Longevity Risk Modelling with Application to Insurer Longevity Risk Based Capital Stress Margins

Carolyn Ndigwako Njenga

School of Actuarial Studies, Australian School of Business

University of New South Wales

A thesis submitted for the degree of

PhilosophiæDoctor (PhD) in Actuarial Studies

August 2011

Declaration

ORIGINALITY STATEMENT

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Signed ....

Date ....

Abstract

Future mortality rates are uncertain and the risk that estimated mortality rates will be higher than observed rates has negative financial implications for providers of living benefits including life annuities and pensions. This thesis studies time trends and cohort trends in mortality rates to determine the number of factors that drive mortality changes. An econometric analysis of mortality improvements is used to give a clearer picture of the stochastic nature of mortality rates in a lower dimensional data space as this thesis uses cointegration analysis for dimension reduction. A multi-country analysis of standardized mortality rates finds evidence of stochastic trends and a significant number of common factors. However, no evidence of common stochastic trends is found. An analysis of Australian mortality rates establishes there are non-stationary and stationary mortality rates by age. The common stochastic trends across age-groups which are exhibited within the Australian data lead to the characterization of mortality rates using a stochastic trend model. Dimension reduction is performed using the Heligman and Pollard (1980) parametric mortality model. The trends in the data are reflected using flexible Vector Autoregressive (VAR) models allowing for correlation between the estimated Heligman and Pollard model parameters. Bayesian Vector Autoregressive (BVAR) models which additionally quantify parameter risk are shown to significantly improve the forecast accuracy when fitting the developed HP-BVAR model to data from 1946-1995 and then comparing its out-of-sample forecasts to observed data from 1996-2007 for Australian mortality rates. Allowing for parameter uncertainty shows it to be a significant component of total risk since the results are realistic probabilistic forecasts. The HP-BVAR model is applied to the calibration of the longevity stress margin of the life insurance capital charge. The structure and magnitude of the current simplification by APRA result in a longevity stress margin that is found to be too prudent and too generalised. An alternative age-dependent simplification is proposed.

Acknowledgements

The completion of my doctoral thesis is without a doubt the most rewarding achievement in my professional life so far. All the glory belongs to God. The advice and guidance that I have received from several people during the course of my PhD has been invaluable, and the first person I must thank is my supervisor, Prof. Michael Sherris, to whom I am indebted. Mike taught me the importance of detail and attention to detail. I thank you for encouraging me to present my work in a variety of forums and helping me to grow into a researcher with a voice.

I sincerely thank the academic staff of the School of Actuarial Studies at UNSW for their comments and suggestions. While undertaking my research I benefitted greatly from Bindya's assistance and that must be credited. Bindya, thank you for not only helping with administrative and technology issues but having a warm smile that made me smile as well and made many days seem much better.

I thank the Australian School of Business, University of New South Wales for providing me with the resources necessary to carry out my research. Without the International Postgraduate Research Scholarship I would never have been able to attend the University of New South Wales.

I thank colleagues and good friends who I met as a PhD student - most of all Séverine Gaille whose visit inspired several intellectual discussions; Katja Hanewald for her helpful insight as I wrote my thesis; and Amandha, Brian, Craig, Jack and Jimmy who all made UNSW an interesting place. I also thank all my friends (especially all my on-call babysitters) from my fellowship group for their prayers, support and encouragement.

Finally, I take this opportunity to express my grateful acknowledgment to my family for their emotional support over the past few years. This thesis is dedicated to them. Most of the time that I spent on this thesis, belongs to my dearest husband, Njenga, who lived through this PhD with me, and my wonderful daughter, Neema, who joined us during the course of the PhD and made life all the more interesting. My deepest gratitude to them for their patience, encouragement, support and caring every step of the way. I am also grateful to all my parents - regardless of whether by blood or by law - for their never ending prayers and encouragement, and to my mummy as she fights the greatest fight of her life - I wish I was there in person for you as you have always been for me.

Regardless of all the above support any errors and/or omissions are solely my own.

Contents

List of Figures

List of Tables

1

Introduction

Longevity risk, the risk that experienced mortality rates are consistently lower than expected mortality rates, has been recognized as one of the significant risks impacting on the financing of ageing populations around the world. It has implications for insurance companies issuing life annuity and other products providing longevity insurance, pension funds as well as governments with social security pension obligations. Mortality rates have been changing globally at different rates and in different directions for different age groups although overall there has been a general decline in mortality rates. This is a problem because there is dependence across age in a country. Therefore, longevity risk is a systematic risk as it cannot be pooled and diversified away. Measuring the uncertainty in changes in levels, trends and volatility of mortality rates, and, specifically, capturing and quantifying longevity risk is challenging but vital for capital planning.

Actuaries and demographers typically assume that the past observed rates can be used to extrapolate future mortality rates. This assumption is challenged if past declines or past increases in mortality rates are not sustained in the future leading to a significant difference between expected and experienced mortality rates. This uncertainty is a significant risk in the life underwriting process. The financial stability of insurance companies that offer annuity products, pension funds and government sponsored social security schemes is reflected by the amount of capital used to protect against uncertainty. There is increased uncertainty about longevity risk and the adequate amount of capital to hold in case of adverse experience. This is because a serious consequence of people living longer than anticipated is that they draw on funds for longer than expected. In fact, deviations of future mortality projections from expected values can be financially devastating. Evidence of this is the world's oldest life insurance company, Equitable Life Assurance Society in the

United Kingdom, which became insolvent not due to the World Wars and the Great Depression but due to unprecedented improvements in mortality[1] (Lord Penrose, 2004).

In light of the above, three problems that are tackled in this thesis present themselves. First, it is necessary to study and understand the historical features of mortality levels, trends and volatilities. This thesis carries out this task on a broad international scale using five countries with comparable living standards from different parts of the world. In order to get a clearer picture it is necessary to adjust the focus of the lens through which mortality is viewed. A variety of dimension reduction techniques are used to adjust the viewing conditions and bring out information about different features of the mortality rates and their changes. A main shortcoming of most of the existing literature is that the observable forces (principal components) and the unobservable forces (factors) that underlie mortality are not clearly distinguished from each other. This ambiguity is often found in both the analysis of past mortality patterns and the subsequently developed mortality models. This thesis focuses on analysing and quantifying trends and volatility in mortality rates while allowing for systematic risk due to the dependence across ages. Various techniques are used to bring out the unobservable and observable variables that explain mortality change. Recent literature has shown that although mortality trends cannot be predicted well, uncertainty in the mortality trends can be quantified based on past observations. Therefore, findings in the first study are used as the basis for a model to quantify the uncertainty in mortality trends in the second study.

Secondly, a mortality model that quantifies longevity uncertainty without neglecting the effects of its components, especially parameter risk, needs to be estimated. There exist several mortality models that attempt to measure the uncertainty in mortality improvements but their performance is limited by various problems such as interdependency in their parameters which make it difficult to capture mortality uncertainty. The solutions to some of the limitations have been recently developed or can be found by innovative interdisciplinary collaborations. Shortcomings of existing mortality models need to be identified and solutions to these problems solved by transferable skills from other fields that are used to solve similar problems. A deep understanding of the features of past mortality patterns will be useful to identify similar problems that challenged model developers in other disciplines and aid in selecting the most relevant tool to borrow. In the second study

[1] Lord Penrose (2004) also specifies other reasons for the insolvency of Equitable Life but longevity risk was a significant factor

of this thesis, a set of econometrics techniques are innovatively transferred into demography and used to account for the interdependencies between the parameters of an analytic mortality law. This thesis transforms a static parametric model into a dynamic parametric mortality model.

Finally, the quantification of capital to adequately reflect the underlying risks is currently under review in Australia. APRA's on-going review of capital standards aims to improve the risk sensitivity and appropriateness of the capital standards. To date, there has been no analysis to check if the APRA-specified longevity stress margin meets this aim. The proposed longevity risk based capital stress margins for insurers have the same structure as Solvency II's but the magnitude is different. Therefore, APRA's specified margin is certain to be subject to several of the same shortcomings of the Solvency II stress margin. It is necessary to analyse the recently proposed method for determining the risk-based capital requirements for longevity risk. An analysis of this type requires a mortality model that comprehensively quantifies longevity risk. The suitability of the simplification that a constant decline in mortality rates across all ages is equivalent to the 99.5% VaR for immediate and deferred life annuity products due to the structure and magnitude of the longevity stress are analysed.

1.1 Motivation and Structure of the Thesis

There is increasing recognition of the need for actuaries and demographers to work together with researchers and practitioners in other disciplines to effectively solve problems such as mortality modelling. For example, mortality models have become more sophisticated by using financial models and exploiting the similarities between mortality rates and financial quantities such as interest rates (see Cairns et al., 2006a, for more details). In the same spirit, actuaries and demographers can collaborate with econometricians to improve existing mortality models. However, only a limited number of investigations into how econometric techniques can be used to enhance mortality modelling have been published. The field of econometrics has several tools to offer that will be used in the analysis and modelling of mortality rates in this thesis. For example, in econometrics, cointegration is used to explain the variability in data by using common trends if they exist.

In this thesis an assessment of longevity risk using econometric analysis of age-specific death rates in a number of developed countries including Japan and Australia from the Asia-Pacific region is provided. This is a vital step in setting up the set of assumptions that govern a mortality model. The analysis considers the statistical evidence in the data in order to determine if models for longevity risk should assume trend or difference stationary processes and to understand if these vary across country and across age. It also provides estimates of the number of factors driving the changes in the age-specific death rates for different countries and considers the extent to which these factors are common across the countries.

In order to quantify the three aspects that measure mortality improvement (level, trend and volatility) stochastic mortality models are necessary because mortality improvements are stochastic (Cairns et al., 2006a). The process of developing a model that quantifies the uncertainty in longevity begins with an analysis of historical mortality trends and changes in these historical mortality trends. In this thesis two general multivariate statistical tools, factor analysis and principal components analysis[1], are used to reduce the number of variables that describe the changes in mortality rates. A parsimonious model that is based on a parametric mortality model is developed whereby the parameters of a parametric mortality model are projected using econometric models in an innovative way. The correlation between the parameters of the parametric model through time is considered which leads to capturing trends from young ages and makes it possible take into account the richer age structure of mortality improvement from young ages to middle and then into older ages by utilizing the dependencies between ages. A range of models is considered by applying econometric techniques to analyse historical mortality rates. In chapter 6 a stochastic mortality model is developed. A Bayesian framework is used to ensure that parameter risk is quantified. Another advantage of the Bayesian method chosen is also reduces the number of parameters in the model.

The developed mortality model is used to investigate the suitability of the equivalence assumptions used to simplify the longevity stress insurance charge in the regulatory framework that the Australian Prudential Regulatory Authority (APRA) proposes to introduce in Australia on 1 January 2013[2].

1.1.1 Aim and Scope

The aim and scope of this thesis are based on the three main research questions that are investigated.

1.1.1.1 Research Question I

How do mortality rates for different countries with comparable standards of living behave in relation to each other? What does this imply

[1]Although principal components analysis is often considered to be a factor analysis, strictly speaking, factor analysis and principal component analysis are separate (this is explained in details in advanced statistics texts such as Härdle and Simar, 2007)

[2]According to a letter dated 16 December 2010 to all CEOs (or equivalent) and Appointed Actuaries of general insurers and life insurers the effective date of the new standards is proposed to be 2013 due to a delay caused by the need for a second Qualitative Impact Study.

for assumptions made while developing mortality models?

The first aim of the thesis is to investigate the development of trends in mortality rates across different countries. The purpose is to provide an analysis of trends and volatility of the historical longevity data across ages and for a number of countries expected to have experienced similar longevity improvements.

It is common to model mortality by projecting factors that affect mortality. Different models include different numbers of factors and in different ways. This poses interesting questions including: How many factors and how many Principal components are necessary to model a population's mortality? Are these factors and principal components comparable across different countries? Do mortality trends exhibit common stochastic trends?

Answering these questions will quantify and give an in-depth understanding of mortality trends and volatility. The existence of common factors means that there is systematic risk across countries. Common factors also quantify the potential for diversification. Additionally, gained insight regarding the assumptions made by various models will be achieved. For example, the Lee-Carter model is based on the assumption that one principal component is sufficient to explain the variation in mortality rates. Principal component analysis will test the validity of this assumption for the countries studied. It is of interest to investigate the nature of trends, number of factors driving volatility and apply cointegration models to research in mortality change.

1.1.1.2 Research Question II

Can existing parametric mortality models be extended to capture the effects of common trends in a given population? Can uncertainty from parameter risk be measured as well?

The second aim of this thesis is to develop a dynamic parametric mortality model that is parsimonious and captures longevity risk which is systematic by nature. This aim is based on the assumption that given an existing static parametric mortality model and appropriate econometric models including VAR models, VECM models (and cointegration tests) and Bayesian VAR models the uncertainty in future mortality can be projected. Intuitively, modelling all the parameters of a parametric mortality model simultaneously has the advantage of capturing the interdependencies in the parameters.

The portion of uncertainty in longevity risk that is contributed by uncertainty from parameter estimation (parameter risk) is usually ignored. Parameter risk arises due to the availability of a limited amount of data for use in model estimation.

The uncertainty from the parameters and the models chosen will be estimable because probability distributions are used to explicitly incorporate data and uncertainties in parameter estimation and model choice coherently and transparently. This leads to realistic probabilistic projections.

1.1.1.3 Research Question III

Is the insurance capital regulator's assumption regarding the longevity stress adequate? Does the magnitude and structure of the APRA specified simplification of longevity stress margin lead to over/under-capitalization?

Finally the model that is developed is applied to the new risk-based framework that is being introduced in Australia. The insurance regulator, APRA, has specified an insurance risk charge which has the purpose of covering the risk that experience is worse than the best estimate of various risks including longevity risk (where longevity relates to the mortality of lifetime annuitants). The stress margins are calculated have 99.5% probability of sufficiency over a 12 month period (i.e. a 1 in 200 year event). The insurance risk charge margins for mortality, morbidity and event stress are determined by an actuary but not the longevity stress margin and the expense stress margin which are specified by APRA. In particular, the proposed simplification that a constant 25% decline in mortality at all ages is equivalent to the 1 in 200 year event is tested.

The goal is to quantify uncertainty in the levels, trends and volatility of mortality improvements using a model that measures and presents uncertainty in the data, parameters of the model and the model through probability distributions of the probabilistic forecasts. The quantified uncertainty is applied to insurer longevity risk based capital stress margins. By addressing these research questions using data and interdisciplinary techniques this goal is achieved. Three studies are conducted that are interrelated as each subsequent one builds upon the results of the previous one.

1.2 Thesis Overview and Contributions of the Thesis

This thesis brings together contributions from different areas of research including Actuarial Science, Demography and Econometrics in an innovative way. After analysing mortality rates and developing a mortality model that can accurately project future mortality rates, the model that is developed is applied to regulatory authority capital requirements and the proposed simplifications to calculating capital requirements in a risk-based capital framework.

The research questions were selected in such a way that firstly, an understanding of mortality trends and changes in mortality trends is given. Factor and Principal Components Analyses are conducted on data to check the validity of some assumptions of popular models such as the Lee-Carter model. Further analysis is conducted using econometric techniques such that the existence of common trends is investigated. This understanding leads to generation of a set of assumptions that are used as a basis for formulating a mortality model. Complex Bayesian econometric techniques are used to improve the mortality model by enabling it to explicitly incorporate data and uncertainties in parameter estimation and model choice. The result is projections of future mortality rates and probability distributions which measure the uncertainty in the projections. Finally, a longevity risk stress margin is calibrated from mortality rates projected using the Bayesian-econometric mortality model (which is referred to as the HP-BVAR model in later chapters).

The contributions of this thesis are answers the research questions above from three interrelated studies.

In the first study a historical analysis of multi-country mortality rates is performed. There has been no previous econometric analysis of mortality for a range of countries investigating the nature of trends, number of factors driving volatility as well as applying cointegration models. The purpose of this study is to understand the interplay of variables that influence mortality improvements. Classical statistical techniques are applied and in order to gain a better understanding of the data's trend, level and volatility, econometric techniques are used. Econometric techniques are more advanced because for many years econometricians have developed and used them to solve economic problems which involve analysis of trends, levels and volatility of economic variables. Economic variables and mortality variables (e.g. age-specific death rates) possess many similar characteristics. The method used to answer this question is a transfer of the techniques used to address the problems in economic variables to mortality variables. Additional information that is given by using econometric techniques includes the potential existence common mortality trends. Understanding common stochastic mortality trends is useful for projecting future mortality rates and is useful to reinsurance companies.

In the second study Bayesian econometric techniques are applied to the parameters of a parametric model for Australia. The parametric model reduced the dimensions of the mortality data set. The parameters are highly correlated and this has been a limitation in previous studies. Using Bayesian econometric methods, the problems due to correlation of the parameters are mitigated. Additionally, common stochastic trends in the parameters are used to improve their predictability. The Bayesian analysis is useful for training the model based on recent data and addresses issues like structure breaks in mortality trends. The resulting model gives realistic projections as evidenced by an out-of-sample analysis.

Finally, the third study involves calibration of the longevity stress margin for a portfolio of life annuitants using the Bayesian Mortality model developed in the previous study.

Chapter 5 is the first study and its results based on a multi-country analysis. The econometric analyses of the trends in the data are considered. Following that the results from a principal components analysis (PCA) for the number of factors that affect mortality across age and country are summarised. Cointegration as a modelling approach is reviewed and applied to assess cross country common trends.

Chapter 6 is the second and its results. The conclusions from Chapter 5 are used as a basis for using a parametric model to model Australian mortality. A mortality model that involves using two stages (parametric mortality modelling and econometric modelling of the estimated parameters) is developed. The results of using an econometric model that captures parameter risk are compared to those of using a similar econometric model that does not capture mortality risk. The results are summarised by comparing the projected probability of death in each case to the observed probability of death.

The final study and its results as presented in Chapter 7 is an application of the models developed in the previous chapter are applied to calculating the amount of capital required for a portfolio of annuities (immediate and deferred) under the risk-based framework proposed in APRA (2010b).

2

Review of Prior Research

Introduction

This chapter reviews literature that is currently available on modelling longevity risk. While there exists a broad range of models including actuarial, demography and financial mortality models, a review of actuarial and demography models is presented because they are the most relevant to this research. The chapter is presented as follows, first, the measures of interest in this thesis and their notations are presented in section 2.1. Secondly, an exploration of mortality models is presented. There are several different ways to classify mortality models (Tabeau, 2002; Booth, 2006; Booth and Tickle, 2008). In this chapter mortality models are classified based on their foundations such as principal components analysis. Aspects of the methodology of these models which are extended in this thesis will be considered further in chapter 4. Cointegration analysis as a dimension reduction technique is useful for analysing large data sets therefore some literature on cointegration is reviewed in section 2.3.

2.1 Measure of Interest

By convention, probability that an individual aged x survives to age $x+1$ is denoted as p_x in actuarial texts. In this thesis an additional dimension, time,t, is required as well. Therefore, conventional notation is modified and the probability that an individual aged x at time t survives to time $t+1$ is denoted as $p_{x,t}$). The remaining life expectancy of this individual is a random variable denoted by $T_{x,t}$. The relationship between these quantities is represented mathematically as:

$$p_{x,t} = \mathbb{P}[T_{x,t} \geq 1 | T_{x,t} > 0] \tag{2.1}$$

In this thesis, the measure of longevity of interest is the probability of death,$q_{x,t} = 1 - p_{x,t}$. It is noteworthy that although the title of this thesis suggests that the analysis is of models of longevity (probability of survival), the application of the mortality models developed within this thesis to valuing benefits that are paid out to individuals that are living form the longevity aspect[1].

Denote the true and unobserved probability of death between times t and $t+1$ for individuals aged x as $q_{x,t}^{true}$. Let $m_{x,t}$ denote the observed mortality rates for individuals aged x in year t. It is calculated as $m_{x,t} = \frac{D_{x,t}}{E_{x,t}}$, where $D_{x,t}$ and $E_{x,t}$ is the recorded number of deaths and the recorded number of exposed-to-risk respectively, of individuals aged x in year t. Dependence between lifetimes at a given point in time is induced by their exposure to environmental factors and this leads to correlation of mortality rates at different ages (Denuit and Frostig, 2009). Also, a limited (finite) amount of data that is available for analysis. Parameters are estimated based on this data. The smaller the data size the less the confidence in the parameters estimated. Therefore, let θ_t denote the estimated set of parameters that is used to describe the factors driving mortality at time t.

Denote using μ_x the instantaneous rate (force) of mortality. Assuming that the force of mortality is constant at integer age, x, and time, t, and that the population is stationary the force of mortality, $\mu_{x,t}$ is approximated by the age specific central death rates, $m_{x,t}$ such that for a given set of parameters, θ_t:

$$m_{x,t} \simeq \mu_x(\theta_t) = \mu_{x+\xi,t+\tau} \qquad 0 \leq \xi, \tau < 1$$

The observed probability of death, $q_{x,t}$, is estimated from $m_{x,t}$ by either assuming that exposure is linear in x or assuming that the force of mortality is constant and equal to the observed mortality rates. Suppressing θ_t, under the latter assumption the link between the crude death rates, $m_{x,t}$, and the probability of death, $q_{x,t}$, is $q_{x,t} = 1 - \exp -m_{x,t}$.

2.2 Modelling Mortality Trends

A relationship between mortality rates and time that is approximately log-linear, decreasing mortality improvements according to age and an increasing trend in the relative rate of mortality change over age are three key features of mortality trends projections are emphasized in Wong-Fupuy and Haberman (2004) as they studied recent developments in projecting mortality trends for the United Kingdom and the United States of America. Uncertainty in mortality trends revolves around whether or not these key features will be sustained.

Cairns (2000) pinpoints the principal sources of uncertainty in modelling mor-

[1]Longevity and mortality are to sides of the same coin. Lower mortality implies higher longevity and vice versa.

tality trends. Firstly, any given model is stochastic by nature because it uses a random process to generate future mortality trends. The results are contingent and sensitive to the model chosen. This comprises model risk. Model risk includes the risk that the assumptions that a mortality model is based on are not valid for the data set used. For example, the Lee-Carter model is based on the assumption that the mortality index has an almost constant (linear) decline. Therefore it does not detect or adjust to structural changes in the age pattern of mortality (Carter and Prskawetz, 2001; Li et al., 2011). Parameter risk is the second source of uncertainty. It is the risk that the estimated parameters in the mortality model are erroneous. This is inevitable because the amount of data available is finite. Although there has been little attention given to parameter risk, the impact of including parameter risk in mortality models is demonstrated in Blake et al. (2008). In general, including parameter risk as an additional source of uncertainty results in broader confidence intervals of forecasted mortality rates. In this thesis the issues of model risk and parameter risk as sources of uncertainty in modelling longevity risk are addressed.

Studies (such as Andreev and Vaupel, 2006; Willets, 2004) have shown that an increase in mortality improvement is usually followed by a decrease in mortality improvement because the gains made against a given cause of mortality will eventually stop. The time the gains will stop is uncertain. This implies that the improvements in mortality in one year can be due to a policy change or medical advancement in (a) previous year(s). It is therefore reasonable to assume that with the passage of time there exists a relationship between the factors (that capture the level of mortality at a given time) in a mortality model.

Quantitative research evidence has shown that extrapolative models such as the Lee-Carter model do not convey information about the forces that drive the changing shape of mortality. This major shortcoming is because extrapolative models rely on the observed past trends. Additionally, in Lee-Carter the general level of mortality in the population captured by a single parameter, k_t, is often extrapolated as a linear function. This implies a fixed mortality trend.

This research focuses on a parametric model (analytical law of mortality) that captures the shape of mortality over the whole life span. It is fitted at a sequence of points in time and the resulting parameters are extrapolated to project the changing shape of the mortality curve as time goes by. This mitigates the shortcoming of extrapolative models that they do not convey information about the forces that drive the changing shape of mortality. This research is based on the assumption that the parameters of the analytical law of mortality adequately reflect the forces that shape the mortality curve.

A key problem that has been encountered in past studies is that the parameters are correlated (Hartmann, 1987). Econometric techniques including VAR models and Cointegration Analysis are used to project economic parameters that are often highly correlated such as Gross National Product, Gross National Expenditure, Unemployment Rates and Interest Rates (Brooks, 2008). In this thesis, these econometric techniques are extrapolated and applied to the parameters in a parametric mortality model.

There has been limited previous econometric analysis of mortality for a range of countries investigating the nature of trends, number of factors driving volatility as well as applying cointegration models (Denton et al., 2005; Hanewald, 2010; Sherris and Gaille, 2010b).

A suitable model should capture model uncertainty (that the model assumptions are accurate) and parameter uncertainty (that the model captures uncertainty due to the limited data available) as it quantifies uncertainty in the level, trend and volatility of mortality rates.

"Excessive reliance on expert opinion-present to some extent in all methods-has led to systematic underestimation of mortality improvements" as shown in Wong-Fupuy and Haberman (2004) and an alternative is to let the data "speak" by using Bayesian methods. This research will incorporate Bayesian methods in modelling the uncertainty in mortality trends.

