria Incidence in Northern Malawi. *Acta Tropica*, 102, 126-137.

[99] King, G., and Soneji, S., (2011) The Future of Death in America. *Demographic Research* 25, 1-38.

[100] Koissi, M.C., Shapiro, A.F., & Hognas, G. (2005) Evaluating and Extending the Lee-Carter Model for Mortality Forecasting: Bootstrap Confidence Interval. *Insurance: Mathematics and Economics* 38: 1-20.

[101] Kuntsche, E., Rehm, J. & Gmel, G. (2004), Characteristics of binge drinkers in Europe, *Social science & medicine*, vol. 59, no. 1, pp. 113-127.

[102] Kuulasmaa, K., Tunstall-Pedoe, H., Dobson, A., Fortmann, S., Sans, S., Tolonen, H., Evans, A. & Ferrario, M. (2000), Estimation of contribution of changes in classic risk factors to trends in coronary-event rates across the WHO MONICA Project populations, *The Lancet* vol. 355, no. 9205, pp. 675-687.

[103] Lee, R.D. & Carter, L.R. (1992), Modeling and Forecasting U. S. Mortality, *Journal of the American Statistical Association*, vol. 87, no. 419, pp. 659-671.

[104] Lee, R. D., and Miller, T., (2001) Evaluating the performance of the Lee-Carter method for forecasting mortality, *Demography* 38, 537-549.

[105] Leon, D.A. (2011), Trends in European life expectancy: a salutary view. *International Journal of Epidemiology*, vol. 40, no.2, pp. 271-277.

[106] Li, J.S.H., Chan, W.S. & Cheung, S.H. (2011) Structural Changes in the Lee-Carter Mortality Indexes: Detection and Implications, *North American Actuarial Journal*, vol. 15, 13-31

[107] Lo, A. W., and A. C. MacKinley (1988) Stock prices do not follow random walks: Evidence based on a simple specification test. *Review of Financial Studies* 1: 41-66.

[108] Lo, A. W., and A. C. MacKinley (1989) The size and power of the variance ratio test in finite samples: A monte carlo investigation. *Journal of Econometrics* 40: 203-38.

[109] Murphy, Michael J. (2010) Reexamining the dominance of birth cohort effects on mortality. *Population and development review*, vol. 36, no.2. pp. 365-390.

[110] McGilchrist, C. A., and Aisbett, C. W. (1991) Regression with Frailty in Survival Analysis. *Biometrics*, 47, 461-466.

[111] Miller, R.D.J. & Frech, H.E.I. (2000), Is There a Link Between Pharmaceutical Consumption and Improved Health in OECD Countries?, *PharmacoEconomics*, vol. 18, no. 3, pp. 33-45.

[112] Moscone, F. & Tosetti, E. (2010), Health expenditure and income in the United States, *Health Economics*, vol. 19, pp. 1385-1403.

[113] NIRSA (2010) Northern Ireland Multiple Deprivation Study. `http://www.nisra.gov.uk/deprivation/archive/Updateof2005Measures/NIMDM_2010_Report.pdf`

[114] Ngianga-Bakwin, K., and Madise, N. (2004) The Spatial Epidemiology of Childhood Diseases in Malawi and Zambia. *African Population Stuides*, 19, 191-218.

[115] O'Hare, C., and Li, Y. (2012), Explaining young mortality, *Insurance: Mathematics and Economics*, vol. 50, no. 1, pp. 12-25

[116] Organization for Economic Cooperation and Development (2009), *OECD Health Data 2009: Statistics and Indicators for 30 Countries*, Paris.

[117] Ostro, B.D. (1983), The effects of air pollution on work loss and morbidity, *Journal of Environmental Economics and Management*, vol. 10, no. 4, pp. 371-382.

[118] Plat, R., (2009), On stochastic mortality Modeling. *Insurance: Mathematics and Economics* vol 45 no. 3, pp. 393-404.

[119] Ploberger W., Krmer W. (1992), The CUSUM test with OLS residuals, *Econometrica*, 60, 271-285.

[120] Renshaw, A.E., Haberman, S., (2003), Lee-Carter mortality forecasting with age-specific enhancement.*Insurance: Mathematics and Economics* vol 33, pp. 255-272.

[121] Renshaw, A.E., Haberman, S., (2006) A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics* vol 38, pp. 556-70.

[122] Richards, S., and Jones, G. (2004) Financial Aspects of Longevity Risk. *Staple Inn Actuarial Society, London*.

[123] Richards, S. J. (2008) Applying Survival Models to Pensioner Mortality Data. *Institute of Actuaries Sessional Meeting Paper*.

[124] Rosen, S. (1974) Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *The Journal of Political Economy*, 82, 34-55.

[125] Ruhm, C.J. (2004), Macroeconomic Conditions, Health and Mortality, *National Bureau of Economic Research*.

[126] Sargent, T.J. and Sims, C. (1977), Business Cycle Modeling Without Pretending to Have Too Much a Priori Theory *New methods of Business Cycle Research, ed. C. Sims, Federal Reserve Bank of Minneapolis, Minneapolis*.

[127] Schwartz, J. & Dockery, D.W. (1992), Increased mortality in Philadelphia associated with daily air pollution concentrations, *American Review of Respiratory Disease* vol. 145, pp. 600-604.