2.2.1 Understanding Mortality Improvements

It is vital to understand a problem in order to develop an effective solution. Understanding and quantifying longevity risk is big problem for entities such as insurance companies as mentioned in 1. This thesis explores mortality trends by investigating, visualising and making inferences about the levels of mortality rates and their transformations. The first analysis in this thesis gives a better understanding of mortality trends and particularly the time trend and cohort trend in mortality rates. The conclusions of this analysis give insight about the validity of assumptions in existing mortality models. The results are also used to make assumptions that form the basis of development of a new a model for quantifying longevity risk.

A deep understanding of mortality trends is obtained by evaluating mortality improvements and the underlying dynamics that drive them (Andreev and Vaupel, 2005). In their research the year on year mortality improvements for age x in a simple form would be:

$$\text{Mortality improvement from year t-1 to year t} = \frac{m_{x,t}}{m_{x,t-1}} \tag{2.2}$$

However, due to the fact that mortality rates at certain ages, especially the elderly, subject to high variation (Shang et al., 2010), it is common practise to analyse the logarithms of mortality rates, $\ln m_{x,t}$, (Lee and Carter, 1992; Shang et al., 2010). It is also advantageous to model the logarithms of mortality rates because it ensures that they remain positive (Booth and Tickle, 2008). Therefore, Andreev and Vaupel (2005) calculate the rate of mortality improvements for age x as:

$$\begin{aligned}\text{Rate of Mortality improvement from year t-1 to year t} &= -\ln \frac{m_{x,t}}{m_{x,t-1}} \\ &= \ln m_{x,t-1} - \ln m_{x,t}\end{aligned} \tag{2.3}$$

Andreev and Vaupel (2005) analyse the surface of the mortality improvements. Mortality trends are studied as part of the process of developing mortality models or to analyse mortality models. A vast array of publications that consider mortality trends for a variety of countries exist (Willets, 2004; Wong-Fupuy and Haberman, 2004; Sherris and Gaille, 2010a,b; Sherris and Njenga, 2009, 2011; Tuljapurkar and Edwards, 2011). This thesis will extend the body of knowledge about the behaviour of mortality trends by determining the number of factors that are required to explain the variation in mortality time trends.

2.2.2 A search for a simple pattern in a complex set

Mortality rates data form a high dimensional data set. In order to obtain a clear picture of the stochastic nature of mortality rates a lower dimensional data space is required. Dimension reduction is done in a variety of ways including factor analysis and principal components analysis. Mortality rates are described as a complex system of many variables that affect the time of death. In any given year, t, the mortality of an individual is affected by observable variables such as age, gender and location and also by unobservable variables.

In this section, principal component analysis (PCA) as a method and how it has been applied in mortality models is reviewed. In particular, the significance of number of principal components that are represented in a mortality model is explained. This gives a background to the analysis of mortality rates that is performed in chapter 5. As a starting point, it is important to distinguish between factor analysis and principal component analysis. Both methods have been used as foundations for development of mortality models. Additionally, mortality models such as the Lee-Carter model have been classified as both a factor model (e.g. in Tabeau, 2002; Booth, 2006) and a principal component model (e.g. in Shang et al., 2010; Girosi and King, 2008; Hyndman and Ullah, 2007; Hyndman and Booth, 2008; Yang et al., 2010).

Principal factor analysis (PCA) is the most common technique of doing factor analysis (Härdle and Simar, 2007; Rachev et al., 2007). However, PCA is in many ways different from factor analysis itself (Härdle and Simar, 2007; Rachev et al., 2007). Both factor analysis and principal component analysis use linear combinations of variables to describe sets of observations of many variables (Härdle and Simar, 2007). If all variances and covariances exist and are finite in a data set, the both factor analysis and PCA can be done. If the error terms in the factor analysis model are found to have the same variance, then the factor analysis and principal component analysis are equivalent and there is no variability due to common factors.

A factor is a single variable that is not directly observable which describes the variations in several observed variables. Factor analysis involves searching for a smaller number of variables (factors) that describe the variation in a large number of observed variables. In particular, factor analysis explains correlation between observable variables through (directly) unobservable factors (Härdle and Simar, 2007). Factor analysis is therefore a useful tool for dimension reduction in data sets such as mortality data where many observed variables can be explained by a few factors (Booth and Tickle, 2008). According to Thurstone (1934) the purpose of factor analysis is parsimony and simplicity in explaining a complex system. The goal of factor analysis is to find the simplest, most parsimonious factor pattern which can account for the $\frac{n(n-1)}{2}$ inter-correlations among n variables (Cureton and D'Agostino, 1993).

A statistical model that attempts to explain a complex phenomenon using a small number of basic factors (causes) is a factor model (Härdle and Simar, 2007). Factor models make estimation possible by reducing the dimensions of the data and finding the "true" causes that drive the data.

Consider, x, a random $p \times 1$ random vector with p observable variables. Assume that x can be written as:

$$x = \mu + \Lambda f + u \tag{2.4}$$

μ is the mean of x and Σ is the covariance matrix of x. The mean centred matrix, X, is therefore:

$$X = x - \mu = \Lambda f + u \tag{2.5}$$

Equation (2.4) is a k-factor model for x if $\Lambda = \{\lambda_{ij}\}$ is a $p \times k$ matrix of constants (factor loadings), f is a random vector of the k underlying common factors and u is a random vector of the p unique factors associated with the original observed variables. In the factor model the variance, Σ, is a sum of a factor covariance matrix (common variation) and an error covariance matrix (unique variation) decomposed as follows:

$$\Sigma = \Lambda\Lambda' + \Psi \tag{2.6}$$

The diagonal of the factor covariance matrix ($\Lambda\Lambda'$) is $h_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2$, $i = 1, \ldots, p$, and represents the communalities (common variation in the factors) in x. The elements of $\Psi = VAR(u)$, ψ_{ii}, are the uniqueness (idiosyncratic variances measuring the variation in x_i that is not shared with the other variables). When the factor model is estimated by maximum likelihood it is possible to perform a hypothesis test with H_0 : The specified number of factors is adequate to explain the model[1]. It is important to note that the solution of equation (2.6) is not unique but a unique solution can be found by obtaining an appropriate rotation such that ΛG is a matrix of rotated loadings. A plot of the loadings gives a visual illustration of which variables are loaded strongly on each factor.

A factor model is a linear regression model defined as:

$$x_i = \mu_i + \sum_{j=1}^{K} \lambda_{ij} f_j + u_i \tag{2.7}$$

where X_i is a set of n random variables, f_j is a set of K common factors, λ_{ij} are the factor loadings representing the influence of factor j on variable i and u_i is the noise in the i-th variable.

PCA does not assume that the data has a strict factor structure while factor analysis does. This is because in factor analysis the covariance matrix is represented as a function of the covariances between factors plus idiosyncratic variances (equation (2.6)). PCA does not have this restriction and for any non-singular covariance matrix, a PCA can be performed as an exact linear transformation of the series.

It must be stressed that in PCA the variables themselves are of interest. The process of PCA involves computing a standardised linear combination (SLC)[2] of the variables and then finding a set of the standardised linear combinations that are orthogonal and explain all the variance of the original data when taken together.

For PCA, equation (2.4) has a covariance vector $\Sigma = \Lambda\Lambda' + \Xi$. $\Lambda = \Gamma'\Sigma\Gamma$ where

[1]A likelihood ratio test is used. It can be improved using Bartlett's Correction (Bartlett, 1954). Details of the hypothesis test are explained in Härdle and Simar (2007).

[2]A standardised linear combination (SLC) of the variables is a weighted average of the variables such that $\sum_{i=1}^{p} \delta_i x_i$ with $\sum_{i=1}^{p} \delta_i^2 = 1$. Härdle and Simar (2007) explains that the SLC of x with the largest variance is the 1^{st} PC. The set of SLC's that are uncorrelated with the 1^{st} PC is determined and the 2^{nd} PC is the SLC with the largest variance from this set.

Γ is orthogonal. The elements in Λ are ordered loadings $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. In this case the error matrix, Ξ, is not diagonal but is small and contains the sums of squares of the elements. PCA measures the lack of fit using the sum-of-squares in u_i. The sums of squares are minimized using either eigen-decomposition of the covariance matrix or singular value decomposition on the mean centred equation (2.5). The principal component transformation from x (the original random vector) to y (the vector of principal components) is:

$$x \rightarrow y = \Gamma'(x - \mu) \tag{2.8}$$

$$y_i = \gamma_1'(x - \mu) \tag{2.9}$$

y_i is the i^{th} principal component of x and γ_i is the i^{th} vector of PC loadings (the i^{th} column of Γ).

In PCA, the more the dimensions are reduced, the less the percentage of variation is explained. However, in factor analysis the exact factor structure of the data is revealed by explaining what factor explains what process.

After principal components analysis, each x_i can be written in terms of principal components as a linear combination of the principal components, e_i, and their loadings, β_{ji}:

$$x_i = \mu + \sum_{j=1}^{K} \gamma_{ij} y_j \tag{2.10}$$

The number of principal components to be retained can be visually determined in a straightforward way from a scree plot using the Cattell scree test (Cattell, 1966) where the scree plot is viewed as an elbow with a drop and a less steep decline when viewed from left to right and all components after the one starting at the elbow are cut off. When a clear visualisation is not possible other methods of determining the number of principal components (or factors) to retain including the Kaiser-Guttman Rule (Kaiser, 1960)[1] and the Variance explained criterion[2] are used (Dunteman, 1989).

2.2.3 Principal Component Analysis in Demography

Principal Components Analysis in demography started notably in the analysis of the dimensions of mortality in Ledermann and Breas (1959) who found that mortality for 1955 United Nations mortality data three factors were sufficient to explain over 90% of the variation in the data. In particular, one factor was interpreted to

[1] It is also known as Kaiser's criterion and states that the number of factors retained is equal to the number of factors with eigenvalues greater than 1

[2] Retain enough principal components to account for a given percentage of variation.

reflect the general change in the level of mortality with age, the second factor interpreted the relationship between mortality of the very old and the very young and the third factor was interpreted to be the changes in mortality of the extremely old. After this, Bozik and Bell (1987) and Bell and Monsell (1991) proceeded to reduce the dimensionality of age-specific fertility and mortality data using principal components analysis. Both these papers find that the approximation error is less than that of its contemporaneous studies such as the mortality investigation in Rogers (1986). Lee and Carter (1992) then introduced the popular method that has formed a cornerstone of modern mortality models (Deaton and Paxson, 2001; Girosi and King, 2007).

2.2.3.1 Models with a foundation in Principal Components Analysis

All the variables (observable and unobservable) that affect age-specific death rates (ASDRs) are the dimensions of the data set of the age-specific death rates. Suppose that there are L principal components. The ASDRs are written as a linear combination of the principal components in order to reduce the dimensions of the data set as follows:

$$\ln m_{x,t} = a_x + b_{x1}k_{1t} + b_{x2}k_{2t} + \cdots + b_{xL}k_{Lt} \tag{2.11}$$

Girosi and King (2007) outlines the loss of information from using only the first principal component and shows that the variation in the mortality rates from the other principal components is ignored. By truncating equation (2.11) from the third term onwards and replacing those terms with an error $e_{x,t}$ this equation results in equation (2.12) below.

The Lee-Carter Model is simple to implement and is defined as:

$$\ln m_{x,t} = a_x^{LC} + b_x^{LC} k_t^{LC} + e_{x,t} \tag{2.12}$$

The Lee-Carter model is over parameterized (Haberman and Renshaw, 2008). In order to make this model identifiable and unique, Lee and Carter (1992) imposed certain restrictions on the parameters a_x^{LC}, b_x^{LC} and k_t^{LC}. The restrictions are

$$\sum_{t=1}^{T} k_t^{LC} = 0 \qquad \sum_{x=0}^{N} b_x^{LC} = 1 a_x^{LC} = E[\ln m_{x,t}]$$

a_x^{LC} is a vector of the mean by age across time of the logarithm of mortality rates and it is independent of time. b_x^{LC} reflects how rapidly the mortality at each age varies with the general level of mortality as measured by κ_t^{LC}. $e_{x,t}$ is the error term (short term fluctuations) assumed to have a Normal Distribution in its original formulation in the classical Lee-Carter Model.

The Lee-Carter model is based on the first principal component extracted by using singular value decomposition (SVD). Lee and Carter (1992) extract a single time varying parameter, k_t^{LC}, which is the level of mortality rates and induces the volatility in the mortality model and drives the mortality process. In practice, the first principal component, k_t^{LC} is modelled as a random walk with drift using:

$$k_t^{LC} = k_{t-1}^{LC} + \delta + \varepsilon_t \tag{2.13}$$

This assumption that the first principal component is sufficient has been discussed in Girosi and King (2008, 2007); Hyndman and Ullah (2007); Booth et al. (2002b).

The Lee-Carter model was specified for all-cause mortality rates in the U.S.A. from 1933 to 1987. It has since been applied to other populations and cause-specific mortality data in studies including Tuljapurkar et al. (2000) and Booth et al. (2002b). The population of the USA is one of the largest in the world - it is the largest in the OECD countries. When applied to smaller populations, the Lee-Carter model is found to perform poorly (Jarner and Kryger, 2009). This is due to the assumption that mortality rates improve at a constant rate over time as the forecasted k_t^{LC} is linear (for details see Girosi and King, 2007). This assumption is valid for most large populations but not for small populations. In particular, k_t^{LC} is linear

If all the principal components are included in the model then no information is lost. In this form, however, equation (2.11) lacks parsimony as some of the principal components contain little information and use up degrees of freedom. The obvious solution is to use only some of the principal components and add an error term, like the Lee-Carter model does. The challenge is to find the optimal number of principal components that are required to give a parsimonious model.

Therefore, before electing to apply the Lee-Carter model one must investigate the data set to confirm that it satisfies the assumption that one principal component is sufficient to explain the variation in the data set. The performance of the Lee-Carter model is attributed to its underlying assumptions. Significantly, as a consequence of the assumption that one principal component is sufficient to explain the variation in mortality rates, the age-specific reaction to shocks, β_x^{LC}, is assumed to be fixed through time (Booth et al., 2002b). Goodness of fit changes with different age ranges. This is because the changes in the pattern of mortality are not reflected well in the Lee-Carter model. A major phenomenon that has affected mortality is the rectangularization of mortality rates.

The volatility in Lee-Carter model is driven by the mortality index - first principal component. It models trend with one random factor k_t^{LC} that drives changes (Booth et al., 2002b) and is often assumed to be a random walk or an $AR(1)$ (Tuljapurkar et al., 2000; Booth et al., 2002b). k_t^{LC} drives the Lee-Carter model. It is assumed that k_t^{LC} is modelled adequately by a univariate time series method - a random walk. The Lee and Miller (2001) fix (based on using a base year of 1950) to the problems in the Lee-Carter model by excluding earlier time periods that have a different structure of mortality patterns does not work for Australian data (Booth et al., 2002b). This may be because the structure of mortality patterns in Australia changed in the 1970s (Booth et al., 2006). This is further explained in 4.1.1.1 and 4.2.2.1.

The Lee-Carter model assumes that the sensitivity of mortality at each age,b_x^{LC}, to changes in the general level of mortality, k_t^{LC}, does not change with time(Booth et al., 2002b). If the rate of decline in a population's mortality during a certain time period is different than the one exhibited in another time period; and if the decline in mortality in different age groups changes in different time periods, the Lee-Carter model's assumption on b_x^{LC} is not valid. Booth et al. (2002b)'s investigation uncovered that there were different patterns of mortality decline for different age groups in developed countries.

The Lee-Carter model is also criticised as lacking age-time interaction terms (Pedroza, 2006; Booth and Tickle, 2008; Haberman and Renshaw, 2011; Dowd et al., 2010; Lee and Miller, 2001; Booth et al., 2002a) since it has a time-invariant age component and no cohort effect. This is important because substantial age-time interactions have been observed (Willets, 2004). However, this is interpretation is arguably distorted as shown in Alai and Sherris (2011) where the period effect with age interaction is seen to be essentially the cohort effect because it confounds the true age and period effect. The age-period interactions can be captured more effectively by including the second and higher principal components of equation (2.11). Further, $e_{x,t}$ do not often have a Normal Distribution (Haberman and Renshaw, 2008).

Several improvements to the model have been proposed such as including more factors, allowing for cohort effects, including more general error distributions and applying more efficient estimation techniques. A number of publications including Shang et al. (2010); Yang et al. (2010) provide a discussion of these model improvements. In this thesis the main interest is the number of principal components that are required. This is important because information is lost in the Lee-Carter model. This is explained in section 2.2.3.1.

In this thesis the inclusion of more than one principal component will be considered. The Hyndman-Ullah functional data approach (Hyndman and Ullah, 2007) is a generalization of the Lee-Carter model primarily by adding sets of (k_t, b_x) components to (2.12) to yield an equation similar to equation (2.11). Hyndman and Booth (2008) use six (6) principal components to model Australian mortality because the find that in order to explain over 90% percent of the variation. These additional five (5) principal components reduce the errors in the forecasted values.

However, a major criticism of the use of principal components methods for the analysis of mortality rates is that principal components are simply linear combinations of all variables (usually age groups) in the data set (Hatzopoulos and Haberman, 2011). If the data is additionally found to be non-stationary and the movements in the data are in the same direction this leads to diminished relevance and interpretability of the principal components because the main variation that is captured by the first principal component for example is simply due to the similar pattern of movement in mortality rates at different ages (Hatzopoulos and Haberman, 2011; Lansangan and Barrios, 2009). Therefore, factor analysis is an alternative that is a useful tool for analysis of mortality trends because rather than explaining the variation by looking for a linear combination of the variables it seeks to explain the variation by seeking hidden unobservable variables. The use of factor analysis in demography will be discussed in the next section.

Further to an analysis of mortality trends, this thesis also goes further to analyse the variation in mortality trends by considering period trends and the cohort trends. The use of PCA for analysis of the variation in mortality rates has not been done before but a study of period trends by considering the splines and the surface of the mortality trends has been done in an analysis of patterns of mortality improvement in Andreev and Vaupel (2005).

2.2.4 Factor Analysis in Demography

The Lee-Carter model has been classified as a two-factor model in Booth (2006) with the factors being age and period effects. Alternatively, it has been classified as a one-factor model in Wolf (2004). Based on the categorization in Booth and Tickle (2008); Tabeau (2002), a factor is viewed as a categorical variable which is intrinsic to the data. The resulting categories of models (Booth and Tickle, 2008) are:

- Zero-factor models: Treating each age independently with no specified underlying model

- One-factor models: Treating mortality rates as a function of age (e. g. Gompertz model, Heligman-Pollard model)
- Two-factor models: Treating mortality rates as a function of age and time (e. g. Lee-Carter Model)
- Three-factor models: Treating mortality rates as a function of age, time and cohort

In chapter 5 the number of principal components required to explain the different percentages of variation in mortality data sets for different countries will be studied. In addition to performing factor analysis and principal components analysis on the levels on the mortality rates, the two analyses are performed on the first differences to give a clearer picture of the factors that underlie and combinations of variables that describe the changes in mortality rates.

A study that is almost similar to that in this thesis is performed in a first difference specification of the Lee-Carter model presented in Wolf (2004). However, that investigation is a model variation that simply amalgamates equations (2.12) and (2.13).

A recent publication on principal components in mortality is O'Hare and French (2011) where rather than adding more cohort and/or period effects or extracting a larger number of principal components they generalise a model by using a dynamic factor approach. Specifically, they generalize the Lee-carter model which they classify as a basic one static factor model. Their approach exploits the dynamics of the data when extracting factors. Dynamic principal components analysis is used to estimate the principal components more efficiently at different lags. This method is an improvement over estimating static principal components over a single period. Interestingly, O'Hare and French (2011) find that a smaller number of dynamic factors (strictly speaking these are principal components) than static factors was required to capture the same amount of variation in the data.

When factor models are considered, one of the most widely discussed factor model is presented in Cairns et al. (2006b). It is a model with two factors, $A_1(t)$ and $A_2(t)$ that are estimated using least squares on mortality rates transformed to a simple (linear) parametric form $\frac{qx}{px} = A_1 + A_2 x + error$. This form is selected from a variety of curves and is chosen because for higher ages it gives a fit that is significantly better. Their justification for the use of two factors is twofold. The first is to get the best fit for ages 60-90; the second is to adequately capture the uncertainty in longevity. The first factor affects mortality-rate dynamics at all ages in the same way (a downward trend interpreted as general improvements in mortality over time at all ages) , whereas the second factor affects mortality-rate dynamics at higher ages much more than at lower ages (it is an increasing trend interpreted as the curve is getting slightly steeper over time: that is, mortality improvements have been greater at lower ages). While this model it attractively simple, it has some limitations in that it is estimated for use based on the United Kingdom population and may not be suitable for use in other populations if they are affected by additional factors.

From the discussion above, it is seen that in most mortality “factor” models, the word “factor” is used ambiguously and does not strictly adhere to the definition of a factor from the statistical analysis sense as being a hidden latent unobservable variable. In this thesis, analyses of factors using this strict definition will be used to search for unobservable variables that drive mortality trends and volatility. Principal components analysis will also be used to search for independent linear combinations of observed variables that explain the variation in mortality trends and volatility.

2.2.5 Parametric Mortality Models

A mathematical expression that describes mortality as a function of age is known as a parametric mortality model (also a mortality law, a parameterized mortality schedule or an analytical law of mortality). The entire age pattern of mortality rates over a lifespan is described in one step by a parametric mortality model. Parametric mortality models enable modellers of mortality rates to easily reduce a large body of data into a few parameters that are easy to interpret and easy to manipulate for analysis (Rogers, 1986; Congdon, 1993). Parametric models are also used for comparisons of mortality rates over time and by region. Further, they produce smooth estimates of the probability of death.

Static parametric mortality models are made dynamic by fitting the parametric model at a series of time points for which data is available to obtain a series of time dependent parameters and then using time series or econometric techniques to quantify and project the parameters.

2.2.5.1 Static Models

The genealogy of parametric mortality models dates back to 1725 with Abraham de Moivre’s publication, ”Annuities upon Lives”, formulating the conditional probability that an individual aged x would survive to at least age $x + t$, subject to some maximum attainable age, ω, as:

$$ {}_tp_x = \frac{\omega - (x+t)}{\omega - x}, \qquad 0 \le t < \omega - x, \tag{2.14} $$

In 1825 another milestone in parametric mortality models was developed by Benjamin Gompertz whereby the conditional probability ${}_tp_x$ was given as:

$$ {}_tp_x = \exp\left[-\int_x^{x+t} BC^y \,\mathrm{d}y\right] \tag{2.15}$$

Several authors developed this model further with extensions in publications including Makeham (1860) and Heligman and Pollard (1980).

With the advent of non-parametric models such as Lee and Carter (1992) research shifted towards non-parametric models. Non-parametric models avoided the problems encountered in some parametric models such as the correlation of parameters. Solving the correlation problem by incorporating the correlation into the formulation of parametric mortality rates models would enhance the significance of parametric mortality models in capturing changing mortality over time. This is done by using econometric models.

Non-parametric models perform best for modelling old ages mortality when they are fitted to ages excluding the young ages. For example, the Lee-Carter model is popular for modelling old age mortality and performs best when estimated using data from middle ages to older ages thereby ignoring the implications of the increased probability at birth of reaching age 65. The changing levels of mortality at different ages affect the mortality rates at subsequent ages and, consequently, ignoring the effect of young age mortality on old age mortality in effect fails to adequately capture the full impact of the rectangularization phenomenon whereby declining mortality rates at lower ages lead to a concentration of deaths at higher ages since more people survive to these higher ages resulting in higher mortality rates (Fries, 1980) and these improvements in mortality rates at lower ages contribute to increasing mortality improvements. Excluding the mortality rates of those age 64 and under when modelling the mortality rates of older people excludes the effects of the improvement in the age structure of mortality that is due to improvement in mortality at younger ages. This reduces the richness of the age structure of mortality and masks information. Therefore, parametric models still have a significant place in modelling mortality trends.

2.2.5.2 Heligman-Pollard Model

Mortality curves should reflect four main features according to Heligman and Pollard (1980); Rogers (1986), a high death rate in the first year of life that declines in childhood then rises in late teenage years to form a hump that is more pronounced in males due to accidents and less pronounced in females due to maternal causes. At old ages the death rate increases. The mortality schedule has this basic fundamental shape that has persisted through time so it is practical to use a parametric model to describe changes in mortality rates (Congdon, 1993).

Let $\theta_t' = (A_t, B_t, C_t, D_t, E_t, F_t, G_t, H_t)$ be a set of parameters at time t. Heligman and Pollard (1980) developed a non-linear model that represents mortality across the entire age range at a given time where the probability of death $q_{x,t}$ at time point t is modelled using the Heligman-Pollard equation (1a):

$$q_x(\theta_t) = q_{x,t} = A_t^{(x+B_t)^{C_t}} + D_t \exp[-E_t(\log\{\frac{x}{F_t}\})^2] + \frac{G_t H_t^x}{1 + G_t H_t^x} \qquad (2.16)$$

The parameters A, B, C, D, G all lie in [0,1], E, F in [0,∞) and H in [0,15].

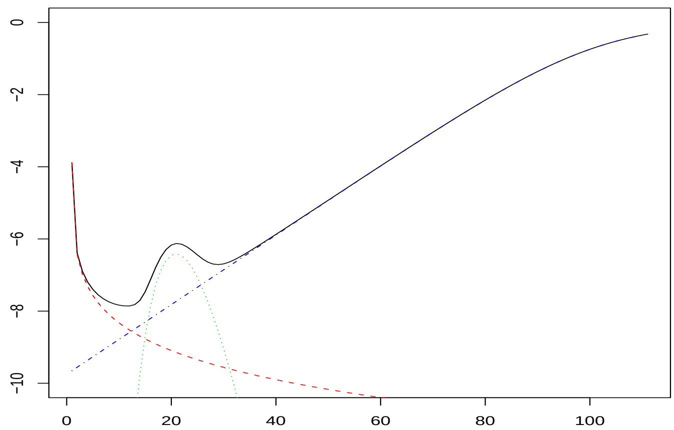


Figure 2.1: Heligman and Pollard Model. Red (broken line) = childhood mortality. Green (dotted line) = adult mortality. Blue (dot-dashed line) = old age mortality.

The parameters of the Heligman-Pollard model are interpreted as factors that affect different ranges of ages (Sherris and Gaille, 2010a). Therefore, from the factor analysis aspect, the Heligman-Pollard model is essentially an 8-factor mortality model.

This model is used to model mortality at a fixed point in time for all ages. Each of the parameters has a demographic interpretation. The first component, $A^{(x+B)^C}$, is a rapidly declining exponential to reflect the fall in mortality during the early childhood years. The middle term in the model reflects accident mortality for males and accident and maternal mortality for females. The third term in the model, GH^x is the Gompertz exponential, and reflects the rise in mortality in adults due to the aging of the body.

Mortality rates are dynamic by nature and are affected by several factors. The effect of each factor at a given time is reflected in the mortality schedule at that time. This is adequately reflected by a parametric mortality law such as the Heligman-Pollard first law of mortality (equation (2.16)).

q_x for x=0,1,...,85 are used to capture the features of the Heligman-Pollard model at a given time, t. The parameters are estimated by minimizing the weighted sum of squared errors between observed, q_x and the fitted, $\hat{q}_x$:

$$S^2 = \sum_{x=0}^{85} \frac{1}{q_x^2}(\hat{q}_x - q_x)^2 \tag{2.17}$$

Consider the first law of mortality from Heligman and Pollard (1980). When the research area of interest is modelling old age mortality, the third term of equation (2.16) is sufficient for estimating $q_{x,t}$ because it is the Gompertz model. In particular:

$$\frac{q_{x,t}}{p_{x,t}} = G_t H_t^x \tag{2.18}$$

Assume that $p_{x,t} \to 1$. The trends in the relative rate of increase in mortality represent the rate at which G is declining (Heligman and Pollard, 1980). As shown in Wong-Fupuy and Haberman (2004) the relative rate of change of mortality rates is:

$$\frac{q'_{x,t}}{q_{x,t}} = \ln(H)$$

A small $\ln(H)$ implies that there has been a small decline in G while a larger $\ln(H)$ implies that there has been a larger decline in G.

Alternatives to the Heligman-Pollard model include a double exponential accident hump in Rogers (1983) function which has nine parameters, Carrière (1992) which has eight parameters or the models in Hannerz (2001a).

The main shortcoming of the Heligman-Pollard model is the correlation of its parameters (Hartmann, 1987) and therefore is often difficult to identify. In some instances the parameters of the Heligman-Pollard model have been criticized as erratic. This is because of numerical instability induced by using weighted squares estimation in S^2. The Heligman-Pollard model assumes that the coefficient of variation is constant across age (Dellaportas et al., 2001) since $\sigma_x^2 \propto q_x^2$ which implies $\sqrt{\frac{\sigma_x^2}{q_x^2}} \simeq \frac{\sigma_x^2}{q_x^2}$. To mitigate this source of instability there has been research into use of Bayesian methods to estimate the parameters (Dellaportas et al., 2001) but for the model that combines the Heligman-Pollard model and Bayesian VAR (HP-BVAR) that is developed later in this thesis it is assumed that the parameter estimates are adequate. Other critics of this model have been that at extremely old ages the model plateaus (Thatcher, 1999) and fails to adequately reflect mortality at those oldest ages. The Heligman-Pollard model also fails to capture multiple mortality humps that are sometimes seen in female probability of death.

2.3 Time Series

2.3.1 The Significance of Stationary Time Series in Models

The stationarity or non-stationarity of a time series is very important in developing an appropriate model. A mortality rate series may have a deterministic trend around which the series fluctuates and the trend-stationary time series has the form

$$y_t = \delta + \phi y_{t-1} + u_t \tag{2.19}$$

with $|\phi| < 1$. After fitting the trend the model errors would then be stationary.

Alternatively the mortality rates may have a stochastic trend and the rate of change in mortality would be stationary with drift or trend. The random walk with drift takes the form

$$y_t = \delta + y_{t-1} + u_t. \tag{2.20}$$

Using the backshift operator $\mathbb{B}$ such that $\mathbb{B}y_t = y_{t-1}$ the difference operator is $\Delta y_t = (1 - \mathbb{B})y_t = y_t - y_{t-1}$. The random walk with drift is written as:

$$y_t = \delta + y_{t-1} + u_t \tag{2.21}$$

$$y_t - y_{t-1} = \delta + y_{t-1} - y_{t-1} + u_t \tag{2.22}$$

$$(1 - \mathbb{B})y_t = \delta + u_t \tag{2.23}$$

Δy_t is a stationary variable and stationarity has been induced by differencing once. The characteristic equation is $(1 - \mathbb{B}) = 0$ and has a root $\mathbb{B} = 1$, hence y_t is referred to as having a unit root.

Differentiating between these two situations is important in fitting mortality trends since the nature of the trends and shocks will have quite different implications for modelling future rates. To illustrate the importance of this consider a stationary $AR(1)$ mean adjusted series y_t as:

$$y_t = \phi y_{t-1} + u_t \tag{2.24}$$

where u_t is a random mean zero shock. This is written in terms of lagged values as:

$$\begin{aligned} y_t &= \phi(\phi y_{t-2} + u_{t-1}) + u_t && (2.25)\\ &= \phi^2 y_{t-2} + \phi u_{t-1} + u_t && (2.26)\\ &= \phi^2(\phi y_{t-3} + u_{t-2}) + \phi u_{t-1} + u_t && (2.27)\\ &= \phi^3 y_{t-3} + \phi^2 u_{t-2} + \phi u_{t-1} + u_t && (2.28) \end{aligned}$$

which becomes

$$y_t = \phi^{T+1} y_{t-(T+1)} + \phi^T u_{t-T} + \phi^{T-1} u_{T-1} + \ldots + \phi^2 u_{t-1} + \phi u_{t-1} + u_t \qquad (2.29)$$

If $\phi < 1$ then as $T \to \infty$ the effect of the past random shocks gradually diminishes since $\lim_{T\to\infty} \phi^T = 0$, which implies y_t is stationary.

If $\phi = 1$ then the series has a unit root and as $T \to \infty$ the effect of the shocks persists since $\lim_{T\to\infty} \phi^T = 1$ and they accumulate as stochastic trends in the series:

$$y_t = y_0 + \sum_{t=0}^{\infty} u_t \qquad (2.30)$$

For the trend stationary model it is necessary to estimate the trend as part of a stationary model using the levels of the series. In the case of the series with the unit root it is necessary to take differences and to model the differences as a stationary series. For the difference stationary series, the series is said to be integrated of order 1 or I(1). Unit root tests are critical in determining the model assumptions. For difference stationary models, shocks to the series have permanent effects and the variance increases with time. With a trend stationary model the shocks around the trend have constant variance and shocks are transitory.

If y_t is a stationary time series then the shocks gradually die away and their effect reduces with time. If y_t is a non-stationary time series then the effect of the shock will persist infinitely.

To determine if the series are trend or difference stationary there are econometric tests that have been developed in the econometric and financial literature since this is a common feature in many economic and financial series.

2.3.1.1 Unit Root Tests

There exist various statistical tests for unit roots tests including the Dickey-Fuller (Dickey and Fuller, 1979) and the Augmented Dickey Fuller Test as well as the Phillips-Perron test (Phillips and Perron, 1988). These tests consider the null hypothesis that the series is non-stationary and require evidence to reject the null hypothesis.

The assumption for the standard Dickey Fuller test is to write the series with a deterministic linear trend as:

$$y_t = \phi y_{t-1} + \delta + \gamma t + u_t \tag{2.31}$$

which after subtracting y_{t-1} from each side becomes:

$$\Delta y_t = (\phi - 1) y_{t-1} + \delta + \gamma t + u_t \tag{2.32}$$

The null hypothesis is that the coefficient on y_{t-1} is zero. If this null is rejected then the series is modelled as stationary but if it is not rejected then the series is modelled as difference stationary. Non-standard test statistics are required under the null hypothesis.

The standard Dickey Fuller test is only valid if the white noise u_t terms are not autocorrelated. This situation is handled in the augmented Dickey Fuller model by including a number of lags, p, for y_t and the model becomes:

$$\Delta y_t = \psi y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta y_{t-i} + u_t \quad , \qquad \psi = \phi - 1 \tag{2.33}$$

where the number of lags, p are usually selected either based on the frequency of the data, where for monthly data 12 lags would be used or for quarterly data 4 lags, or based on an information criterion to select the number of lags that minimizes the value of the information criterion.

It is important to select the number of lags with care since including too few lags will not remove all the autocorrelation while including too many lags reduces the power of the test. The mortality rate time series were found to be sensitive to the lag length.

Phillips and Perron (1988) introduce a test that allows for autocorrelated residuals. In the Dickey Fuller Test u_t are assumed to be independent and identically distributed while in the Phillips-Perron test u_t are assumed to be serially correlated. The Phillips-Perron test is usually more powerful than the Augmented Dickey-Fuller test but it is also more sensitive to miss-specification of the order of the lag of its autoregressive and moving average components.

These tests have been known to have low power if the process is stationary but with root close to 1. For these series it is difficult to determine if they have long run trends or are random walks with stochastic trends. Tests such as the Kwiatkowski et al. (1992) test are then used since they assume stationarity as the null hypothesis ($H_0 : y_t \sim I(0)$) and require evidence of non-stationarity. The joint use of unit root tests and stationarity tests places checks on the standard unit root tests and provides a stronger basis for determining if trends are deterministic or stochastic.

There are four possible outcomes of this analysis using both unit root tests along with stationarity tests:

1. Unit root test - Reject H_0; Stationarity test - Do not reject H_0 (Stationary)
2. Unit root test - Do not reject H_0; Stationarity test - Reject H_0 (Non-stationary)
3. Unit root test - Reject H_0; Stationarity test - Reject H_0 (Inconclusive)
4. Unit root test - Do not reject H_0; Stationarity test - Do not reject H_0 (Inconclusive)

Stationary (I(0)) processes are short memory processes since after long lags observations at different times are independent while Integrated (I(1)) processes are long memory processes since after long lags observations are not independent so differencing data leads to loss of information.

2.3.1.2 Autoregressive Moving Average Models

An autoregressive-moving average model, ARMA(p,q), is defined by a p-th order stochastic difference equation:

$$y_t = \delta + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \tag{2.34}$$

where ϵ_t $WN(0, \sigma^2)$

ARMA(p,q) models are based on the assumption that the process y_t is stationary. If the data is trending then Box and Jenkins (1976) propose differencing the data d times to eliminate the trend and hence suitable for modelling using ARMA(p,q). Such a model is an ARIMA(p,d,q) (Hamilton, 1994).

In McNown and Rogers (1989) a parameterized time series is modelled as described in Thompson et al. (1989) using the techniques of Box and Jenkins (1976). A unit root in the time series of the parameters required differencing the time series to achieve stationarity (McNown and Rogers, 1989). The parameters for an ARIMA models were then extrapolated to obtain a series of Heligman-Pollard curves with time varying parameters. Because of the assumption of independence of the Heligman-Pollard parameters in the process given in Rogers (1986) the forecasts are not accurate and will be inconsistent (Lee, 1992). Lee (1992) also notes the absence of confidence intervals which is attributed to problems that arise from the independence assumption.

Sims (1980) developed a Vector Autoregression(VAR) model with p lags, VAR(p) for expressing a set of variables as a weighted linear combination of each variable's past values and the past values of the other variables in the set. The VAR(p) models are more flexible than AR models and have a rich structure that captures more features of the mortality system. Literature and details on the VAR models are in the following section.

2.3.2 Econometric Dimension Reduction Techniques

Section 2.2.2 explained PCA and FA. Dimension reduction using PCA involves a linear data transformation. Econometrics presents an additional tool that is used to reduce the dimensionality of a data set with times series' with a unit root by seeking a stationary linear combination. This phenomenon is known as cointegration. In this section, the concepts behind testing cointegration are systematically built up. VAR(p) and VECM models are explained.

2.3.3 VAR, VECM and Cointegration

Denote by y_{it}, $i = 1, \ldots, n$, the univariate time series that are in a vector θ_t. θ_t, $t = 1, 2, \ldots, T$ is a column vector of n variables. The unrestricted VAR(p) model is:

$$\theta_t = c + \sum_{l=1}^{p} \Omega_l \theta_{t-l} + \epsilon_t \tag{2.35}$$

$c = (c_1, \ldots, c_n)'$ is an $n \times 1$ vector of unknown constants. Ω_l is an unknown $n \times n$ matrix of coefficients of θ_{t-l} at lag l. $\epsilon_t, t = 1, \ldots, T$ are independent identically distributed errors distributed as $N_n(0, \Sigma)$ as $\mathbb{E}(\epsilon_t) = 0$, $\mathbb{E}(\epsilon_t \epsilon_t') = \Sigma$ and $\mathbb{E}(\epsilon_t \epsilon_{t-l}) = 0$ for $l \neq 0$. ϵ_t measures the degree to which the contemporaneous vector θ_t is determined by the VAR(p). ϵ_t is a variable that is influenced by the number of lags, p in the VAR model and the choice of coefficients, $\upsilon = (c, \Omega_1, \ldots, \Omega_p)$, that give weights in the linear combination that forms the VAR(p).

To illustrate, for $n = 3$ and $p = 2$,the VAR(2) is written:

$$\theta_t = c + \Omega_1 \theta_{t-1} + \Omega_2 \theta_{t-2} + \epsilon_t \tag{2.36}$$

Equation (2.36) is written in matrix form as:

$$\begin{bmatrix} \theta_{1t} \\ \theta_{2t} \\ \theta_{3t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} \omega_{11}^1 & \omega_{12}^1 & \omega_{13}^1 \\ \omega_{21}^1 & \omega_{22}^1 & \omega_{23}^1 \\ \omega_{31}^1 & \omega_{32}^1 & \omega_{33}^1 \end{bmatrix} \begin{bmatrix} \theta_{1t-1} \\ \theta_{2t-1} \\ \theta_{3t-1} \end{bmatrix} + \begin{bmatrix} \omega_{11}^2 & \omega_{12}^2 & \omega_{13}^2 \\ \omega_{21}^2 & \omega_{22}^2 & \omega_{23}^2 \\ \omega_{31}^2 & \omega_{32}^2 & \omega_{33}^2 \end{bmatrix} \begin{bmatrix} \theta_{1t-2} \\ \theta_{2t-2} \\ \theta_{3t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{bmatrix} \tag{2.37}$$

Equation (2.37) is written equation-by-equation as:

$$\begin{aligned}
\theta_{1t} &= c_1 + \omega^1_{11}\theta_{1t-1} + \omega^1_{12}\theta_{2t-1} + \omega^1_{13}\theta_{3t-1} + \omega^2_{11}\theta_{1t-2} + \omega^2_{12}\theta_{2t-2} + \omega^2_{13}\theta_{3t-2} + \epsilon_{1t} \quad (2.38)\\
\theta_{2t} &= c_2 + \omega^1_{21}\theta_{1t-1} + \omega^1_{22}\theta_{2t-1} + \omega^1_{23}\theta_{3t-1} + \omega^2_{21}\theta_{1t-2} + \omega^2_{22}\theta_{2t-2} + \omega^2_{23}\theta_{3t-2} + \epsilon_{2t}\\
\theta_{3t} &= c_3 + \omega^1_{31}\theta_{1t-1} + \omega^1_{32}\theta_{2t-1} + \omega^1_{33}\theta_{3t-1} + \omega^2_{31}\theta_{1t-2} + \omega^2_{32}\theta_{2t-2} + \omega^2_{33}\theta_{3t-2} + \epsilon_{3t}
\end{aligned}$$

The VAR(p) can be written in concise matrix notation which is useful for specifying the prior distribution of a VAR(p). Let $\theta{=}(\theta_p, \theta_{p+1}, \ldots, \theta_T)$, $\upsilon{=}(c, \Omega_1, \ldots, \Omega_p)$, $X{=}(X_0, \ldots, X_{T-1})$, $X_t{=}(1, \theta'_t, \ldots, \theta'_{t-p+1})$, $\xi{=}(\epsilon_p, \ldots, \epsilon_T)$. Then, the VAR(p) is:

$$\theta = X\upsilon + \xi \qquad (2.39)$$

The coefficients of the VAR(p) are estimated using the OLS estimator:

$$\upsilon^{OLS} = (X'X)^{-1}X'\theta \quad \text{or} \quad \upsilon_i^{OLS} = (X'X)^{-1}X'\theta_i \qquad (2.40)$$

where θ_i is a $T \times 1$ vector with the t-th element θ_{it}.

The vector autoregressive (VAR) model extends univariate autoregressive models to dynamic multivariate and provides better forecasts than univariate time series models (Zivot and Wang, 2006). VAR models are used to describe and forecast multivariate time series for stationary time series.

For non-stationary time series a Vector Error Correction term is added to form a vector error correction model (VECM) and it is necessary to test for the existence of a stationary linear combination of the non-stationary terms (cointegration). Cointegration relations are not directly apparent from a VAR(p) in levels such as Equation (2.35). It must be transformed into Vector Error Correction Model (VECM) by taking the first difference of θ_t so that:

$$\Delta\theta_t = c + \mathbf{\Pi}\theta_{t-1} + \sum_{k=1}^{p-1}\Gamma_k\Delta\theta_{t-k} + \epsilon_t, \qquad k = 1, \ldots, p-1 \qquad (2.41)$$

where $\Pi = \sum\Omega_i - \mathbf{I}_n$ and $\Gamma_k = -\sum_{i=k+1}^{p}\Omega_i$.

From a VAR model it is possible to analyse the impacts that the y_{it} have on each other over time and how the variables respond to other unobservable factors. The VAR is based on the assumption that having observed a set of variables over time, the underlying structure that connects them is unknown but the variables are all generated by this unknown underlying relationship. In this thesis it is assumed that how the changes to the underlying structure of mortality occur is unknown, that is, the manner in which the log-mortality rates (Chapter 5) and the parameters of the Heligman-Pollard model (Chapter 6) change from one year to the next is unknown. The VAR model is considered as a tool that allows the data to describe the underlying structure. The VAR model also assumes that each variable (time series) is regressed on its own values and also on the values of the other variables in the system.

The procedure for estimating VAR(p) models is presented in section 4.1.1.2.

2.3.3.1 Lee-Carter as a VAR model and as a time trend model

Girosi and King (2007) present the Lee-Carter model as a special case of a multivariate random walk with drift (which they refer to as the RWD model). The difference in the specification of the Lee-Carter model and the RWD model is in the structure of the covariance matrix.

The Lee-Carter model is re-written in the following form for age x with a random walk assumption for trends given by $\kappa_t = \delta + \kappa_{t-1} + u_t$:

$$\begin{aligned}
\ln m_t &= a^{LC} + b^{LC}\kappa_t + e_t \\
\ln m_{t-1} &= a^{LC} + b^{LC}\kappa_{t-1} + e_{t-1} \\
\Delta \ln m_t &= \ln m_t - \ln m_{t-1} \\
&= b^{LC}(\delta + u_t) + e_t - e_{t-1} \\
\ln m_t &= b^{LC}\delta + \ln m_{t-1} + e_t^* \\
\text{where} \quad e_t^* &= b^{LC}u_t + e_t - e_{t-1}
\end{aligned}$$

Only in this case, with the random common mortality trend for the log of the rates assumed to be a random walk, will the Lee and Carter (1992) model corresponds to a difference stationary model. Volatility is modelled with a single common factor $k(t)$ and independent noise $\epsilon(x,t)$ (Alho, 2000; Lee and Miller, 2001).

Hari et al. (2008) extend the approach taken by Girosi and King (2007). Given a set of normality assumptions on the residuals, the time trend $\ln m_{x,t} - \ln m_{x,t-1}$ is assumed to be normal based on the information from observations up to time $t-1$. The advantage of the technique in Hari et al. (2008) over the original Lee-Carter model and the re-formulation in Girosi and King (2007) is that it is less sensitive to the choice of period when estimating the long run mortality trend. In Hari et al. (2008) the use of more than one latent underlying factor is considered but no formal analysis into the number and form of the factors that drive time trends is done. This thesis adds to existing research by formally and systematically investigating the factors (assumes a strict factor structure and searches for unobservable variables that explain the variation in the observable variables) and principal components (does not assume a strict factor structure and searches for independent linear combination of the variables themselves) that explain and drive the variation in mortality trends and additionally time trends and cohort trends. Since a multi-country view is taken, the results are significant in making inference about the behaviour of the mentioned trends in different countries. Importantly, the common variation due to the factors is brought out by factor analysis and the amount of variation explained the principal components is brought out by principal component analysis.

2.3.3.2 Cointegration in Mortality Models: Lee-Carter

Some studies have analysed cointegration in mortality but with emphasis on how cointegration affects Lee-Carter (Chan, 2002) and the cointegration of the parameters in the Lee-Carter model (Darkiewicz and Hoedemakers, 2004). Darkiewicz and Hoedemakers (2004) suggest that cointegration analysis can be used as a diagnostic check of the validity of the Lee-Carter model. They do cointegration analysis of England and Wales log-mortality rates. Lazar (2004) finds that for Romanian mortality rates at high ages (63+) and given Lee-Carter's $\ln m(x,t)$ and $k(t)$, if the ages are pairwise cointegrated then the Lee-Carter model is the cointegration relation. The Lee-Carter model is written as a cointegration relation when x is fixed as:

$$\theta_{1t} = \delta + \beta^C \theta_{2t} + \epsilon_t \tag{2.42}$$

as compared with

$$\ln m_{x,t} = a_x^{LC} + b_x^{LC} \kappa_t + \epsilon_{x,t} \tag{2.43}$$

θ_{1t} (or $\ln m_{x,t}$) and θ_{2t} (or κ_t) evolve together in time with a long term equilibrium disturbed by random shocks with short-term effects.

The first study in this thesis uses principal components analysis and cointegration analysis to study mortality trends and the volatility in mortality trends as measured in period trends and cohort trends. An interesting relationship to point out is that cointegration can be viewed as a generalization of PCA since when PCA is performed on cointegrated variables the first principal component is the common stochastic trend Alexander (1999).

2.3.4 Bayesian Vector Autoregression Models

VAR models are often over parameterized since they impose no theoretical restrictions to guide the specification of the model and are consequently not parsimonious (Litterman, 1986; Zivot and Wang, 2006; Sims and Zha, 1998; Robertson and Tallman, 1999b; Brandt and Freeman, 2006; Baltagi, 2002). A VAR requires the estimation of $n+pn^2$ coefficients (parameters). The estimates of the coefficients of the VAR model, c and Ω_l, when estimated using unrestricted VAR(p) model are considered to be fixed quantities. These estimates of coefficients do not accurately reflect the underlying relationship because some of the estimated coefficients of the VAR model are non-zero purely by chance when estimated by OLS so restrictions may be imposed to reduce the number of parameters being estimated. This makes it seem like a parameter in the model affects the estimated mortality rates while in reality it does not.

Bayesian inference is based on the premise that the data is fixed but the population parameters are random and requires some knowledge of the distribution of these random parameters. In the case of the Bayesian VAR there is uncertainty regarding the distribution of the coefficient matrices that is reflected in the prior and resulting posterior distribution of the coefficients. One way of doing this is by giving the non-zero coefficients of recent observations more weight (Robertson and Tallman, 1999b). Litterman (1986) develops a Bayesian method that views the coefficients c and Ω_l as random variables rather than viewing them as fixed quantities like in the unrestricted VAR(p). Litterman (1986) specified the form of the prior distributions by giving them specific mean values and measuring the variation (the "tightness" of the distributions) around these given prior mean values using a set of hyperparameters that shall be explained shortly. This is known as Litterman's Prior or the Minnesota Prior because it was part of Litterman's work at the Federal Reserve Bank of Minneapolis and the University of Minnesota (Robertson and Tallman, 1999b). This method was extended by Sims and Zha (1998) to give the prior that shall be used in the HP-BVAR model described in this thesis.

What follows is the formulation of the Bayesian VAR starting with the unrestricted VAR(p) from Equation (2.35) building it into a Bayesian VAR with Litterman's Prior and a Bayesian VAR with Sims-Zha's Prior.