[128] Shang, H. L., Booth, H., Hyndman, R. J. (2011). Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. *Demographic Research*, 25, 173-214.

[129] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002) Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society*, 64, 583-639.

[130] Stein, M. L. (1999) Interpolation of Spatial Data: Some Theory for Kriging. *New York: Springer*.

[131] Stewart, S.T., Cutler, D.M. & Rosen, A.B. (2009), Forecasting the Effects of Obesity and Smoking on U.S. Life Expectancy, *New England Journal of Medicine*, vol. 361, no. 23, pp. 2252-2260.

[132] Stock, J.H. & Watson, M.W. (2002), Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business & Economic Statistics* vol. 20, no. 2, pp. pp. 147-162.

[133] Thornton, P. E., B. Law, H. Gholz, K. Clark, E. Falge, and D. Ellsworth, et al., (2002) Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests, *Agricultural and Forest Meteorology* 113 185222

[134] Tulijapurkar, S., (2008), Mortality declines, Longevity risk and Aging. *Asia-Pacific Journal of Risk and Insurance* vol. 3(1), pp. 37-51.

[135] Tuljapurkar, S., and Boe, C. (1998) Mortality Change and Forecasting: How Much and How Little Do We Know? *North American Actuarial Journal*, 2, 13-47.

[136] Turrell, G., Kavanagh, A., and Subramanian, S. V. (2006) Area Variation in Mortality in Tasmania (Australia): the Contributions of Socioeconomic Disadvantage, Social Capital and Geographic Remoteness. *Health and Place*, 12, 291-305.

[137] Unal, B., Critchley, J.A. & Capewell, S. (2004) Explaining the Decline in Coronary Heart Disease Mortality in England and Wales Between 1981 and 2000, *Circulation* vol. 109, no. 9, pp. 1101-1107

[138] Vaupel, J. W., Manton, K. G., and Stallard, E. (1979) The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography*, 16, 439-454.

[139] Vaupel, J.W., and Kistowski, K. G. V. (2005) Broken Limits to Life Expectancy *Oxford Institute of Ageing, Ageing Horizons*, No. 3, pp. 613

[140] Vaupel, J.W., (1997) The Remarkable Improvements in Survival at Older Ages. *Philosophical Transactions of the Royal Society of London, B* 352, 17991804.

[141] Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E. (1997) Hierarchical Spatio-Temporal Mapping of Disease Rates. *Journal of the American Statistical Association*, 92, 607-617.

[142] Waller, L. A., Zhu, L., Gotway, C. A., Gorman, D. M., and Gruenewald, P. J. (2007) Quantifying Geographic Variations in Associations between Alcohol Distribution and Violence: A Comparison of Geographically Weighted Regression and Spatially Varying Coefficient Models. *Stochastic Environmental Research and Risk Assessment*, 21, 573-588.

[143] Weir, D.R., (2010) Grand Challenges for the scientific study of aging. *Available at http://www.aeaweb.org/econwhitepapers/whitepapers/DavidWeir.pdf*

[144] Westerlund, J. (2007), Testing for Error Correction in Panel Data, *Oxford Bulletin of Economics & Statistics*, vol. 69, no. 6, pp. 709-748.

[145] Wienke, A. (2003) Frailty Models. *Working Paper, Max Planck Institute for Demographic Research, Germany*.

[146] Wilkinson, D., Hiller, J., Moss, J., Ryan, P., and Worsley, T. (2000) Mortality Variation Across Australia: Descriptive Data for States and Territories, and Statistical Divisions. *Australian and New Zealand Journal of Publich Health*, 24, 226-233.

[147] Wilkinson, R.G. & Marmot, M.G. (2003), Social determinants of health: the solid facts, *World Health Organization, Regional Office for Europe*.

[148] Yang, S.S., Yue, J.C. & Huang, H. (2010), Modeling longevity risks using a principal component approach: A comparison with existing stochastic mortality models, *Insurance: Mathematics and Economics* vol. 46, no. 1, pp. 254-270.

[149] Zimmerman, D. (1993) Another Look at Anisotropy in Geostatistics. *Mathematical Geology*, 25, 453-470.

[150] Zweifel, P., Breyer, F., Kifmann, M. & Kifmann, M. (2009), *Health Economics*, *Springer*.

[151] Zeileis., A., (2000) p values and alternative boundaries for CUSUM tests. *Working Paper 78, SFB Adaptive Information Systems and Modelling in Economics and Management Science*, `http://www.wu-wien.ac.at/am/wp00.htm#78`.

[152] Zeileis, A., Kleiber C., Kramer W., Hornik K. (2003), Testing and Dating of Structural Changes in Practice, *Computational Statistics and Data Analysis*, 44, 109-123.

[153] Zeileis A. (2005), A Unified Approach to Structural Change Tests Based on ML Scores, F Statistics and OLS Residuals. *Econometric Reviews*, 24, 445466.

[154] Zivot, E., & Andrews, D., (1992), Further Evidence of the Great Crash, the Oil-Price Shock and the Unit-Root Hypothesis. *Journal of Business and Economic Statistics* 10, 251270.