Each Ω_l, the individual elements, ω^l_{ij}, are independent normally distributed random variables and each variable follows a random walk with a drift that may be nonzero. The random walk assumption is implemented by giving the following means for the lagged coefficient matrices. At $l = 1$, the mean of the coefficient matrix, Ω_1, is the identity matrix (the prior mean for the coefficient of the $j = i$-th variable in equation i is 1) and at $l \neq 1$, the mean of the coefficient matrix is the zero matrix:

$$\mathbb{E}[\Omega_l] = \begin{cases} \begin{bmatrix} \omega_{11}^1 & \omega_{12}^1 & \omega_{13}^1 \\ \omega_{21}^1 & \omega_{22}^1 & \omega_{23}^1 \\ \omega_{31}^1 & \omega_{32}^1 & \omega_{33}^1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, & \text{if } l = 1; \\ \begin{bmatrix} \omega_{11}^l & \omega_{12}^l & \omega_{13}^l \\ \omega_{21}^l & \omega_{22}^l & \omega_{23}^l \\ \omega_{31}^l & \omega_{32}^l & \omega_{33}^l \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, & \text{if } l \neq 1; \end{cases} \tag{2.44}$$

Denote the variance of Ω_l as Ψ_l such that $\sqrt{\mathbb{V}[\omega_{ij}^l]} = \sqrt{\psi_{lij}}$. The random walk assumption imposed on the VAR(p) is loosened by specifying the standard deviations of the individual elements, ω_{ij}^l, of the coefficient matrices. The prior standard deviations are measures of confidence in the prior means. A high confidence that the estimated coefficients match the prior mean is indicated by a small prior standard deviation.

First, consider the elements along the diagonal of Ω_l, that is, ω_{ij}^l when $i = j$. Litterman (1986) defined a hyperparameter, λ_1^h, to be the prior standard deviation of ω_{ii}^l which controls the extent to which the process is a random walk. As λ_1^h gets smaller, the random walk assumption becomes stronger since there is little variation around the prior mean of ω_{ii}^l which was set at 1 as in Equation (2.44) and ω_{ij}^l when $i \neq j$ will tend to zero.

Litterman (1986) next considered each equation in the VAR(p) system to impose further restrictions on the standard deviations of the prior means.

Using the example with $n = 3$ and $p = 2$ any row in Equation (2.39) can be written as:

$$\theta_{it} = c_i + \omega_{i1}^1\theta_{1t-1} + \omega_{i2}^1\theta_{2t-1} + \omega_{i3}^1\theta_{3t-1} + \omega_{i1}^2\theta_{1t-2} + \omega_{i2}^2\theta_{2t-2} + \omega_{i3}^2\theta_{3t-2} + \epsilon_{it} \tag{2.45}$$

The variation in the i-th variable comes from its past values, the past values of the other variables in the system and the estimated constant. Litterman (1986) took the prior mean of the constant c_i as zero and the standard deviation of c_i to be a weight or fraction, denoted by the hyperparameter, $\lambda_4^h > 0$, of the variation in the i-th variable, σ_i and set the standard deviation of c_i to be $\sigma_i\lambda_4^h$. A small λ_4^h implies that there is little variation in c_i from zero.

Let σ_i be the standard deviation of the i-th variable and σ_j be the standard deviation of the j-th variable. The ratio, $\frac{\sigma_i}{\sigma_j}$ affects the value of the coefficient of

θ_{jt-1} in equation i. If $\sigma_i < \sigma_j$ then less weight will be given to the coefficient of θ_{jt-1} in the i-th equation and this coefficient will tend to zero. The contribution of the variable's own lags to the variable's variation and the contribution of the other variable's lags to the variable's variation is measured by a hyperparameter $\lambda_2^h, (0, 1]$. $\lambda_2^h = 1$ implies that the contribution to the variation from the variable's own lagged values and the variation from the other variables lagged values are the same. A small λ_2^h implies that ω_{ij}^l when $i \neq j$ will tend to zero. The coefficients at lags $l > 1$ have a prior mean of zero. Another hyperparameter, $\lambda_3^h > 0$, is used in the form $l^{\lambda_3^h}$. If $\lambda_3^h > 1$, as the number of lags, $l > 1$ gets larger $l^{\lambda_3^h}$ also gets larger.

Finally specifying Litterman (1986)'s prior standard deviations for the ω_{ij}^l in Ω_l as:

$$\sqrt{\psi_{lij}} = \begin{cases} \frac{\lambda_1^h}{l^{\lambda_3^h}}, & \text{if } i = j; \\ \frac{\sigma_i \lambda_1^h \lambda_2^h}{\sigma_j l^{\lambda_3^h}}, & \text{if } i \neq j; \end{cases} \tag{2.46}$$

gives Litterman's Prior.

To illustrate this, consider the system of equations that forms the VAR(p) with $n = 3$ and $p = 2$. The diagonal of the prior covariance matrix for the i-th equation is given by the standard deviations in parenthesis below the coefficients as:

$$\begin{aligned}
\theta_{1t} &= \underset{(\sigma_1\lambda_4^h)}{c_1} + \underset{(\lambda_1^h)}{\omega_{11}^1\theta_{1t-1}} + \underset{(\frac{\sigma_1\lambda_1^h\lambda_2^h}{\sigma_2})}{\omega_{12}^1\theta_{2t-1}} + \underset{(\frac{\sigma_1\lambda_1^h\lambda_2^h}{\sigma_3})}{\omega_{13}^1\theta_{3t-1}} + \underset{(\frac{\lambda_1^h}{2^{\lambda_3^h}})}{\omega_{11}^2\theta_{1t-2}} + \underset{(\frac{\sigma_1\lambda_1^h\lambda_2^h}{\sigma_2 2^{\lambda_3^h}})}{\omega_{12}^2\theta_{2t-2}} + \underset{(\frac{\sigma_1\lambda_1^h\lambda_2^h}{\sigma_3 2^{\lambda_3^h}})}{\omega_{13}^2\theta_{3t-2}} + \epsilon_{1t} \\
\theta_{2t} &= \underset{(\sigma_2\lambda_4^h)}{c_2} + \underset{(\frac{\sigma_2\lambda_1^h\lambda_2^h}{\sigma_1})}{\omega_{21}^1\theta_{1t-1}} + \underset{(\lambda_1^h)}{\omega_{22}^1\theta_{2t-1}} + \underset{(\frac{\sigma_2\lambda_1^h\lambda_2^h}{\sigma_3})}{\omega_{23}^1\theta_{3t-1}} + \underset{(\frac{\sigma_2\lambda_1^h\lambda_2^h}{\sigma_1 2^{\lambda_3^h}})}{\omega_{21}^2\theta_{1t-2}} + \underset{(\frac{\lambda_1^h}{2^{\lambda_3^h}})}{\omega_{22}^2\theta_{2t-2}} + \underset{(\frac{\sigma_2\lambda_1^h\lambda_2^h}{\sigma_3 2^{\lambda_3^h}})}{\omega_{23}^2\theta_{3t-2}} + \epsilon_{2t} \\
\theta_{3t} &= \underset{(\sigma_3\lambda_4^h)}{c_3} + \underset{(\frac{\sigma_3\lambda_1^h\lambda_2^h}{\sigma_1})}{\omega_{31}^1\theta_{1t-1}} + \underset{(\frac{\sigma_3\lambda_1^h\lambda_2^h}{\sigma_2})}{\omega_{32}^1\theta_{2t-1}} + \underset{(\lambda_1^h)}{\omega_{33}^1\theta_{3t-1}} + \underset{(\frac{\sigma_3\lambda_1^h\lambda_2^h}{\sigma_1 2^{\lambda_3^h}})}{\omega_{31}^2\theta_{1t-2}} + \underset{(\frac{\sigma_3\lambda_1^h\lambda_2^h}{\sigma_2 2^{\lambda_3^h}})}{\omega_{32}^2\theta_{2t-2}} + \underset{(\frac{\lambda_1^h}{2^{\lambda_3^h}})}{\omega_{33}^2\theta_{3t-2}} + \epsilon_{3t}
\end{aligned} \tag{2.47}$$

The prior means, $\mathbb{E}[\upsilon']$ are summarised as:

$$\overline{\upsilon}' = \mathbb{E}[\upsilon'] = \mathbb{E}\begin{bmatrix} c' \\ \Omega_1' \\ \Omega_2' \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \overline{\upsilon}_{i=1} & \overline{\upsilon}_{i=2} & \overline{\upsilon}_{i=3} \end{bmatrix} \tag{2.48}$$

Denote the diagonals of the i-th equation's prior covariance matrix under the

Litterman Prior by $\overline{\mathfrak{G}}_i$ (whose elements are the standard deviations in parenthesis in Equation (2.48) e.g. $\overline{\mathfrak{G}}_1 = (\sigma_1\lambda_4^h), (\lambda_1^h), (\frac{\sigma_1\lambda_1^h\lambda_2^h}{\sigma_2}), (\frac{\sigma_1\lambda_1^h\lambda_2^h}{\sigma_3}), (\frac{\lambda_1^h}{2^{\lambda_3^h}}), (\frac{\sigma_1\lambda_1^h\lambda_2^h}{\sigma_2 2^{\lambda_3^h}}), (\frac{\sigma_1\lambda_1^h\lambda_2^h}{\sigma_3 2^{\lambda_3^h}}))$. The posterior mean of υ, denoted υ^{Lit} is used as a point estimator for υ (Lütkepohl, 2005). This is calculated for the i-th equation as:

$$\upsilon_i^{Lit} = (\overline{\mathfrak{G}}_i^{-1} + \sigma_i^2 X'X)^{-1}(\overline{\mathfrak{G}}_i^{-1}\overline{\upsilon}_i + \sigma_i^2 X'\theta_i) \tag{2.49}$$

Litterman's Prior estimates the coefficients for one equation at a time.

Kadiyala and Karlsson (1997) analyse other options of priors that are used in estimating the coefficients of a VAR(p). Sims and Zha (1998) generalized the result into one prior. Sims and Zha (1998) replaced Litterman's Prior with a form of the Normal-Wishart prior as follows.

First, the prior distribution of the coefficients given Σ is assumed to be normal. The prior distribution of the covariance matrix, Σ is inverse Wishart (Drèze and Richard, 1983) with $\overline{S}$ the diagonal scale matrix in the inverse Wishart prior distribution of Σ.

The relation between the Minnesota prior and the Normal-Wishart prior is as follows. Let η be the $(1+np)n \times 1$ vector of stacked intercepts and coefficients in the VAR(p), e.g.:

$$\eta = vec(\upsilon) = (c_1, c_2, c_3, \omega_{11}^1, \omega_{21}^1, \omega_{31}^1, \ldots, \omega_{13}^2, \omega_{23}^2, \omega_{33}^2)' \tag{2.50}$$

Similarly, $\eta^{OLS} = vec(\upsilon^{OLS})$. Also, let ϑ be the $nT \times 1$ stacking of the T observations of the first variable, then the T observations of the second variable and so on:

$$\vartheta = vec(\theta) = (\theta_{11}, \theta_{12}, \ldots, \theta_{1T}, \theta_{21}, \ldots, \theta_{3T})' \tag{2.51}$$

In a similar way, $e = vec(\xi)$. Then, the VAR(p) takes the form:

$$\vartheta = (I_n \otimes X)\eta + e \tag{2.52}$$

where $e \sim N(0, \sigma \otimes I_T)$. $\otimes$ is the Kronecker Product that for two matrices

$\mathbf{A} = \left[\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right]$

and

$\mathbf{B} = \left[\begin{smallmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{smallmatrix}\right]$, $\mathbf{A} \otimes \mathbf{B} = \left[\begin{smallmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{smallmatrix}\right]$

The BVAR assumes the existence of some kind of a prior, a barrier that prevents the coefficients from being non-zero unless they contain information is created (Kadiyala and Karlsson, 1997) and in particular the coefficients of long-term lags shall be closer to zero than the short-term lags. This is considered to be a way of imposing structure on the system. It reaches a compromise between over-parameterization (in VAR modelling) and under-parameterization (in univariate modelling). It will also improve the accuracy of estimates and subsequent forecasts by introducing appropriate prior information into the model. In particular the BVAR model proposes that the standard deviations of the coefficients decrease as the lags increase. Litterman (1986) incorporates a prior into the system by considering the behaviour of each equation in the system on its own. Sims and Zha (1998) and Waggoner and Zha (1999) incorporate prior information into the VAR model by considering the entire system.

In the second study in this thesis, BVAR will be used to model the system of parameters of a parametric mortality model and thereby develop a dynamic parametric mortality model. The BVAR will give a structure that is part-way between the under-parameterized dynamic model in McNown and Rogers (1989) and the over parameterized dynamic model in Sherris and Gaille (2010a).

2.4 Two-Stage Mortality Models: From static to dynamic parametric mortality models

The Heligman-Pollard model has been used as a basis for creating age-period tables (Forfar and Smith, 1987; McNown and Rogers, 1989). According to Wong-Fupuy and Haberman (2004), these age-period table would be described as "a model based on graduating mortality measures with respect to age for specific time periods, involving two stages so that the parameters need to be projected (Congdon, 1993; Forfar and Smith, 1987; McNown and Rogers, 1989)."

Forfar and Smith (1987) fitted the Heligman-Pollard model to the English Life Tables (ELT 1-13) and found there was a "relatively good fit" and, further, there was a regular progression in the parameters of the Heligman-Pollard model. However, it is important to note that the spacing of the time periods when the model was fitted is irregular since ELT 1 and ELT 2 are both based on data from 1841 while ELT 3 is based on data from 1846. Although from ELT 8 to ELT 13 there is a consistent 10 year interval (1911, 1921, . . . , 1971) there was no ELT for 1941. Using the estimated sequence of parameters, Forfar and Smith (1987) estimated possible parameters for 1981 and 1991. After 1981's ELT 14 was developed they compared their results and found that the predicted Heligman-Pollard parameters resulted in a model that had a smaller error in fit for females than for males especially for higher ages. To get a better fit for the males they used an adjusted formula:

$$q_{x,t}^{Adj.HP} = \frac{f_t(x)}{1+f_t(x)}; \qquad f_t(x) = A_t^{(x+B_t)^{C_t}} + D_t \exp[-E_t(\log\{\frac{x}{F_t}\})^2] + \frac{G_t H_t^x}{1+G_t H_t^x} \tag{2.53}$$

These findings were surprising especially because the parameters for females included a year when the parameters had to be forced due to the presence of multiple humps. Also, the adjusted model is quite different from the original equation (2.16).

A study that extends the method of projecting mortality in Forfar and Smith (1987) is McNown and Rogers (1989) where a parameterized time series is modelled as described in Thompson et al. (1989) but using the techniques of Box and Jenkins (1976). Their study used USA mortality data for the years 1900-1985. The Heligman-Pollard model was the parametric mortality model they used to describe mortality patterns of over a lifespan at a sequence of points in time and obtain time series of the Heligman-Pollard model's eight parameters. The time series of the parameters exhibited highly non-stationary behaviour and this made it difficult to select a proper model for extrapolation. Due to the existence of a unit root in the time series of the parameters that were considered McNown and Rogers (1989) differenced the time series to achieve stationarity and as a result could estimate the parameters using univariate ARIMA models which when extrapolated formed a basis to obtain a series of Heligman-Pollard curves with time varying parameters. Due in part to the assumption of independence of the Heligman-Pollard parameters, the forecasts from the process described in Rogers (1986) are not accurate and are inconsistent (Lee, 1992; McNown and Rogers, 1992). Further, Lee (1992) also critics the absence of confidence intervals which is attributed to be due to problems that arise from the independence assumption as well. The gaps in this work are the use of univariate time series models and the assumption of independence of parameters. More recent work such as Sherris and Njenga (2009) and Sherris and Gaille (2010a) use advanced econometric techniques to fill this gap.

Although the approach in McNown and Rogers (1989) has been extended in studies of cause-specific mortality in Rogers and Gard (1991) and McNown and Rogers (1992) the independence assumption is still maintained. However, in 2010 Sherris and Gaille studied cause-specific mortality using econometric techniques and particularly the Vector Autoregressive (VAR) model and the Vector Error Correction Model (VECM) (for details see Sherris and Gaille, 2010a,b). Allowing for common stochastic long-run mortality trends improves the projected mortality curves.

This approach is extended in the second study of this thesis by considering a

Bayesian econometric model.

2.5 Bayesian Mortality Models

Mortality models with a Bayesian analysis have been developed in studies such as Pedroza (2006); Chan and Ting (2011); Cairns et al. (2011); Chunn et al. (2010) and Kogure and Kurachi (2010). An important advantage of using Bayesian methods is probability distributions are used to explicitly incorporate data and uncertainties in parameter estimation and model choice in a coherent and transparent way leading to realistic probabilistic projections (Abel et al., 2010); furthermore Bayesian methods provide a formal way to incorporate expert opinion into the mortality model.

According to Pedroza (2006) Bayesian analysis is a means of tailoring a model to a specific data set. Specification of prior information in terms of a prior distribution presents a challenge in formulating Bayesian mortality models. A probability distribution on the parameters, $p(\theta)$, depicts current knowledge about the model parameters and treats the model parameters as random variables. The data provides additional information about the model parameters and this information is incorporated in the likelihood, $p(data|\theta)$, which is proportional to the distribution of the observed data given the model parameters. A combination of the prior distribution and the additional information results in a posterior distribution, $p(\theta|data)$, which is an updated probability distribution on the model parameters. Given the observed data, the marginal posterior distributions describe the uncertainty about each model parameter completely. Marginal posterior distributions are summarised by posterior means or parameter estimates and posterior standard deviations of the parameter estimates quantifying parameter uncertainty.

The posterior distributions are used to project probabilistic mortality forecasts which are a better measure of the uncertainty in the mortality projections (Abel et al., 2010).

Pedroza (2006) implements the Lee-Carter model for predicting male mortality for the USA in a Bayesian framework and uses simulation (Markov-Chain Monte-Carlo methods and particularly a Gibbs sampler) to obtain the joint posterior distribution of the parameters of equation (2.12). A multivariate normal model for the log-mortality rates is used to provide a joint distribution for all the included age groups. A non-informative flat prior which will have no impact on the posterior is used for a_x^{LC}, b_x^{LC} and the drift term in k_t^{LC}. Non-informative priors will lead to improper posteriors. k_t^{LC} is assumed to have a Normal prior. The use of a Gibbs sampler to produce the posterior eliminates the problems that arise from estimating a high-dimensional joint posterior distribution of the unknown parameters of the Bayesian Lee-Carter model. Other Bayesian studies that are based on the Lee-Carter model include Reichmuth and Sarferaz (2008); Girosi and King (2008).

By combining a Bayesian statistical model and genetic algorithm, Chan and Ting (2011) develop a Bayesian mortality model for sick patients. In the general population, the genetic algorithm is not observable and the method in their model is not applicable for population mortality models.

The research in Abel et al. (2010) is comparable to the work in this thesis because in addition to using Bayesian methods it is for a developed country (England and Wales in their case, Australia in this thesis), is based on relatively good data (obtained from the Human Mortality Database in both cases) and includes autoregression models for time series. However, Abel et al. (2010) is analysing population growth and only considers a single time series of population change. Nonetheless, the desirable quantification of uncertainties in the data, parameters of the model and the model are achieved due to probability distributions of the probabilistic forecasts.

2.5.1 Bayesian Heligman-Pollard Models

Dellaportas et al. (2001) considers the Heligman-Pollard first law to be too restrictive through its use of least squares leading to over-parameterization. In order to resolve this issue a prior distribution for θ, $p(\theta)$, is specified then the posterior distribution of Heligman-Pollard Parameters $p(\theta|q_x) \propto p(\theta)p(q_x)$ is obtained. Markov Chain Monte Carlo simulation is used to fit the parameters and update θ at each iteration. The alternative used as priors include an informative Binomial prior for d_x[1]. A transformation is made on the vector of parameters θ to ensure that the resulting posterior is close to a normal distribution. A log-normal approach is also used in Dellaportas et al. (2001) based on the formulation:

$$\ln\left(\frac{q_x}{1-q_x}\right) = \ln\left(A^{(x+B)^C} + D\exp[-E(\log\{\frac{x}{F}\})^2] + GH^x\right) + \epsilon_x \qquad (2.54)$$

ϵ_x are independent $N(0,\sigma^2)$ variables and $\sigma^2 \sim p(\sigma^2)$ is the prior such that the posterior is $p(\theta,\sigma^2|q_x)$.

Sharrow et al. (2010) is similar to Dellaportas et al. (2001) and uses Bayesian Melding[2] to improve parameter estimates. Although this technique quantifies the uncertainty in the inputs (parameters) and outputs (estimated mortality rates) without making assumptions about relationships among the parameters, it does not

[1] $q_x = \frac{d_x}{l_x}$

[2] A statistical method used to calibrate uncertainty in data and models based on observed data over a longitudinal period with past trends in inputs and outputs in the form of prior distribution (Sharrow et al., 2010).

quantify the volatility of the trends in the parameters.

Both Dellaportas et al. (2001) and Sharrow et al. (2010) use Bayesian techniques to estimate the parameters of the Heligman-Pollard curve but do not explicitly model the time evolution of the parameters.

2.6 Longevity Risk Margins for Risk-Based Capital

In this section, the literature on the calibration of the longevity risk margins for risk-based capital is reviewed. Background information on the risk-based capital requirements is given followed by a review of literature on studies that are similar to the final study in this thesis. Most of the existing literature is about Solvency II since to date there has been no Australian study on the implications of the APRA-specified simplification for calculating the longevity stress margin.

Defined benefits superannuation schemes leave the retiree to bear the market risk and longevity risk and annuities certain leave the annuity provider rather than the annuitant to bear the market risk but the annuitant still bears the longevity risk. In the lifetime annuity, the annuitant does not bear any risk. The annuity income is guaranteed to be paid throughout the whole of the annuitant's lifetime. The annuity provider therefore bears all the risk - and in particular the market risk, longevity risk and the liquidity risk.

After the Global financial recession of 2008, the need to remove the burden of both the longevity risk and market risk from annuitants has become even more significant as several Australian retirees are likely to run out of wealth. The current Australian retirement system fails to cater for the provision of an income stream that will last for the remaining lifetime of an individual at retirement due to the absence of life annuities. Possible reasons for the limited life annuity market include the expense due to being priced high above the actuarially fair price, lack of suitable instruments for hedging the mortality risk and limited understanding of uncertainties that will affect future liabilities of the insurance companies such as longevity risk (Sherris and Evans, 2010). This study considers longevity risk. Mortality of annuitants will possibly improve beyond forecasted levels if cures or more effective treatments for certain terminal conditions such as cancer are discovered. Longevity risk for Australia (and other countries) has been difficult to model and quantify and has formed the basis for several studies (Heligman and Pollard, 1980; Lee and Carter, 1992; Booth et al., 2002b; Cairns et al., 2006b; Booth et al., 2006)

Risk-based capital requirements are being developed and implemented for insurance companies by regulatory bodies around the world. For instance, in Australia, the prudential regulations are being reviewed by the Australian Prudential Regulation Authority (APRA) and the technical papers on its proposed framework were released in July 2010. This new framework will ensure that a certain level of capital is maintained (under Pillar I addressing the quantitative requirements of an insurance company), the capital levels are supervised (under Pillar II where APRA can impose further capital requirements if deemed necessary) and as well as transparency in the insurance industry (under Pillar III's disclosure requirements). The first pillar is of interest in this thesis. In particular, the calculation of the prescribed capital amount (equivalent to Solvency Capital Requirement) that is determined by quantitative rules. Under APRA's proposed new requirements Life Insurers will be required to hold capital determined by a single measure in a risk-based framework that is more sensitive to the risks that a firm is exposed to. This is in line with the first purpose of the review "improving the risk sensitivity and appropriateness of the capital standards, in general and life insurance"(APRA, 2010a). This will be implemented by a capital base for life insurers.

APRA's requirement is similar to that in Solvency II as the level of capital must be sufficient to meet unexpected shocks over a one year horizon and also meet its obligations to its policyholders up to the end of that one year horizon with a probability of 99.5%. Solvency II allows for the use of a simplification in calculating the various sub-modules that are used to calculate the life underwriting risk module.[1] A significant component is longevity risk which is the risk that insured lives on average survive longer than expected.

In APRA's technical specification 'longevity' relates to the mortality of lifetime annuitants (APRA, 2011). Life annuities pay individuals an income stream for the rest of their life. The life annuity market in Australia is almost non-existent (Sherris and Evans, 2010) therefore this is a simulation based study. There exists a need for life annuities and as shown in The Henry Tax Review (Henry, 2009) the government is backing the idea of a private annuity market as well as government issued life annuities. Therefore it is important to consider how the capital requirements for these products will be determined. For capital reserving purposes, the longevity stress margin specified by APRA in its review is a permanent 25% decrease in mortality rates for each age.

APRA (2011) specifies that the simplification must be used in calculating the longevity risk stress margin mortality for the insurance risk capital charge of a

[1]The Solvency II Life Underwriting Risk Module is equivalent to APRA's Life Insurance Risk Charge. The Solvency II sub-modules are equivalent to APRA's stress margins.

portfolio of lifetime annuitants.[1] This does not allow for use of internal models. Although it is argued that there is great uncertainty in mortality models (CEIOPS, 2009), most quantify longevity risk in a realistic way to some extent and not all models are completely wrong.

The specified longevity risk stress margin is appealing due to its ease of computation because it simplifies a stochastic risk measure into a deterministic one.

Nevertheless, the magnitude and structure of the longevity risk stress margin has been criticised in the Solvency II QIS4 report CEIOPS (2008) and also in studies including Börger (2010) and Plat (2010). It is a uniform measure across age and country. This "one size fits all" approach when applied to more than 30 countries is potentially problematic due to the assumptions that underlie the calibration of the longevity risk stress margin. Further, the longevity stress margin serves to stress longevity risk over a one year horizon with the intention of reflecting the 99.5% confidence interval. The number of historical observations used to calibrate the longevity risk stress margin will affect the magnitude deemed to be adequate for a one year horizon. Specific issues regarding the methodology used in calibrating the longevity stress margin are addressed in section 4.3.

In the Australian framework, there is only one country being considered compared to over 30 in Solvency II. Nonetheless, the other shortcomings of the longevity risk stress margin are still inherent in APRA's proposition. For example, the simplification does not have a granular basis[2]

In order to analyse the suitability of the longevity risk sub-module in Solvency II or the longevity stress margin in the insurance risk charge in APRA's new regulatory framework an analysis of annuitant mortality data is required. Australian annuitant data is not available so it is not possible to investigate Australian annuitant mortality. However, Australian pensioner data is available and this gives some insight into features that annuitant data may have. Literature that compares population mortality and pensioner mortality such as Knox and Nelson (2006) and Stevenson and Wilson (2008) will form an important basis for interpreting results based on population mortality data which is readily available and used in studies such as those in this thesis. A study on pensioner mortality was done on mortality, from 2002 to 2005, of pensioners in public sector schemes in various parts of Australia in Knox and Nelson (2006) showing that the amount of income received in retirement is

[1] Margins that are chosen by the actuary include random mortality stress, future mortality stress, mortality event stress and lapse stress. Margins specified by APRA include expense stress and longevity stress.

[2] A granular basis considers the actual characteristics of the annuitants or members of a pension scheme.

inversely proportional to the mortality in retirement with low income retirees having a higher mortality than high income retirees. Notably, although the pensioner life expectancy was higher than the population life expectancy, they observed heavier pensioner mortality than population mortality at older ages (above 85). Stevenson and Wilson (2008) conduct a further analysis that builds on Knox and Nelson (2006) in an extended data set that uses data from 2002 to 2007. Over the additional years (2006-2007) they observed a lower mortality experience in the pensioners studied than for the previous period. A further finding that is consistent with Knox and Nelson (2006) is heavier pensioner than population mortality above the 85-90 age group. Why pensioner mortality exceeds population mortality at higher ages is still being investigated but it is suggested that pensioners retire in relatively good health and are therefore subject to less mortality from disease. This is a peculiar Australian phenomenon that was not observed in a similar study on New Zealand data (Stevenson and Wilson, 2008). Therefore, pensioners at retirements are considered to be "select" lives with the effect of selections decreasing with increased age (or equivalently time since retirement). It is also possible that it is an industry specific occurrence.

The simplifications by APRA and in Solvency II both involve an immediate shock as a measure of the longevity risk stress. This is not appropriate because improvements in mortality rates are often gradual and the results of medical advancements whose effects are seen over longer time horizons (Willets, 2004).

2.6.1 Importance of Distribution of Mortality Rates to Capital Determination

Capital reserves depend on the distribution of mortality rates. The distribution is obtained from a suitable stochastic model for longevity risk. Stochastic mortality models that are suitable for computing a VaR over a one-year time horizon must portray two essential components of longevity risk (Börger, 2010; Plat, 2010). First, the model must reflect the risk that next year's realized mortality may deviate (go higher or lower) from its expectation. Secondly, it must also anticipate that expected mortality beyond next year may also increase or decrease. Assuming a fixed mortality trend fails to reflect the latter component (Börger, 2010) and the implication of this assumption is that it will lead to over/under-capitalization.

It is noteworthy that a VaR over a one-year time horizon as a measure of longevity risk is not appropriate because longevity improvements are often the gradual result of medical advancements, for example, and the effects are seen over longer time horizons.

Several models have been fit to Australian data including the popular Lee-Carter model (Lee and Carter, 1992). The Lee-Carter method does not perform well with Australian data and possible suggestions include the non-linear behaviour of the time component and the changes over time in the age component which is assumed to be fixed (Booth et al., 2002b). The resulting fixed mortality trend will lead to underestimating of mortality rates if the model is fitted over periods with significant mortality decline.

Models with flexible trend are developed in several publications such as Cox et al. (2010); Hari et al. (2008) and Biffis (2005). One way of incorporating a flexible mortality trend is by allowing the parameters in the mortality model to be stochastic and change from year to year. The model developed in Sherris and Njenga (2011) that combines the Heligman-Pollard model and Bayesian VAR models captures both components of longevity risk is used in this study.

The mortality of annuitants in Australia has not been widely studied because of the small size of the annuity market. However, there have been limited studies (e.g. Knox and Nelson, 2006; Stevenson and Wilson, 2008) on the mortality of pensioners. The findings of studies such as these can be used to granularize the longevity stress margin by industry.

2.6.2 Capital Reserves

The capital base (or eligible capital) of a life insurance statutory fund is defined to be the amount of capital that APRA deems to be adequate to satisfy the Prudential Capital Requirements APRA (2010b). In the technical document released in July 2010 the proposed methods for determining the capital base and an additional insurance risk capital charge. Stress margins are determined from the best estimate assumptions. The increase in liabilities results in the insurance risk capital charge. The insurance risk capital charge is the amount of capital required to cover the risk that experience is worse than the best estimate for risks (such as longevity, mortality, expenses and lapses) and caters for adverse experience due to random fluctuations, extreme events, mis-estimation of the mean or the development of adverse trends over time APRA (2010b). The insurance risk capital charge will be the aggregate of the various individual risk's stress margins. APRA's proposed framework will allow for diversification benefits (using a prescribed correlation matrix) after the stress margins are applied.

APRA (2010b) outlines stress margins that will be specified by APRA. The key characteristics of these APRA-specified stress margins is that they should either be the same across the industry, simplify the process or solve the problems of insufficient data or inconsistency for individual insurers. The margin for longevity stress is one of the APRA-specified stress margins and the focus of this study. It is assumed that the longevity stress margin satisfies the characteristics above. Life annuities are classified as L3 (Annuity with Longevity Risk) products in APRA's proposed risk-based regime. The purpose of the capital charge is to buffer adverse experience. This could be due to random stress (random fluctuations in experience that may occur in the next 12 months causing mortality experience to be lighter than anticipated), event stress (extreme events that may occur in the next 12 months causing improved longevity) and future stress (mis-estimation of the mean and adverse trends that may apply from the reporting date for the remaining term of the liabilities). The margins are applied to the best estimate assumptions and are determined at 99.5% probability of sufficiency to meet liabilities. The APRA-specified margin assumes that a permanent 25% decrease in mortality rates at each age is equivalent to the 99.5% probability of sufficiency.

The assumption of a constant percentage decline is not realistic and does not reflect longevity risk accurately because longevity trends are declining at different rates and by different amounts at different ages Sherris and Njenga (2009). However, the purpose of the simplification is to reduce the amount of computation involved in calculating capital requirements so assuming a constant permanent decline in mortality rates is a reasonable assumption. The value of the constant is critical because assuming a value that is too high leads to over-capitalization while a value that is too low will lead to under-capitalization. The cost of the simplification to a company is potentially very high in monetary terms. The initial formulation of the longevity risk capital charge under Solvency II calibrated the permanent decline in mortality that was consistent with the 99.5% probability of having sufficient funds to meet liabilities as 25%. The Committee of European Insurance and Occupational Pensions Supervisors (CEIOPS) Consultation paper 49 attempts to justify how the permanent decrease was set to 25%. This percentage change in mortality was estimated using UK insurance companies' data CEIOPS (2007). Several respondents to Solvency II QIS 4 CEIOPS (2008) and in particular the European insurance and reinsurance federation (CEA) criticised the amount as too high and the approach as too simplistic. Eventually the stress was recalculated and reduced by 5% and currently stands at 20%. The permanent 25% decline in mortality implied over-capitalization for European insurers. This study investigates if it will be a similar case for Australian insurers. This is particularly of interest because the Institute of Actuaries of Australia in the Risk Business Capital Taskforce's paper of recommendations further proposed that a permanent 30% decline in mortality should be used for calculating a more prudent capital requirement.

Increased longevity leads to increased cost of annuities Blake et al. (2008) but for the insurer it also means increased liabilities as it will be making payments to individuals for longer than anticipated. Longevity risk is a significant risk because it when the estimated mortality is higher than the observed mortality such that there are more annuitants alive than was estimated, the benefits that are credited from those who died are less and therefore the cross-subsidy or mutuality effect is reduced.

2.7 Focus of the study and Overview of the Contributions

It is important to study the forces that drive changes in mortality trends. This is crucial for making assumptions that will underlie the method that is used to model and project the changing mortality trends. Understanding the forces that drive changes in the levels, trends and volatility on improvements in mortality rates is also important for calibrating and analysing longevity risk stress margins. Therefore, in this thesis development of a stochastic mortality model will begin with an analysis of mortality trends. The model developed will quantify uncertainty in the level, trend and volatility of future mortality rates. The developed model will consequently be used to analyse the longevity risk stress margin specified by APRA for implementation in 2013.

The problems tackled in this thesis are:

1. Analysing Past mortality trends and volatility: Factor analysis and Principal component analysis describe how many hidden underlying variables and how many independent linear combinations of the variables in a data set of mortality rates are necessary to explain the variation in the data set. The results of this analysis will be used to judge the suitability of using existing models to describe mortality patterns of different populations. Econometric Analysis (and specifically Cointegration Analysis) describes how many common trends exists in the mortality data set.

2. Modelling mortality trends and volatility: Extrapolative models do not convey information about the forces that control the changing shape of mortality. The cause of this major shortcoming is that extrapolative mortality models mainly rely on observed past trends. This problem is overcome by fitting a parametric model that captures the shape of mortality and extrapolating the parameters. In this way, the changing shape of mortality with time is modelled. This is

based on the assumption that the parameters adequately reflect the forces that control the shape of mortality at a given point in time.

3. Applying a mortality model to determine the longevity stress margin: The adequacy of the proposed simplification of the longevity stress margin in APRA (2010b) is gauged using a stochastic mortality model with Bayesian methods. A simple alternative that portrays the increasing longevity risk as age increases will be sought.

3

The Data

Introduction

This chapter presents the data available for analysis of mortality rates that are used in this thesis.

3.1 Data for Cross-Country Analysis

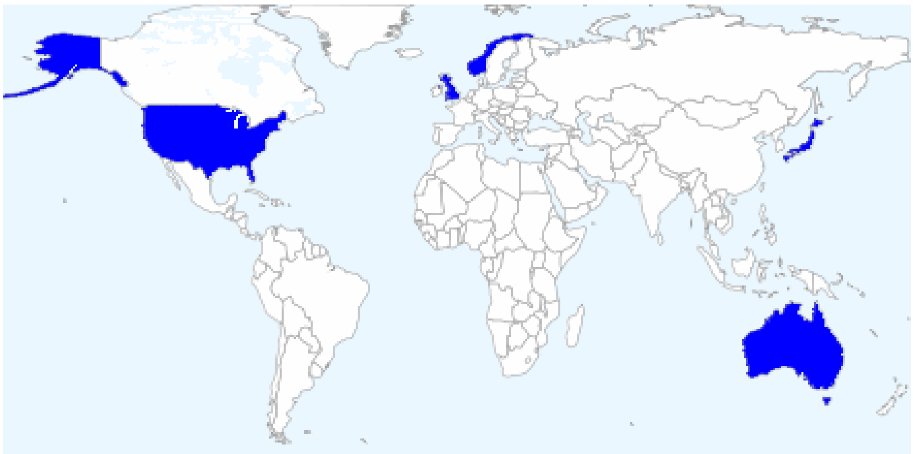


Figure 3.1: Geographical Location of the Countries Studied

This investigation of mortality rates is based on population data obtained from the Human Mortality Database, University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany) for Australia, Japan, Norway, UK and USA. These countries are selected because they provide coverage of different parts of the world, are all developed countries with similar economic and social environments and generally expected to exhibit consistent mortality patterns (Tuljapurkar et al., 2000).

3.1.1 Justification for the Choice of Countries

Australia, Japan, Norway, the United Kingdom (UK) and the United States of America (USA) are all countries from the Organisation for Economic Co-operation and Development (OECD) and are selected in order to capture developing countries from different parts of the world as shown in figure 3.1. These are all countries with a very high HDI (Human Development Index [1] - calculated by data on life expectancy, education and per-capita GNI (as an indicator of standard of living or income). Other statistics about the countries selected are given in table 3.1. From

Country	Population in '000s	GDP per capita	Human Development Index (HDI)	Income Inequality
Australia	21,472	41,985	0.937	0.336
Japan	127,176	34,050	0.884	0.329
Norway	4,762	57,637	0.938	0.25
UK	62,309	35,527	0.849	0.341
USA	310,233	47,005	0.902	0.378

Table 3.1: Some facts about the countries in the study in 2010. These are all countries with a high development index determined by considering literacy (education), gross national income (living standards) and life expectancy (health). Sources: OECD Factbook 2010 and Wikipedia

the 2010 Human Development Report, OECD countries in general have very high HDIs but Norway has the highest (1) (1 in 2009, 2 in 2008), followed by Australia (2), (2 in 2009, 4 in 2008). Of the remaining countries in this investigation, USA ranks fourth (4) (13 in 2009, 15 in 2008), Japan eleventh (11) (10 in 2009, 8 in 2008) and the United Kingdom ranks twenty-sixth (26) (21 in 2009, 21 in 2008). Each country has the highest HDI in its geographical region. Both the UK and Norway are considered because even though they are both from Northern Europe, Norway has the highest HDI in the World (Nordic countries all have outstandingly high HDIs) .

The HDI reports uses data up to 2 years prior (e.g. the 2010 report uses data up to 2008). This index has been criticised (see Wolff et al. (2011)) for reasons

[1] The HDI measures development by considering three dimensions - education, health and living standards of a population. The education component of the HDI is measured by mean of years of schooling for adults aged 25 years and expected years of schooling for children of school going age. The health component is measured by life expectancy. The standard of living component is measured by logarithm of GNI per capita. The scores for the three HDI dimension indices are then aggregated into a composite index using geometric mean i.e. $HDI = \sqrt[3]{I_{Life} I_{Education} I_{Income}}$. For details see Klugman et al. (2010)

including its components and the way they are included in calculating the index. For example, it will not reward high longevity if it is associated with high illiteracy because it assumes high longevity and high literacy occur simultaneously. This is a sound assumption in most countries. This thesis accepts the HDI as an adequate measure of human development.

3.1.2 Mortality Rates

Mortality rates have been changing at different rates for different age groups as shown in the mortality profiles in figures 3.2 and 3.3. Further, the volatilities in the mortality rates vary by age-group and by country as postulated in Tuljapurkar et al. (2000). In addition, as age increases improvements in mortality rates decrease Wong-Fupuy and Haberman (2004). This is visualized in figures 3.2 and 3.3 by comparing the steepness of the curves in the upper subplots (younger ages) to the steepness of the curves in the lower subplots (older ages). The curves for the younger ages are steeper indicating that the mortality improvements were occurring at faster rates.

For the multi-country analysis, data in age-groups of 5 years based on annual observations are used. The key features of the data are as follows:

1. **A log-linear relationship between mortality rates and time.** This feature varies for different countries and different time periods. For example, for Australian mortality, the very young ages in the first rows of figures 3.2 and 3.3 are almost linear. In the young ages in the second row the mortality rates are almost flat for males and linear after 1980 for females. In contrast, for Japanese mortality at most ages the curves look quadratic.

2. **Mortality Improvements by Age** With the exception of the very old ages, there has been a general improvement in mortality as evidenced by the downward slopes in the plots in figures 3.2 and 3.3

3. **The relative rate of mortality change over age.** The gradients (which represent the relative rates of mortality change) on the plots in 3.2 and 3.3 vary for different ages and different countries.

4. **Correlation** The surviving population at any given time is subject to similar policies and conditions. Therefore, mortality rates at some ages are correlated.

5. **Volatility** The mortality curves in figures 3.2 and 3.3 are not smooth. This is due to the volatility that is inherent in mortality rates.

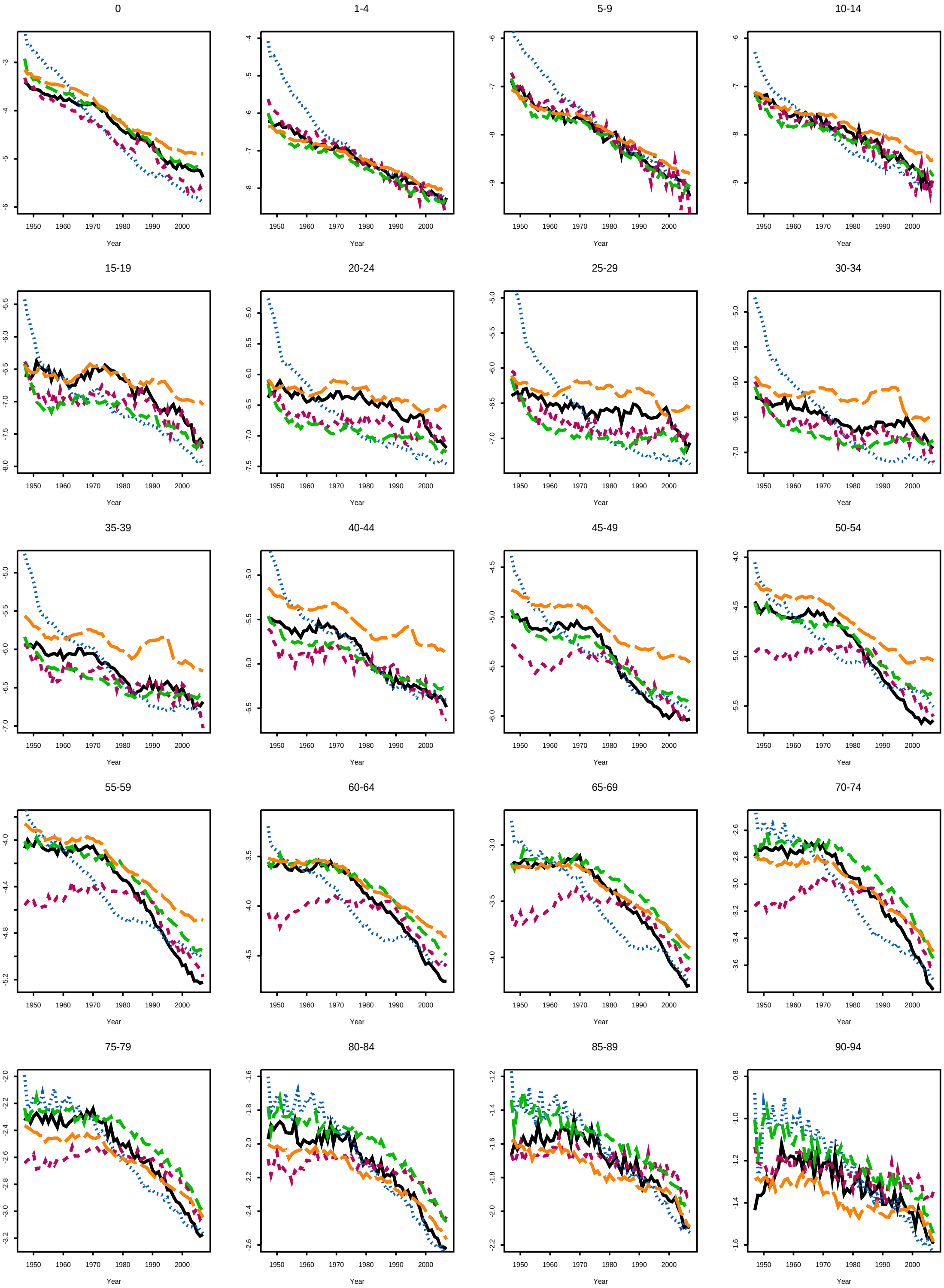


Figure 3.2: Male Mortality Profiles are decreasing. Mortality Rates ($\ln m_{x,t}$) for different age groups. The legend to indicate the countries is shown in figure 3.4.

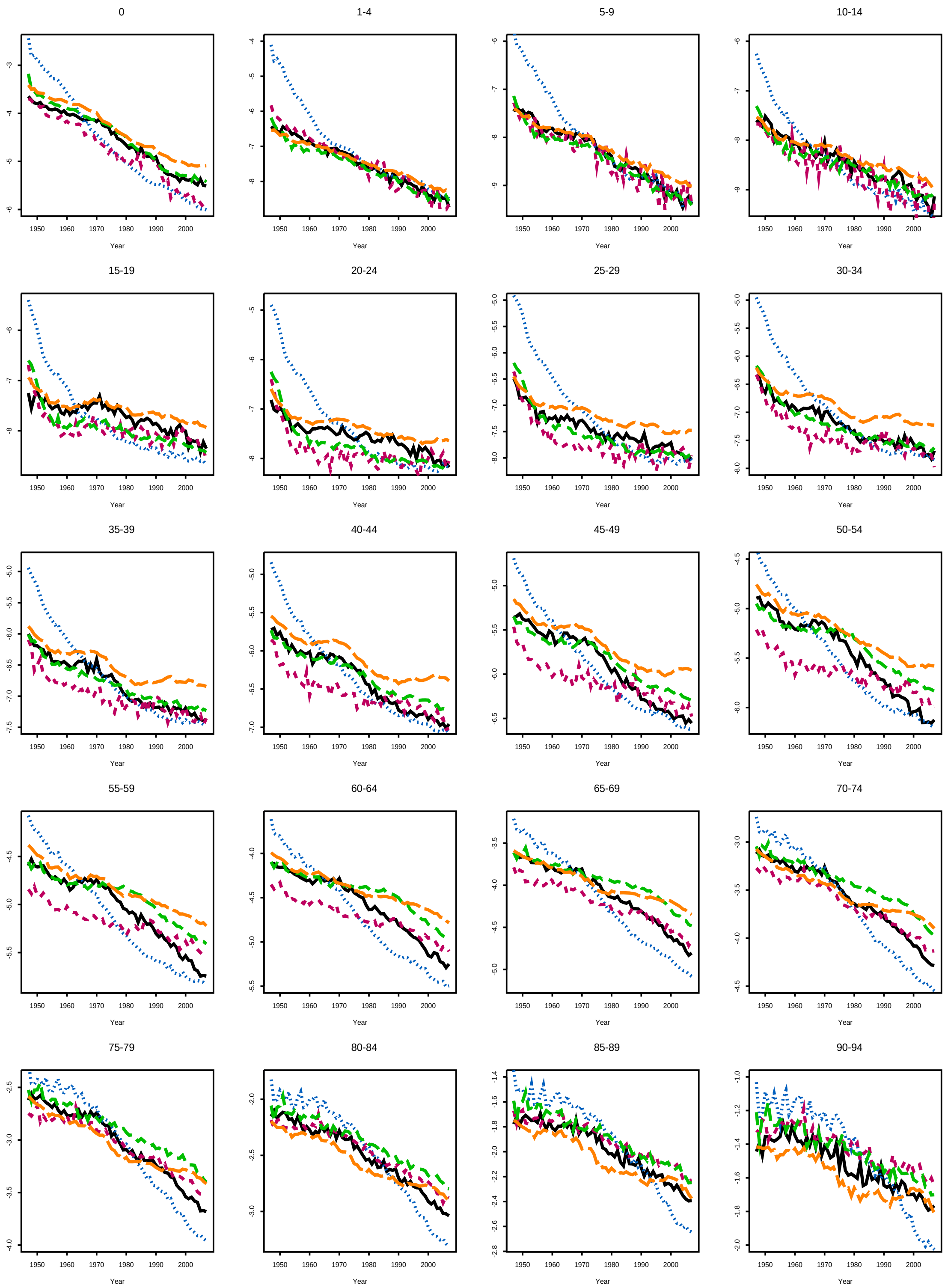


Figure 3.3: Female Mortality Profiles are decreasing. Mortality Rates ($\ln m_{x,t}$) for different age groups. The legend to indicate the countries is shown in figure 3.4.

These five features present the challenges and opportunities that can be used to develop a mortality model that adequately quantifies longevity risk. Further to that, the observation in Andreev and Vaupel (2006) that an increase in mortality improvement is usually followed by a decrease in mortality improvement can be seen (see also figure 5.2). This is interpreted to be partly because improvements in mortality due to a given factor do not affect mortality improvements to the same extent as time goes by since the gains made against a given variable of mortality will eventually stop. For example, when a vaccine such as the influenza vaccine is introduced the improvement in mortality rates would be large initially but as time goes by the year on year improvement due to the influenza vaccine will not be as large as that first improvement. Visual inspection of figures 3.2 and 3.3 intimates that it is important to consider if mortality improvement trends are stochastic, whether there are common trends across countries and also how many factors are required to explain the variation in mortality rates.

3.1.3 Age-Standardised Mortality Rates

The raw data is not suitable for comparing mortality rates across countries. To make the data comparable, standardised mortality rates are computed. The age-standardised mortality rates as shown in figures 3.2 and 3.3 are standardised using World Health Organisation World Standard Values (see table C.1 in the appendix). Age-standardised mortality rates are the weighted average of the age-specific rates for each of the populations to be compared (weighted using the World Standard Values). This useful single age-independent index is not affected by the different sizes of the population in each age-group. The standardised mortality rates capture the declining trend in mortality rates (see figure 3.4).

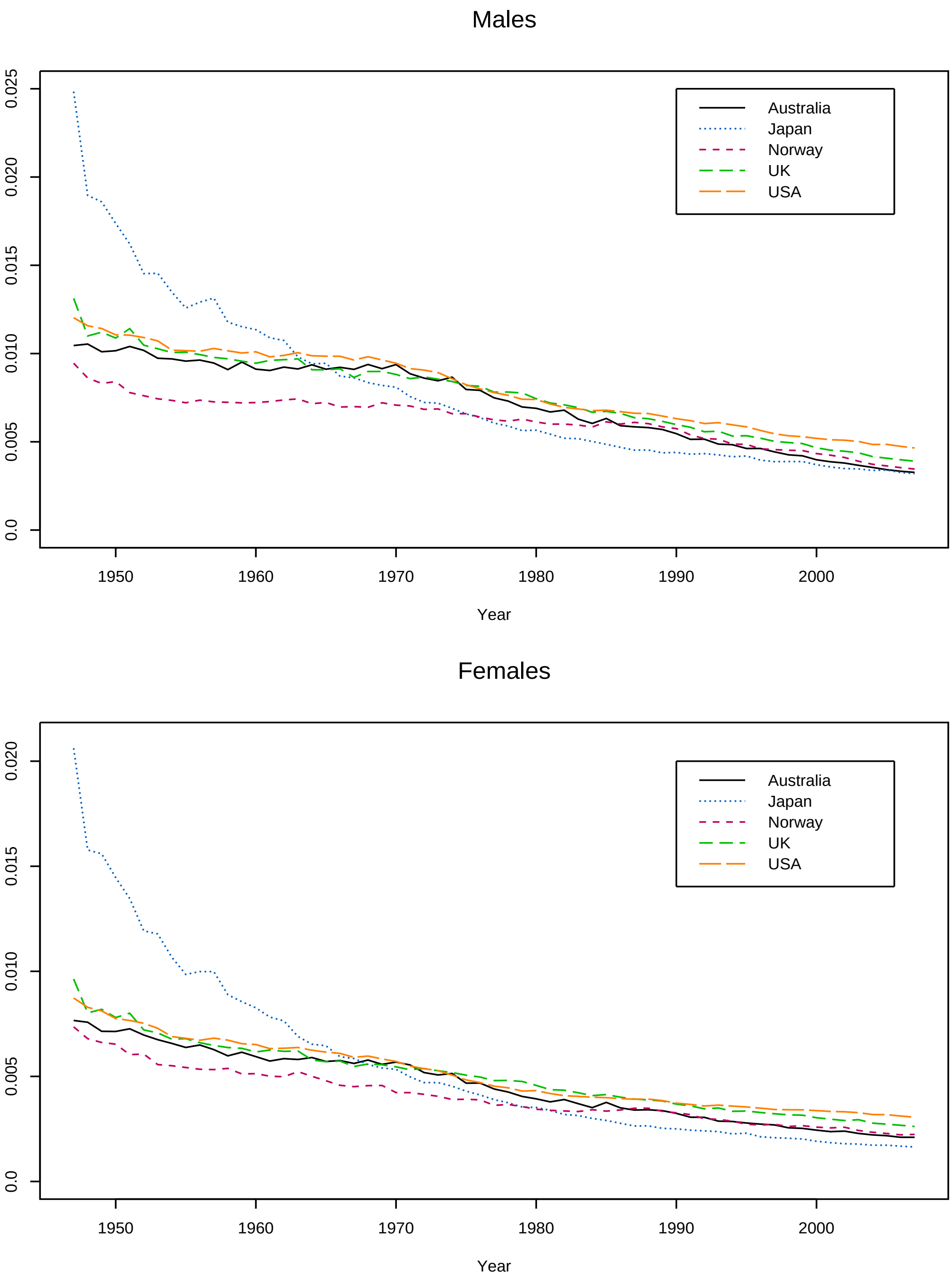


Figure 3.4: Standardised Mortality Rates

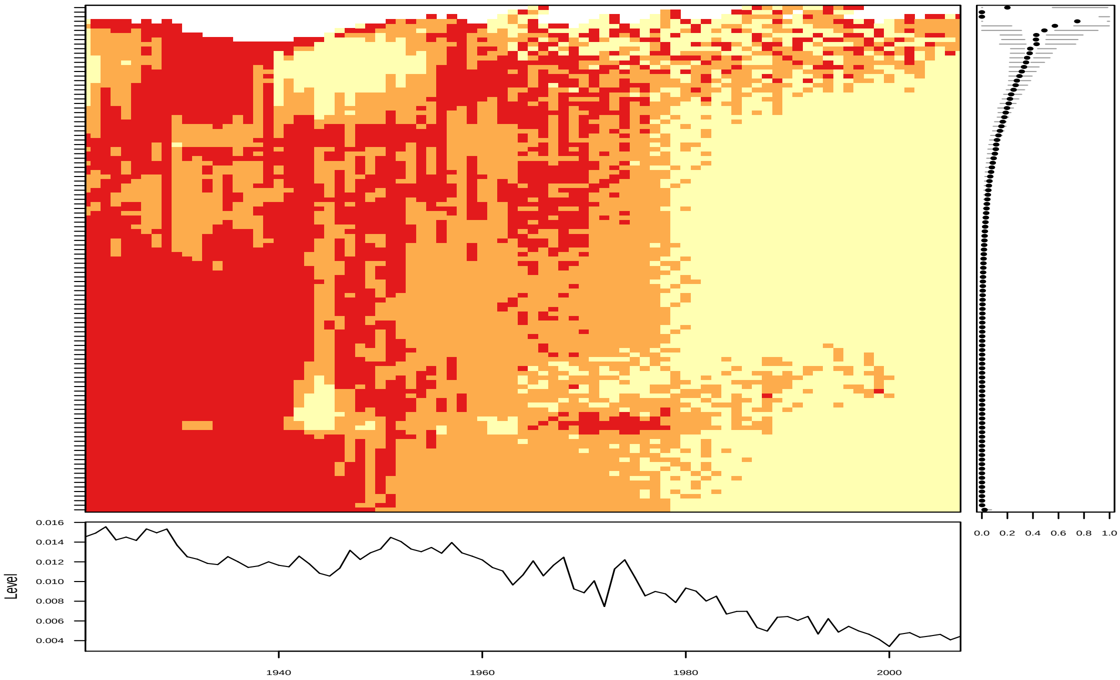
3.2 Data for Australian Analysis

The data set used is Australian population data obtained from the Human Mortality Database (HMD)[1]. The HMD data for ages above 80 are not the actual death rates Wilmoth et al. (2007). At the older ages the volatility of mortality rates increases. The maximum age used for modelling is 89 in order to increase the reliability of the model estimation. For Australian deaths before 1964 the Human Mortality Database data was only provided in 5 year age groups and was split into annual data using cubic splines and smoothed as described in McNeil et al. (1977). Until 1971 the population data was not adjusted for net undercounts and the data from 1971 onwards is of better quality and complete.

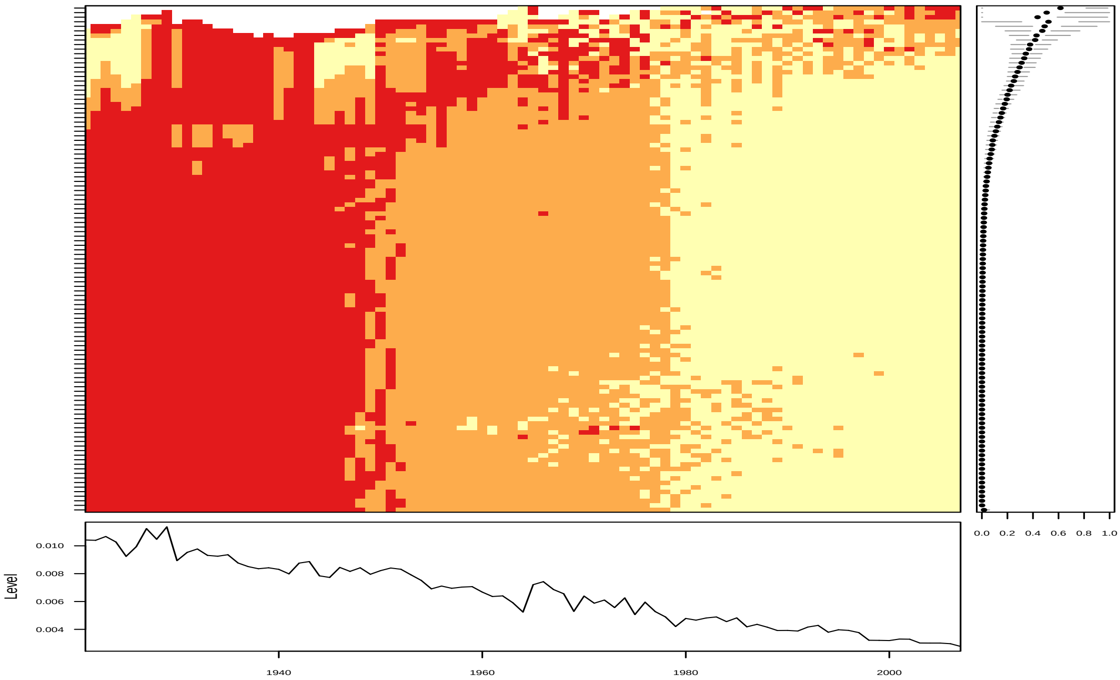
Mortality in Australia is shown in figure 3.5 which is prepared using the method in Peng (2008). The vertical axis on the main plot represents ages with age zero (0) at the bottom increasing to age one hundred and ten and over (110+) at the top. The data values are divided into three categories and represented as low values (Yellow or Very Light Grey), medium values (Orange or moderately grey) and high values (Red or Darker Grey). Missing data for the very old ages is represented by white areas at the top of the main plot.

There are more yellow areas (light shade) on the right hand side of the plot and more red areas (dark shade) on the left hand side of the plot. Mortality rates have declined significantly. From the box plots in the panel on the right side of the main plot, the variation in mortality rates is highest for the older ages, the dots are increasing and the whiskers (lines on either side of the dot) are longer for the older ages. The plot of levels on the bottom is the median values for all the mortality rate time series and quantifies the strong downward trend in mortality rates. In particular, a change in the structure of mortality patterns occurs in the 1970s. Between the 1970s and the early 1990s there is a strong downward trend suggesting that mortality was declining quickly. However, there has been a decline in the rate of decrease in mortality rates in more recent years, from the late 1990s. This implies another change in the structure of mortality patterns. An explanation for this is that in the recent past mortality trend improvements have been subject to a law of "diminishing returns" (Wong-Fupuy and Haberman (2004)). Capturing these changes in trends is an important feature of a mortality model.

[1] www.mortality.org



(a) Males



(b) Females

Figure 3.5: Declining Mortality Rates for Australia from 1921 to 2007 for ages 0 to 110+ from bottom to top. Red=High mortality, Orange=Moderate mortality, Yellow=Low mortality, White=Missing Data. The panel on the right of the main plots has boxplots of the data in each time series. In the bottom panel are median values across all the time series of mortality rates for each time point.

4

Methodology

Introduction

The thesis research questions were outlined in chapter 1. A review of what recent literature has addressed in relation to the research questions is presented in chapter 2. The research strategy that is followed to analyse mortality trends and their volatilities in this thesis is described in this chapter. In section 4.1, the methods used to visualise the patterns that underly mortality trends and mortality improvements are presented. Section 4.2 details the methodology used in the second study of this thesis. The demography model and econometric techniques used to quantify uncertainty in longevity risk are described in detail. Section 4.3 describes the APRA specified longevity stress margin and the procedure that will be used to analyse the adequacy of the specified stress margin.

4.1 Analysis 1: Factor Analysis and Econometric Analysis of Mortality Trends of Multi-Country Mortality Data: Understanding Mortality Trends and Improvements in Mortality Rates

A system of equations is often used for discrete ages when modeling mortality rates by age across time. The trends by age are captured by an age based model such as the Lee-Carter model. Mortality rates exhibit trends across time for a given age (see figures 3.2 and 3.3). At any given time there are trends across age. However as individuals grow older trends occur for a cohort of individuals born in the same year. Trends for a cohort reflect changes in time as well as changes in age. The mortality rates of most interest are the cohort rates since these are used to project future mortality based on each cohort year of birth. These are a combination of trends across time and by age. There is significant variability in the trends.

A great starting point for quantification of longevity risk is an exploration of observed mortality trends and their variability. Rates of mortality improvement can be calculated and then used to analyse changes in mortality over time (Andreev and Vaupel, 2005). Further, Andreev and Vaupel (2005) recommends using age groups rather than single ages because the resulting estimates are often unstable; however, the downside of this is a necessary loss of information on the finer details of mortality dynamics. The choice of method used to explore how mortality trends have changed over time and any underlying forces that drive these changes is therefore a significant factor to consider. Dimension reduction - as described in the literature review - is used as a solution to the problem of pattern recognition in a data set with multiple variables. This involves extraction and visualization of the patterns. Past literature such as Andreev and Vaupel (2005) analyse past improvements by estimating the surface of mortality improvement but due to the high variability in the improvements a direct examination of the mortality improvements is not done. Instead, the univariate rates of mortality improvement were smoothed[1] by age and time to obtain sequence of cubic splines that revealed an underlying pattern of mortality improvement.

In this thesis, a different approach is taken. First, the mortality trends themselves are analysed to seek a visualization of the patterns that describe their changes as well as any unobserveable underlying structures. Then, mortality improvements are analysed directly. As a consequence, a clear pattern of the dynamics of mortality improvements is obtained using dimension reduction techniques including PCA and FA.

4.1.1 Dimension Reduction of a data set of Mortality Rates

The set of data that is analysed is described in chapter 3. The first step in data analysis was to perform a principal component analysis to reduce the dimensionality of the data set of ASDRs by taking into account all variability in the variables. The next step was to perform a factor analysis to reduce the dimensionality of the data set of ASDRs by taking into account the variability which is due to common factors. Factor analysis and Principal Component analysis are both traditional dimension reduction procedures. Econometric dimension reduction is done by looking for common trends. Common trends are revealed by cointegration relationships.

Analysis of mortality trends is useful for anticipating mortality improvements

[1] Andreev and Vaupel (2005) describe smoothing as determining a smooth and slowly changing function of age and time from a distorted highly variably data set of mortality improvements.

that may be exhibited in the future. Therefore, dimension reduction is performed on the logarithms of the level of the mortality rates as well as on horizontal and diagonal differences defined in equations (4.1) and (4.2).

$$\begin{aligned}\Delta_h \ln m_{x,t} &= \ln m_{x,t} - \ln m_{x,t-1} \qquad \text{Horizontal Differences} \\ &= \ln \frac{m_{x,t}}{m_{x,t-1}}\end{aligned} \tag{4.1}$$

and

$$\begin{aligned}\Delta_d \ln m_{x,t} &= \ln m_{x,t} - \ln m_{x-1,t-1} \qquad \text{Diagonal Differences} \\ &= \ln \frac{m_{x,t}}{m_{x-1,t-1}}\end{aligned} \tag{4.2}$$

$\Delta_h \ln m_{x,t}$ shows how mortality rates change over time for a given age x. They reflect time trends only for any given age. When all ages are considered they allow common trends to be identified across time for levels of mortality.

$\Delta_d \ln m_{x,t}$ shows how mortality for a given age cohort changes from one year to the next. These are cohort trends or mortality changes for a set of individuals born in the same year and who experience common factors through time. The changes in the cohort rates include effects from age and from time.

Formal modeling and testing of model assumptions is required to confirm that common trends can be identified. This will be done in the econometric analysis. Principal Components Analysis and Factor Analysis will be used to identify factors driving mortality trends and their volatility after removing the trend through the drift term (after differencing).

4.1.1.1 Estimated Mortality Using the Lee-Carter Model

Consider the plots of the parameters the Lee-Carter model when applied to data from Australia, Japan, Norway, the UK and the USA in figures 4.1- 4.10.

The Lee-Carter model is based on a single improvement factor, k_t with differential impact by age, b_x. In the top right plot in figures 4.1- 4.10 the mortality trends from the Lee-Carter model are given. When k_t is not linear the resulting projections will be inaccurate (Girosi and King, 2007; Li et al., 2011). The gradient of the trend in k_t changes significantly over time. k_t for Norway and UK in figures 4.5 and 4.7 show this clearly. The detection and implications of structural breaks in k_t are discussed in detail in Li et al. (2011). The patterns across the countries vary and, although there is evidence of a common downward trend reflecting mortality improvement

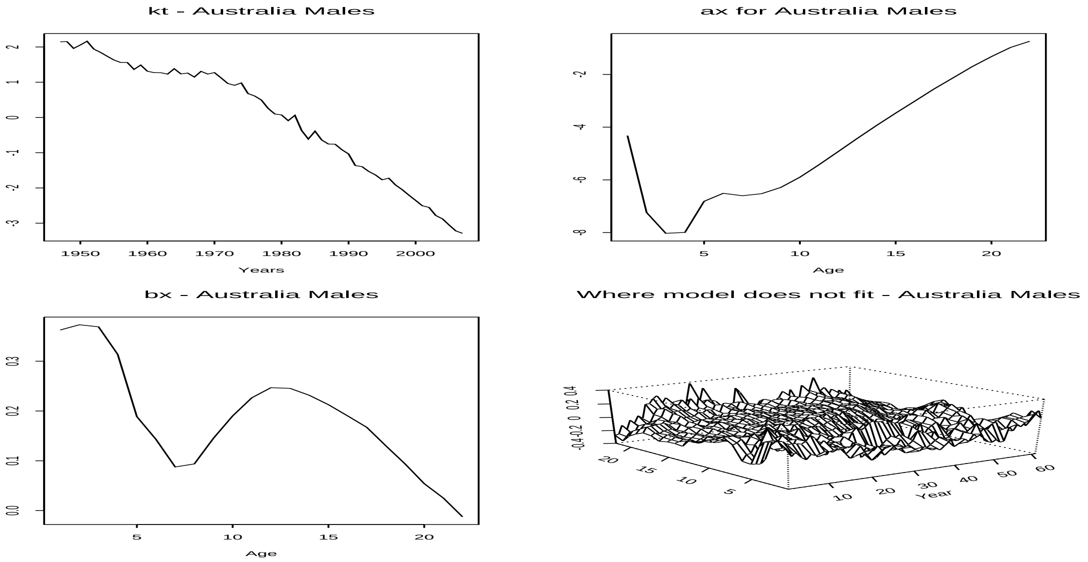


Figure 4.1: Lee-Carter Parameters Australia Males. The errors lie between -0.4 and 0.4 with a rough surface.

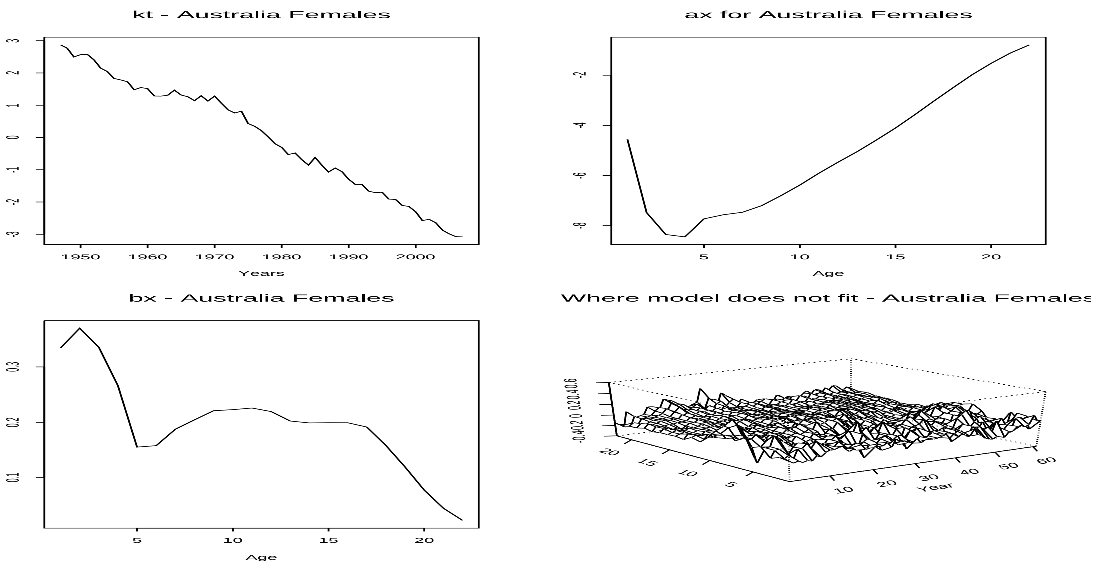


Figure 4.2: Lee-Carter Parameters Australia Females. The errors lie between -0.4 and 0.6 with a rough surface.

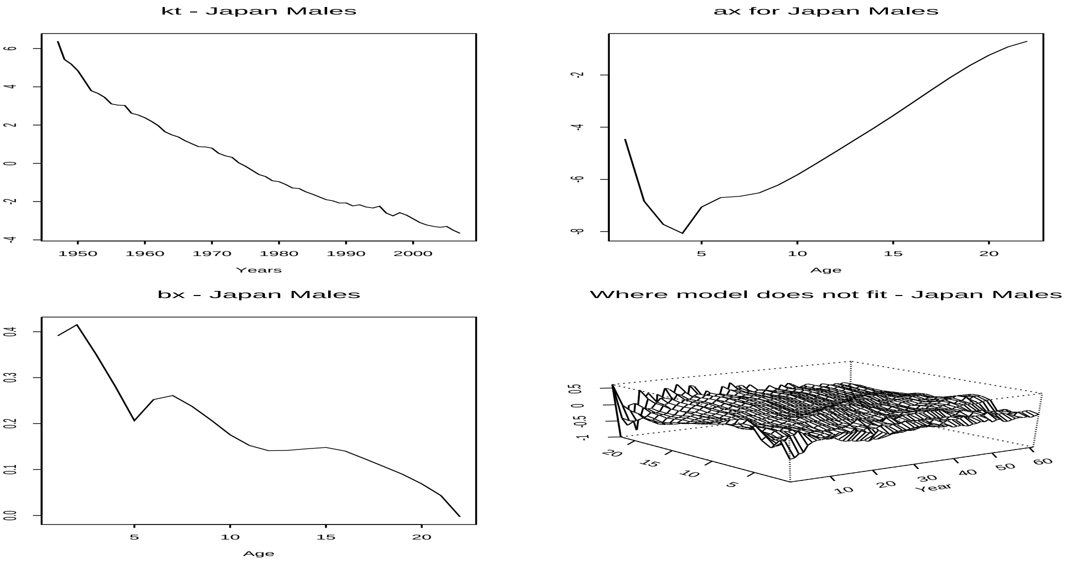


Figure 4.3: Lee-Carter Parameters Japan Males. The errors lie between -1 and 0.5 with a rough surface for the elderly.

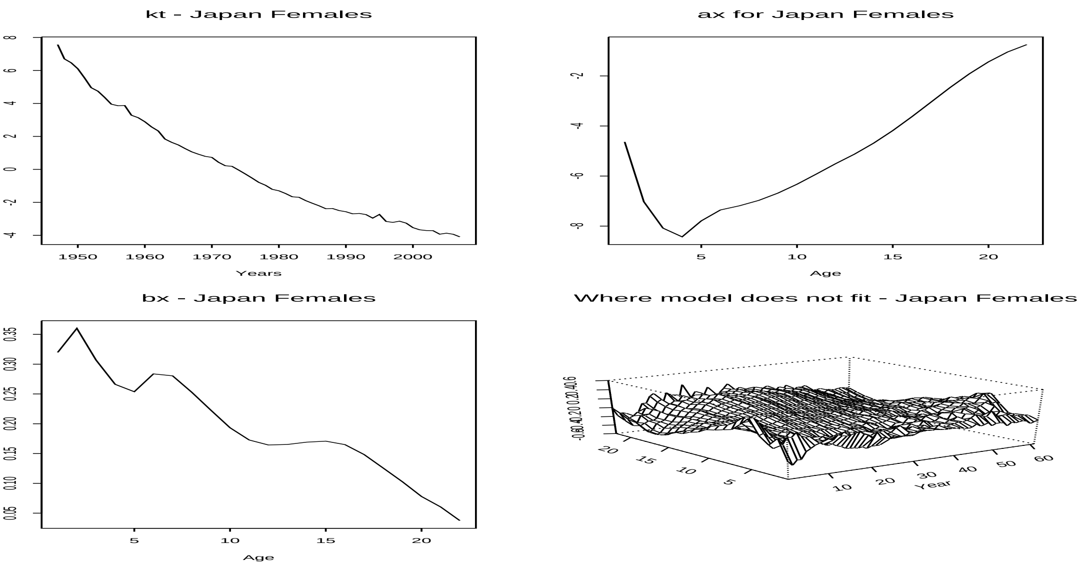


Figure 4.4: Lee-Carter Parameters Japan Females. The errors lie between -0.6 and 0.6 with a relatively smooth surface but in recent years the surface is rough.

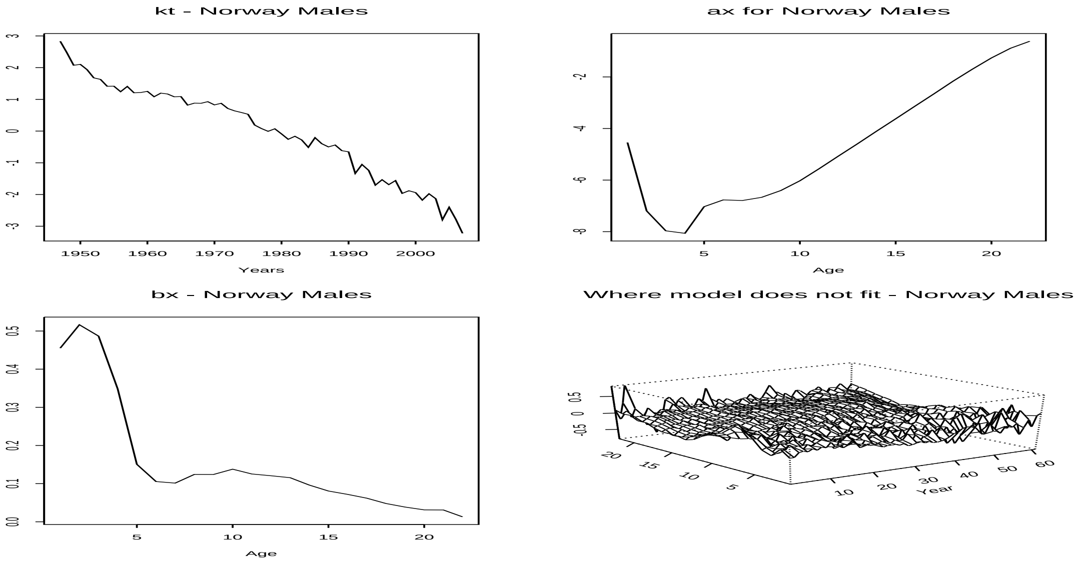


Figure 4.5: Lee-Carter Parameters Norway Males. The errors lie between -0.5 and 0.5 with a rough surface.

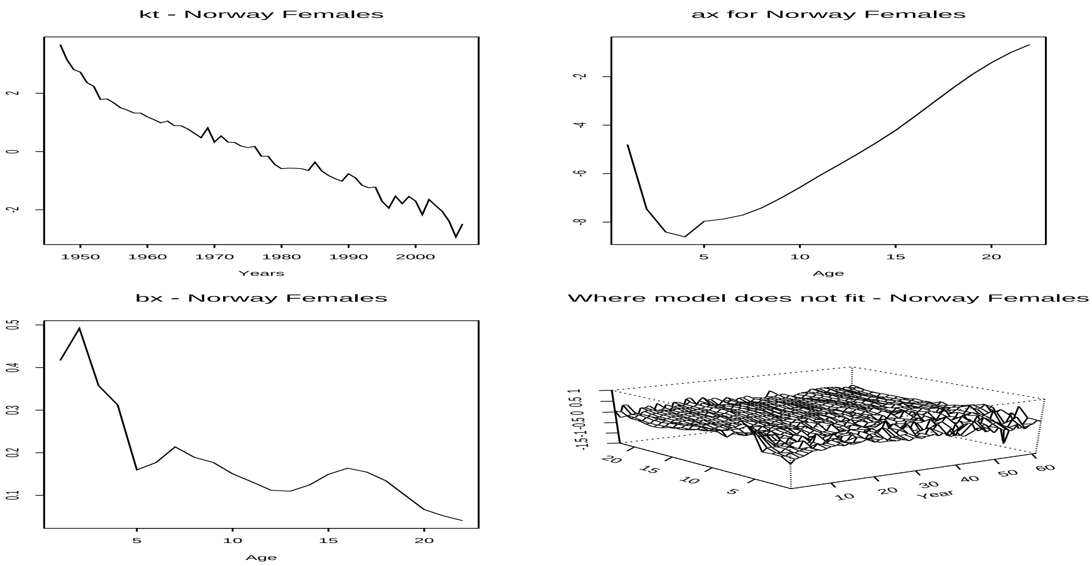


Figure 4.6: Lee-Carter Parameters Norway Females. The errors are quite large and lie between -1.5 and 1 with a rough surface. The Lee-Carter model does not fit well.

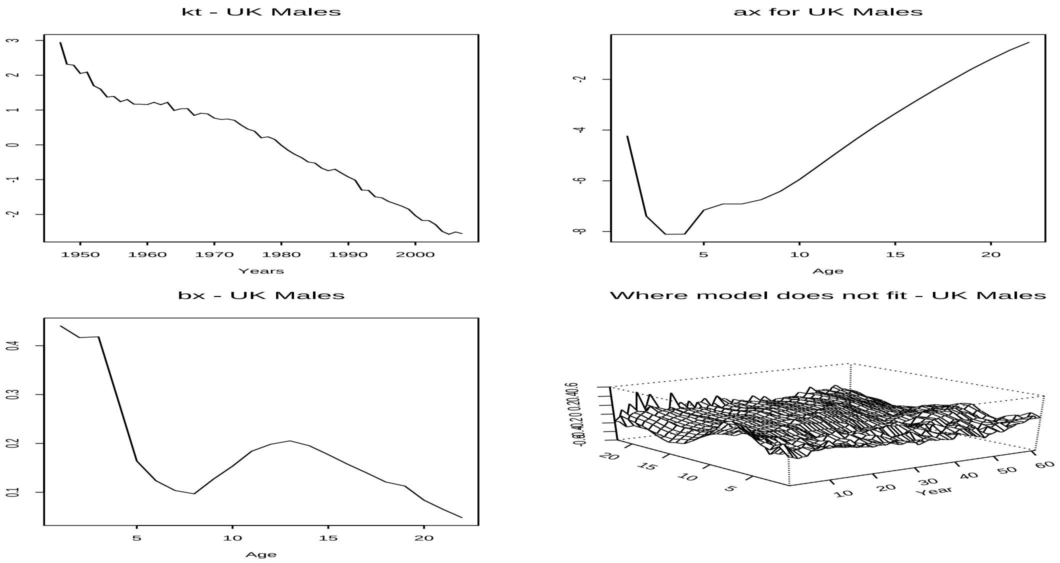


Figure 4.7: Lee-Carter Parameters UK Males. The errors lie between -0.6 and 0.6 with a relatively smooth surface for most years and the rough surface is mostly for the elderly.

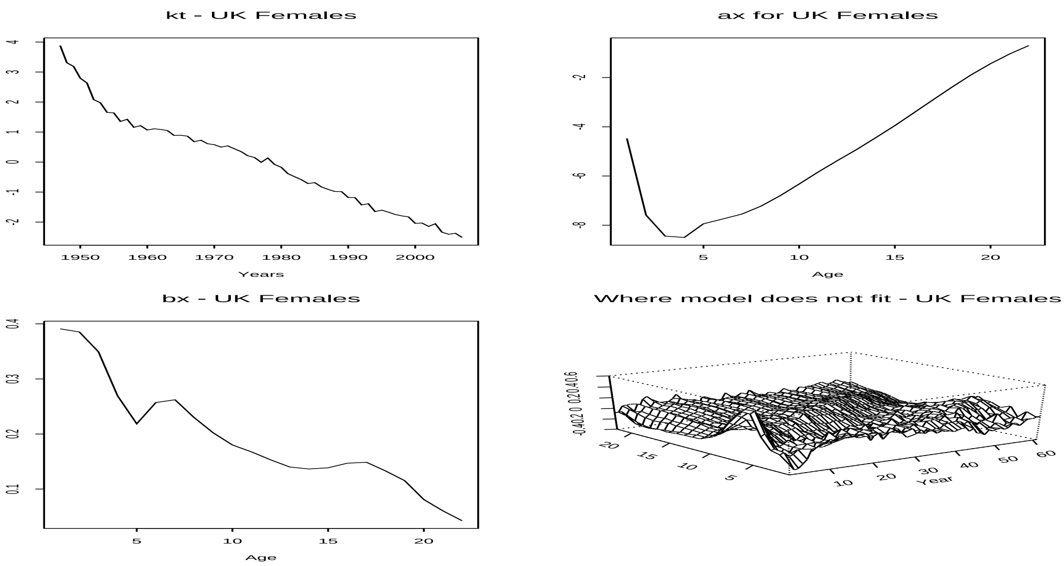


Figure 4.8: Lee-Carter Parameters UK Females. The errors lie between -0.4 and 0.6 with a relatively smooth surface for most years.

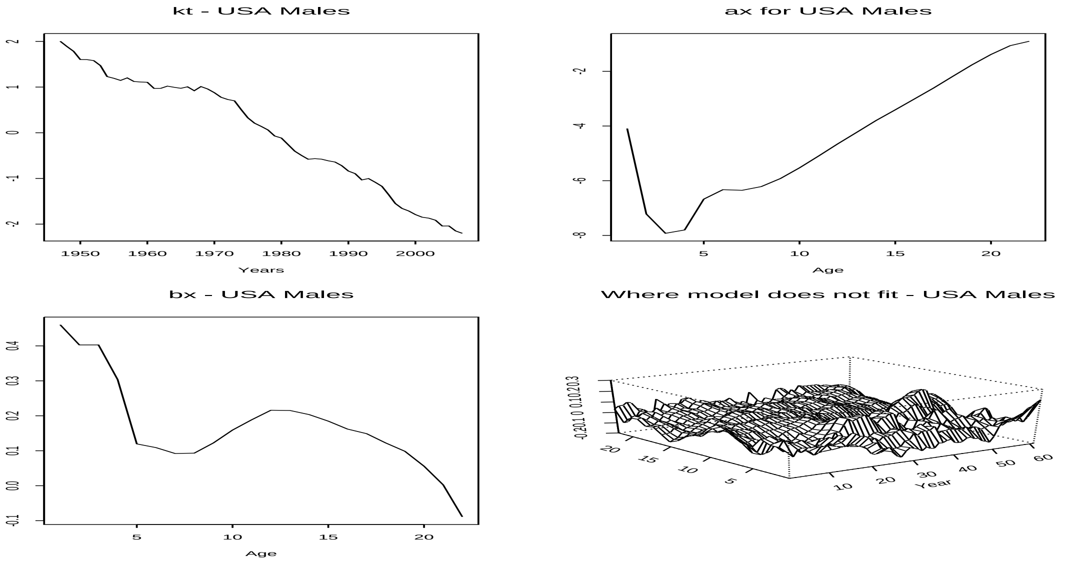


Figure 4.9: Lee-Carter Parameters USA Males. The errors lie between -0.2 and 0.3 with a rougher surface in recent years.

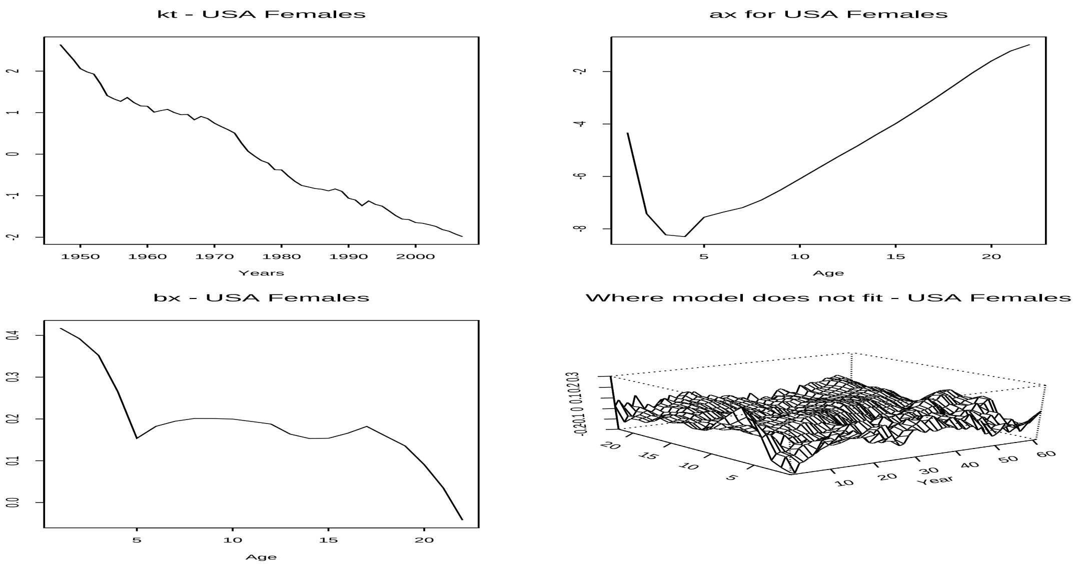


Figure 4.10: Lee-Carter Parameters USA Females. The errors lie between -0.2 and 0.3 with a rough surface for most years.

across these countries it is possible that other trends that are not captured by the Lee-Carter model exist.

Several explanations have been given on why the Lee-Carter model fits some populations better than others. For example, Jarner and Kryger (2009) hypothesizes that the Lee-Carter method is suitable for large populations because they are subject to constant improvement rates over time and performs badly when applied to small populations because their improvements are irregular. The use of a single improvement factor causes a poor fit if the population is not subject to a constant rate of mortality improvement. In the data used in this thesis, the countries when arranged in order of increasing population size are USA, Japan, UK, Australia and Norway. If the Jarner and Kryger (2009) hypothesis contains some truth, the Lee-Carter model should perform best for USA and worst for Norway.

The bottom right plot in each of figures 4.1- 4.10 shows the difference between the fitted values using the Lee-Carter model and observed values. This shows the model error structure indicating that there are trends not captured by the Lee-Carter model. The rougher the surface the poorer the fit. This visual display of the Lee-Carter model results for these countries indicates the need to assess if the mortality improvement trends are stochastic, whether there are common trends across countries and also how many factors and principal components are required to explain the variation in mortality rates.

In this thesis a PCA of the data set of age-specific death rates is performed. This establishes the number of factors that are significant in explaining various percentages of variation in mortality rates. This is important because models such as the classic Lee-Carter model underestimate the number of factors that are required to model mortality rates.

From section 2.2.3.1, the literature suggests that the number of factors or principal components explaining the variation in the mortality data affects the number of factors or principal components that should be included in a mortality model. Including the optimal number of factors or principal components can lead to a reduction in the number of parameters required. This leads to a parsimonious model.

4.1.1.2 Econometric Modeling: VAR, VECM and Cointegration

The stationarity or non-stationarity of a time series is very important in developing an appropriate model. Time series are classified as either stationary or non-stationary (integrated). When time series show non-stationarity in the mean or have a trending behavior the process is integrated. A stationary process is modeled using an equation with fixed coefficients estimated from past data.

A mortality rate series may have a deterministic trend around which the series fluctuates or, alternatively, it may have a stochastic trend and the rate of change in mortality would be stationary with drift or trend.

Unit root tests on $m_{x,t}$, $\ln m_{x,t}$, $\Delta_h \ln m_{x,t}$ and $\Delta_h \ln m_{x,t}$ analyse their stationarity or non-stationarity. Differentiating between these two situations is important in fitting mortality trends since the nature of the trends and shocks will have quite different implications for modeling future rates.

Define y_t, θ_t and VAR(p) as in section 2.3. The choice of VAR(p) (see equation (2.35)) is largely influenced by the lag order, p. Therefore, first step in estimating a VAR is selecting the lag length, p. This decision is based on minimizing a selection criterion (Lütkepohl, 1991). The Bayesian Information Criterion (BIC(p)) and Hannan-Quinn Criterion (HQ(p)) penalize VAR(p) models with large (p) and are:

$$BIC(p) = \ln|\tilde{\Sigma}(p)| + \frac{\ln T}{T} pn^2 \qquad HQ(p) = \ln|\tilde{\Sigma}(p)| + \frac{2\ln\ln T}{T} pn^2 \tag{4.3}$$

where

$$\tilde{\Sigma}(p) = \frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t'$$

Several VAR(p) models are usually estimated. The key assumptions in the VAR(p) model are that there is no serial correlation and no heteroscedasticity in the residuals and that the residuals are normally distributed. The VAR(p) models with different lag lengths, p, are analysed to check that the model's assumptions hold based on diagnostic tests. To test for serial correlation of the residuals a Portmanteau test is performed (Harvey, 1990), to test for heteroscedasticty ARCH tests are performed while to test for normality of the residuals normality tests such as the Jarque-Bera test are performed. Serial correlation of the errors will suggest that a transform of the data is necessary. The Box-Cox transform is often used in econometrics and a special case is the log transform. A similar analysis using a log transform of the standardized rates for each country will be done if the VAR(p) of the levels of the standardized rates has serially correlated residuals.

If all or some of the time series are I(1) while others are I(0) it is necessary to consider a cointegrated VAR since it is possible for the I(1) variables to be cointegrated. The variables in θ_t are at most I(1), $\Delta\theta_t$ are I(0). $\Pi\theta_t$ is the only term that may contain I(1) variables. Since $\Delta\theta_t$ is I(0) then $\Pi\theta_t$ must also be I(0). Π is the long run impact matrix and its rank determines the number of common stochastic trends. Due to the unit root(s) Π is a singular matrix and thus it is not of full rank. Let rank(Π)=r. If r=0 then $\Pi = 0$ and θ_t is I(1) with no cointegration. The VECM is therefore simply a first differences VAR(p-1) in this case. If $0 < r < n$, θ_t is I(1) with r linearly independent cointegrating vectors and $n - r$ common stochastic trends.

Denote the VECM as a function of r by $H(r)$. When $r = 0$ there is no cointegration and when $r = n$ the corresponding VAR(p) is made up of stationary variables. Testing for cointegration involves a nested model:

$$H(0) \subset \cdots \subset H(r) \subset \cdots \subset H(n)$$

Rank(Π) determines the number of cointegrating relations in θ_t. Likelihood ratio (LR) statistics for determining r are determined by the estimated eigenvalues of Π, $\lambda_i^{eigen}, i = 1, \ldots, n$.

Johansen's Likelihood Ratio (LR) statistic test is formulated to test the nested hypothesis:

$$H_0(r) : r = r_0; \qquad H_a(r) : r > r_0 \tag{4.4}$$

The likelihood ratio trace statistic, LR_{trace} is:

$$LR_{trace}(r_0) = -T \sum_{i=r_0}^{n} \ln(1 - \lambda_i^{eigen}) \tag{4.5}$$

The analysis was implemented using R-statistical software using the methodology outlined in Pfaff (2008).

4.1.2 Summary

The first study is an investigation of mortality trends and the factors driving the volatility of mortality using principal components analysis for a number of developed countries including Australia, Japan, Norway, UK and USA. It will give insight into the need for multiple factors for modeling mortality rates across all these countries. Conclusions will be drawn on about whether the basic structure of the Lee-Carter model adequately models the random variation and the full risk structure of mortality changes for these countries. Unit root tests conducted on the mortality trends to test if they are stochastic will be used as a starting point for using econometric models. Finally, this study tests for the existence of common trends and cointegrating relationships.

The proposed procedure for analysing mortality trends of multi-country data is outlined in Figure 4.11:

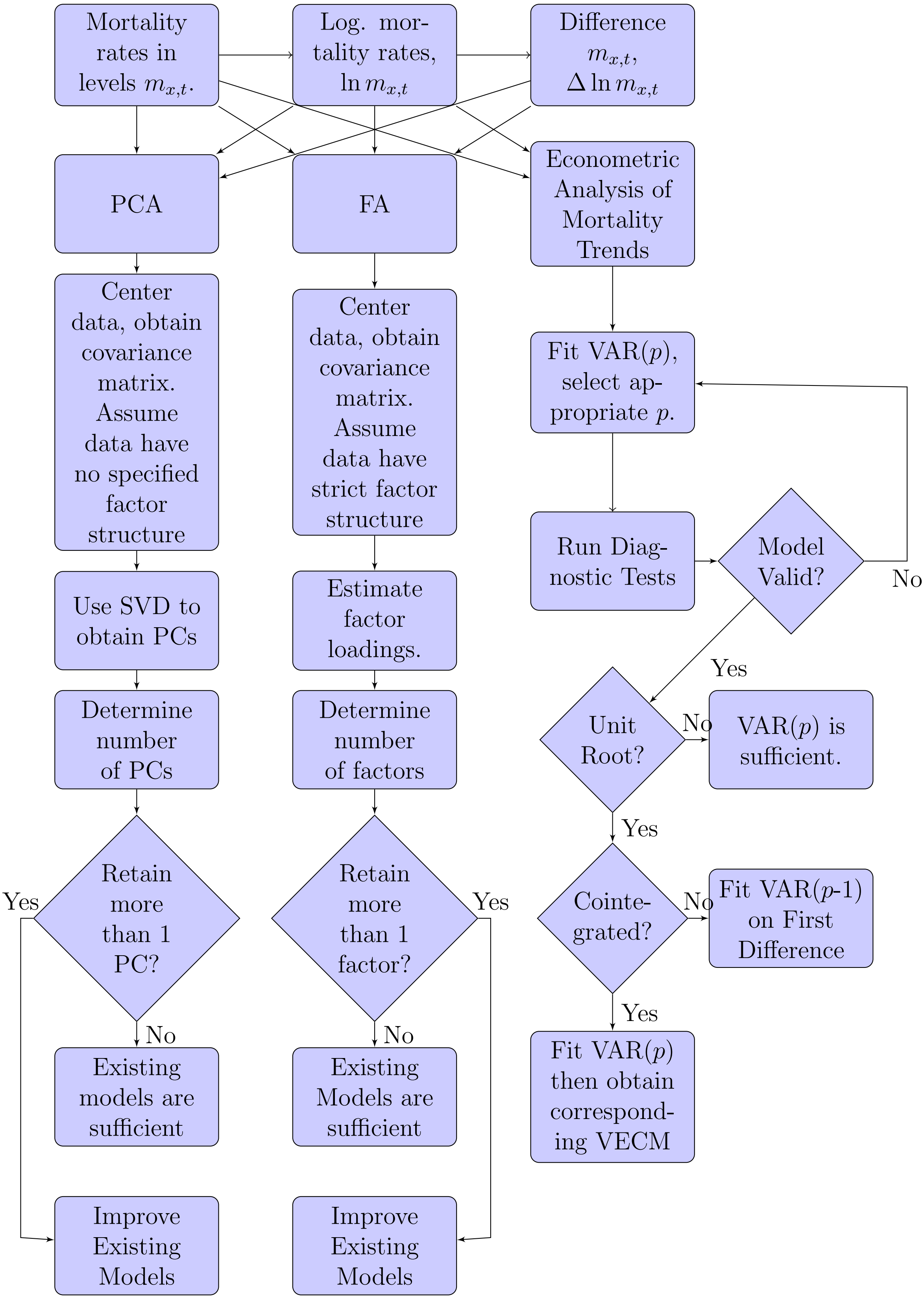


Figure 4.11: Flow chart illustrating the methodology of this thesis' 1st Study: Analysis of Mortality Trends

4.2 Analysis 2: A Two-Stage Bayesian Mortality Model

4.2.1 Modeling mortality using Econometric Techniques

Economic data is similar to mortality data. Using the properties of economic data mentioned in Brooks (2008) economic data and similar to mortality data because both types of data are reported annually. This also leads to the similar problem of small data sets. If data is collected annually for 25 years then only 25 data points will be available.

Economic data and mortality data are both less noisy than financial data (such as asset prices) which makes it is easier to separate random effects from existing underlying trends (patterns). Further, the variables are often from a non-Normally distribution yet belonging to a Normal distribution is a common assumption of most time series models (Brooks, 2008; Robertson and Tallman, 1999b).

These similarities motivated the techniques in this thesis where econometric techniques are used to analyse and project mortality data. The general idea is to simultaneously model the parameters from a parametric mortality model and capture their correlation using a parsimonious model.

From the literature in the previous sections, a key use of the Lee-Carter model is to extrapolate the long-run trends in mortality rates. In order to forecast accurately a data set that spans over quite a long historical period is required (Booth et al., 2006) and it is also necessary to use data from periods where the mortality patterns have similar structures. This is not always available or easy to determine. Also, in some populations such as the Australian population, what may be classified as a type of structural shift in mortality patterns was observed in the 1970's. In such a case, if a long data set is required, a technique that does not give more weight to more recent observations will be prone to bias since the older data is "irrelevant" (Booth et al., 2002b, 2006). Econometric techniques such as the VAR, VECM and BVAR extract and project long-run trends in mortality rates. The Bayesian VAR model in particular will be used to capture parameter risk and because the parameters are updated based on available information more weight is given to more recent observations. The procedure used to transform a static parametric mortality model into a dynamic parametric mortality model using a Bayesian-VAR is in the following sections.

4.2.2 Parametric Mortality Models

In order to reduce the number of random factors driving mortality changes over time a parameterized mortality model is cross sectionally estimated at a series of points in time and the evolution of the parameters is modeled as a VAR/VECM system. This not only reduces the dimension of the random variability but allows for smoothing across ages and improved forecasting performance of the model.

Following McNown and Rogers (1989) the eight parameter model proposed in Heligman and Pollard (1980) is used to model the probability of death of an individual aged x in the next year, q_x. Heligman-Pollard is appropriate because based on past observations it is reasonable to assume that the shape of mortality will persist for the foreseeable future. It is possible to use another parametric model.

4.2.2.1 Heligman-Pollard Model

The parameters of the Heligman-Pollard model are estimated at a series of points in time, t, and the evolution of the parameters is shown in figure 4.12. The model fits the data well and provides a consistent basis for smoothing across age. The fitted parameters are shown in 4.12. They show how the trends in the different mortality experiences for different ages have varied through time and also highlight the variability in the trends.

A_m and A_f are both declining. This implies that the approximate mortality of children at age 1 has been steadily declining. The gradient of the parameters is not as steep in recent years as it was between 1950 and 1970. This is one indicator of the change in the structure of the mortality patterns in the 1970s.

The difference between mortality at age one and mortality at age 0 as represented by B_m and B_f and the rate at which mortality decreases during childhood C_m and C_f are both fairly erratic with no distinct trend although the Bs are higher for more recent years while the Cs are lower.

The intensity of young adult mortality is reflected in D_m and D_f. D_m is consistently higher than D_f. It is noteworthy that between 1960 and 1970 D_m increased and started to decline from about 1980 and has generally been declining although there was a short period of increase in the mid to late 1990s. This is another indicator of the change in the structure of the mortality patterns in the 1970s.

An increase in D implies that mortality for the young adults had increased. E_m and E_f vary inversely with the spread of the young adult mortality hump. An increase in E means the hump is tight with little spread around the modal age while a decrease in E implies that hump is more spread around the modal age. Since the 1970s E_m and E_f have been declining although E_f has not been consistent. This

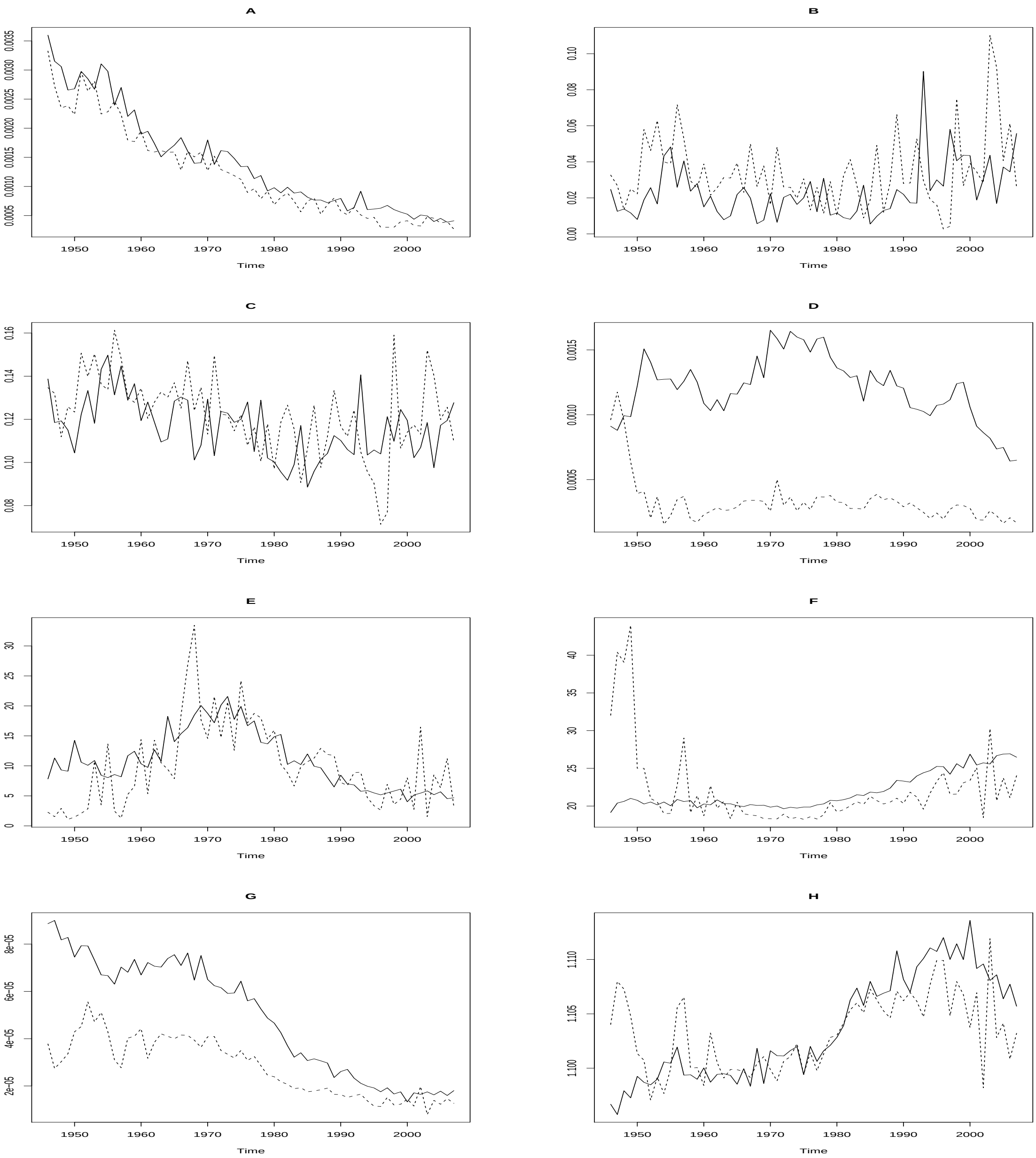


Figure 4.12: Heligman and Pollard Parameters for Australia Male and Females from 1921 to 2007. Solid Line=Males; Dotted Line=Females

implies that the spread of young adult mortality around the mortality hump has a wide spread. F_m and F_f is the location or modal age of the accident hump (and maternal hump for females). Before 1960, this location is higher for females than males but after 1965 F_f is lower than F_m. This implies that the accident hump for males occurs at a higher age than the accident and maternal hump for females. For Australian females, the location of the maternal hump is influenced greatly by policy changes such as the introduction of the baby bonus in 2002. The introduction of the paid parental leave scheme is also tipped to influence the location and severity of the maternal hump as an increase in the number of women having children during different stages of their careers will be observed.

G_m and G_f are the intercept of the Gompertz curve at age 0. They represent the base level of senescent mortality. G_f is lower than G_m with a narrowing gap. H_m and H_f are the slope of the Gompertz curve. They represent the rate of increase of G_m and G_f. Until the late 1990s, late life mortality for both males and females had been declining at an increasing rate. However, from 2000 late life mortality for both males and females has been declining at a decreasing rate. This is consistent with the findings of Risk Management Solutions medical-based longevity risk model that the level of mortality improvement experienced in the last 30 years is likely to tail off in 15 to 25 years (Risk Management Solutions, Inc. , 2010; Risk Management Solutions, Inc., 2010). This is a characteristic that is not easily captured when mortality is modeled without giving more weight to recent observations and leads to underestimating of mortality rates.

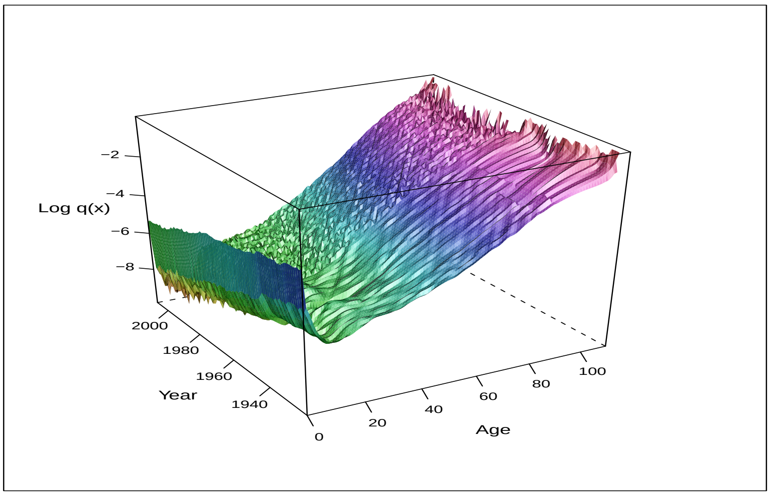
The Heligman-Pollard model fits mortality rates at a time point well and yields a sequence of annual estimates of parameters that forms a time series. Each of the parameters of the Heligman-Pollard model has a straightforward interpretation. This makes the model simple to explain.

Even though when looking to model old age mortality the main parameters of interest are those that are directly concerned with old ages it is also beneficial to look at the mortality of the other ages as there may some common factors that affect them as well. The result may be that some relationships may be identified for the variable of interest that involves the variables in the entire system. This is easily analysed by using econometric techniques.

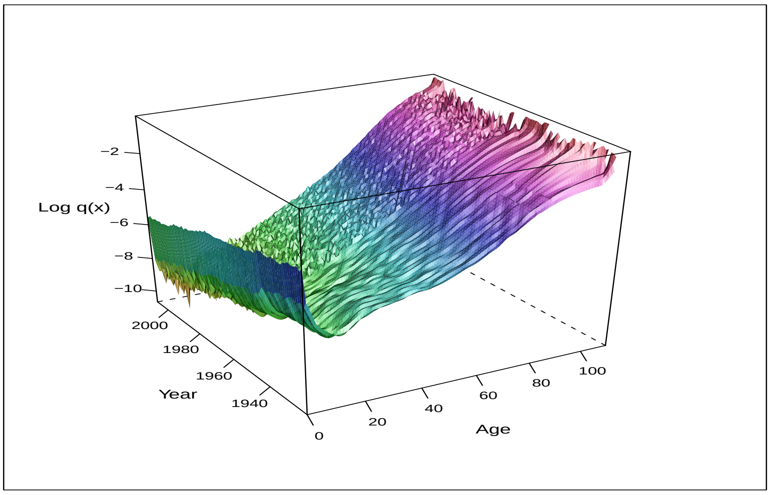
There is dependence in q_x that is inherited from the positive dependence between the components of θ. Let $\Theta = (\theta_{t_0}, \ldots, \theta_{t_T})$ be a matrix of the estimated parameters at times $t_0, \ldots, t_T$:

$$\Theta = \begin{pmatrix} A_{t_0} & A_{t_1} & \ldots & A_{t_T} \\ B_{t_0} & B_{t_1} & \ldots & B_{t_T} \\ & \vdots & \ddots & \\ H_{t_0} & H_{t_1} & \ldots & H_{t_T} \end{pmatrix} \tag{4.6}$$

Mortality rates have maintained the shape in figure 2.1 over long time periods. Parametric models that capture the shape mortality rate curves over time allow efficient dynamic models to be fitted to the data (Congdon, 1993).



(a) Log q_x for Australian Males



(b) Log q_x for Australian Females

Figure 4.13: The probability of death for Australian Males (left) and Females (right) using data from 1921 to 2007.

4.2.2.2 Limitations and Strengths of Similar Studies

McNown and Rogers (1989) use univariate time series models and the assume that the parameters are independent. The implications of this are:

1. Differencing the time series to achieve stationarity eliminates the trend (Box and Jenkins, 1976).

2. Forecasts from the process described in Rogers (1986) and McNown and Rogers (1989) are not accurate and are inconsistent (Lee, 1992; McNown and Rogers, 1992).

3. Confidence intervals are not easily computed (Lee, 1992)

The weights in equation (2.17) were used in Heligman and Pollard (1980) and are based on the assumption that the coefficient of variation is constant across age as it is assumed that the variance of $q_{x,t}$, denoted by $\sigma^2_{x,t}$, is directly proportional to $q^2_{x,t}$. Dellaportas et al. (2001) argue that equation (2.16) is too restrictive as it assumes a deterministic relationship between $q_{x,t}$ and θ_t and consequently not all the variation in $q_{x,t}$ is explained by θ_t. This requires assumptions about the distribution of $q_{x,t}$ and therefore Dellaportas et al. (2001) consider different probability distributions and apply Bayesian techniques to estimate θ_t. Although Dellaportas et al. (2001) consider the uncertainty in the estimated parameters at a point, the time series of the resulting estimated time series of the HP-parameters are not modelled.

These two studies, McNown and Rogers (1989) and Dellaportas et al. (2001), are both based on Heligman and Pollard (1980). McNown and Rogers (1989) recognizes that a dynamic parametric mortality model can be generated by using the Heligman-Pollard model and time series methods while Dellaportas et al. (2001) recognizes that there is uncertainty in estimation of HP-parameters.

In this thesis, a solution that generates a dynamic parametric mortality model and quantifies the uncertainty in estimation of HP-parameters is presented.

4.2.3 Quantification of uncertainties in the data, parameters of the model and the model using Bayesian Vector Autoregression Models

Recall equation (2.35). For the Heligman-Pollard parameters θ_t is $\theta_t = (A_t, B_t, \ldots, H_t)'$ with $n = 8$.

The unknown relationship that generates the multivariate time series is the evolution of q_x over time. When two variables simultaneously influence each other they are referred to as endogenous variables. Since all the parameters are explained within the mortality system the parameters of the Heligman-Pollard model are all treated as jointly endogenous variables. There is uncertainty regarding the underlying structure of the changes in mortality over time. Therefore the use of a VAR model to analyse and predict the evolution of the log-mortality rates or parameters of the Heligman-Pollard schedule of mortality is plausible. The past values of one of the variables in the mortality schedule may have a delayed effect on the values of another variable. When two variables simultaneously influence each other they are referred to as endogenous variables. This means that, for example, since all the parameters of the Heligman-Pollard model are explained within the mortality system the parameters will all be treated as endogenous variables. The VAR(p) models are more flexible than AR models and have a rich structure that captures more features of the mortality system as revealed by the parameters of the Heligman-Pollard model.

From a VAR model it is possible to analyse the impacts that the HP-parameters have on each other over time and how the HP-parameters respond to other unobservable factors. It is also possible to assess if knowing about the past values of some of the HP-parameters tell us about the future values of other HP-parameters. Key questions that may be asked include whether the HP-parameters that model the mortality schedule are related to each other, if the current and past values of a HP-parameter affect other HP-parameters, how a change in one HP-paramter relates to a change in another HP-parameter, if the change in the other HP-parameter is immediate or does it occur with a delay and, finally, how shocks to the mortality system affect the variables.

The procedure for estimation of a VAR model was outlined in section 4.1.1.2. This method is extended by using Bayesian Vector Autoregression Models as described in 2.3.4. Univariate methods (such as those used in McNown and Rogers (1989)) oversimplify the modelling process while unrestricted VAR models (such as those in Sherris and Gaille (2010a)) lead to over-fitting. Bayesian VARs provide a middle ground. Univariate models are unable to capture interactions between the variables and VAR models capture interactions between the variables but due to estimation by OLS some interactions are spurious.

Bayesian inference begins with specification of a prior distribution for use to analyse the data. The knowledge and the experience of the researcher are used to choose the prior distribution. Several books (such as Bauwens et al. (1999)) and papers (such as Robertson and Tallman (1999b); Brandt and Freeman (2006); Kadiyala and Karlsson (1997); Sims and Zha (1998); Joiner (2001); Sevinç and Ergun (2009)) provide knowledge about the different priors available.

Denote the sampling density of ϑ as $p(\vartheta|\eta, \Sigma)$. This sampling density can be broken up into two parts. Firstly, the distribution for η given Σ:

$$\eta|\Sigma, \vartheta \sim N(\eta^{OLS}, \Sigma \otimes (X'X)^{-1}) \tag{4.7}$$

Secondly, $p(\vartheta|\eta, \Sigma)$ also has a part where Σ^{-1} has a Wishart distribution:

$$\Sigma^{-1}|\vartheta \sim W([(\vartheta - Xv^{OLS})'(\vartheta - Xv^{OLS})]^{-1}, T - (1 + np) - n - 1) \tag{4.8}$$

The Litterman (or Minnesota) prior (Litterman, 1986) is commonly used in

Bayesian VAR analysis Sevinç and Ergun (2009); Kadiyala and Karlsson (1997). However, in this analysis the Sims-Zha (or Normal-Inverse Wishart) prior (Sims and Zha, 1998) is used. Both the Litterman prior and the Sims-Zha prior are informative priors that produce closed form posteriors (Brandt and Freeman, 2006) but the Sims-Zha prior is selected because studies such as Robertson and Tallman (1999b) have shown that its provisions for unit roots and cointegration improve the performance of forecasts in systems that are based on non-stationary variables. A (multivariate) normal distribution for the prior of the variables, θ, is advantageous because it is a rich and flexible family of priors; specifically it enables the researcher to specify varying degrees of strength in the beliefs of the prior location hyper parameters (Sinay, 2008; Kadiyala and Karlsson, 1997). The Inverse Wishart distribution is a natural candidate for the prior of Σ, the covariance matrix of the residuals of the VAR (Bauwens et al., 1999). It is a natural candidate because it is the conjugate prior for the covariance matrix of a multivariate normal distribution. The Inverse Wishart distribution is also easy to use. In comparison, the Litterman prior assumes an independent normal prior (Kadiyala and Karlsson, 1997) and that all variables follow a random walk with drift. Further, the fact that it is specified equation-by-equation means it is not a proper prior for the VAR model (Brandt and Freeman, 2006). Modifications of the Litterman prior have been published, in Ni and Sun (2003) for example, where the Litterman prior is analysed in different forms. Ultimately, the recommendation for choice of prior in Kadiyala and Karlsson (1997)[1] forms the basis of the decision to use the Sims-Zha (Normal-Inverse Wishart) prior in this analysis.

Under the Normal-Wishart Prior, the posterior distribution is also Normal-Wishart and the posterior mean or point estimate of υ is:

$$\upsilon^{NW} = (\overline{\mathcal{H}}^{-1} + X'X)^{-1}(\overline{\mathcal{H}}^{-1}\overline{\upsilon} + X'\theta) \tag{4.9}$$

The estimate of the error covariance matrix, Σ is:

$$\Sigma^{OLS} = T^{-1}(\theta'\theta - \upsilon'^{NW}(X'X + \overline{\mathcal{H}}^{-1})\upsilon^{NW}) + \overline{\upsilon}'\overline{\mathcal{H}}^{-1}\overline{\upsilon} + \overline{S} \tag{4.10}$$

$\overline{S}$ is a diagonal matrix with elements $((\frac{\sigma_1}{\lambda_0^h})^2, (\frac{\sigma_2}{\lambda_0^h})^2, (\frac{\sigma_3}{\lambda_0^h})^2)$

[1]From Kadiyala and Karlsson (1997) the Normal-Inverse Wishart prior should be chosen over the Litterman prior when "the prior beliefs are of the Litterman type" such that the amount of prior information needs to regulated and computational effort is of major importance

Sims-Zha's prior standard deviations for the ω_{ij}^{l} are specified as:

$$\sqrt{\psi_{lij}} = \frac{\lambda_0^h \lambda_1^h}{\sigma_j l^{\lambda_3^h}} \tag{4.11}$$

$\overline{\mathcal{H}}$ has diagonal elements $(\lambda_0^h \lambda_4^h)^2, (\frac{\lambda_0^h \lambda_1^h}{\sigma_1})^2, (\frac{\lambda_0^h \lambda_1^h}{\sigma_2})^2, (\frac{\lambda_0^h \lambda_1^h}{\sigma_3})^2, (\frac{\lambda_0^h \lambda_1^h}{\sigma_1 2^{\lambda_3^h}})^2, (\frac{\lambda_0^h \lambda_1^h}{\sigma_2 2^{\lambda_3^h}})^2, (\frac{\lambda_0^h \lambda_1^h}{\sigma_3 2^{\lambda_3^h}})^2)$

$\overline{\mathcal{H}} \otimes \overline{S}$ gives a matrix with elements equal to the variances under Litterman's Prior when $\lambda_2^h = 1$. The Normal-Wishart Prior makes λ_2^h redundant because the scale matrix of the inverse Wishart distribution, $\overline{S}$ has the constant weight, λ_0^h in its denominator and prevents the prior from distinguishing between the lags of the i-th variable and those of the other variables in the i-th equation.

The basic VAR(p) in levels model assumes that the series θ_t are stationary (Hamilton, 1994). Many economic time series are non-stationary and in some cases cointegrated. In order to incorporate this into the Bayesian VAR models, Sims and Zha (1998) added two hyperparameters that incorporated non-stationarity and cointegration into the priors. The first hyperparameter is based on Doan et al. (1984) which restricts the sum of the coefficients on the lagged values of the dependent variable in each equation to 1 (i.e. $\sum_{l=1}^{p} \Omega_l = I_n$) and the sum of the coefficients on each of the other variables to 0 in a BVAR(p) in first differences. This restriction is referred to as the sum of coefficients prior.

From Equation (2.41) this implies that the BVAR in first differences can be written as:

$$\Delta\theta_t = c + \sum_{k=1}^{p-1} \Gamma_k \Delta\theta_{t-k} + \epsilon_t, \qquad k = 1, \ldots, p-1 \tag{4.12}$$

since $\mathbf{\Pi}\theta_{t-1} = \sum_{l=1}^{p} \Omega_l - I_n = 0$ because $\sum_{l=1}^{p} \Omega_l = I_n$ under the sum of coefficients prior.

Eliminating the Π term implies that there is no cointegration but there are non-stationary variables in the system. The sum of coefficients prior is controlled by using the hyperparameter, $\mu_5^h \geq 0$. Sims and Zha (1998) assume that a good forecast of θ_i at some point $t = \tau$ is the average of the lagged values of θ_i, denoted as $\overline{\theta}_i$, but $\overline{\theta}_i$ does not help to predict the values of θ_j for $j \neq i$. They introduced n dummy variables, $\overline{\theta}_i^0$ at the start of each data set where $\overline{\theta}_i^0$ is the mean of the first p values of the i-th variable. μ_5^h assigns weights to the $\overline{\theta}_i^0$ such that a large $\overline{\theta}_i^0$ implies a high likelihood of $\Pi = 0$ such that in the long run the values of $\theta_i^0 \to \overline{\theta}_i^0$ (Doan et al., 1984; Sims and Zha, 1998; Robertson and Tallman, 1999b; Summers, 2001).

The weighted dummy observations for the case n=3 and $p = 2$ is:

$$\begin{bmatrix} 0 & \mu_5^h\overline{\theta}_1^0 & 0 & 0 & \mu_5^h\overline{\theta}_1^0 & 0 & 0 \\ 0 & 0 & \mu_5^h\overline{\theta}_2^0 & 0 & 0 & \mu_5^h\overline{\theta}_2^0 & 0 \\ 0 & 0 & 0 & \mu_5^h\overline{\theta}_3^0 & 0 & 0 & \mu_5^h\overline{\theta}_3^0 \end{bmatrix} \tag{4.13}$$

The sum of coefficients prior implies that as $\mu_5^h \to \infty$, $\Pi \to 0$ the number of unit roots is the same as the number of variables in the system and there is no cointegration. The additional hyperparameter μ_6^h allows the BVAR model to incorporate cointegration by assuming there is one cointegrating relation and sets the value of the constant $c_i = 1$. The matrix of the weighted dummy observations for $n = 3$ and $p = 2$ is:

$$\begin{bmatrix} \mu_6^h & \mu_6^h\overline{\theta}_1^0 & \mu_6^h\overline{\theta}_2^0 & \mu_6^h\overline{\theta}_3^0 & \mu_6^h\overline{\theta}_1^0 & \mu_6^h\overline{\theta}_2^0 & \mu_6^h\overline{\theta}_3^0 \\ 1 & \theta_{12} & \theta_{22} & \theta_{32} & \theta_{11} & \theta_{21} & \theta_{31} \\ 1 & \theta_{13} & \theta_{23} & \theta_{33} & \theta_{12} & \theta_{22} & \theta_{32} \end{bmatrix} \tag{4.14}$$

Sims-Zha's prior has a hyperparameter λ_0^h for the overall tightness of the standard deviation of the errors ϵ_{it} and their intercorrelations. This prior also allows for non-stationary time series and cointegrated relations.

This prior is suitable for modeling the parameters of the Heligman-Pollard model through time and this is the prior used in this thesis to implement the HP-BVAR model.

$\lambda_0([0, 1])$ controls the overall tightness of the prior on the error covariance matrix. As it increases the coefficients have increased variance in the structural form such that the model strays further from a random walk.

$\lambda_1(> 0)$ specifies how tight the random walk prior specification is. As $\lambda_1 \to 0$ the diagonal elements of the coefficient matrix for the first lag tend to one and all other elements tend to zero. This restriction is implemented only on the lagged matrices. As the value increases the random walk prior will not be enforced as strictly and the model will stray further from a random walk in the lags.

Increasing the value of $\lambda_3(> 0)$ shrinks the coefficients of higher order lags to zero and by allowing the parameters contained in these lags to vary less around their conditional mean of zero.

λ_4 controls the tightness of the prior on the constants and as $\lambda_4 \to 0$ the constants tend to zero. It is conditional on λ_0.

μ_5 controls the unit root prior. As it increases the likelihood that the model can be expressed in first differences also increases.

μ_6 controls the co-integration prior such that as it increases co-integration will be given more weight.

Choice of values for the hyperparameters is determined by obtaining values from previous studies or by evaluating the forecast performance of the model over a range of hyperparameters (Joiner, 2001).

The choice of prior used in the Bayesian VAR in this model is such that a closed form posterior is obtained. An advantage of analytic (including closed form) posteriors is that they allow Gibbs sampling to be used. A Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990) is a special case of the Metropolis-Hastings algorithm (Hastings, 1970) and is used to to draw samples from the joint posterior distribution. In the Gibbs sampler algorithm, several iterations are performed where each iteration cycles through the unknown parameters. A sample of one parameter conditioned on the latest values of all the other parameters is drawn. After a sufficiently large number of iterations the samples that have been drawn of one parameter represent its marginal posterior distribution.

Combining tight priors on the long-run properties of the VAR through the hyper parameters leads to improved accuracy in forecasts from the BVAR (Robertson and Tallman, 1999b). By extension, the hyper-parameters are used to provide an easy and accurate assessment of uncertainty. In the unrestricted VAR the forecasted variables attain equilibrium faster than those of the Bayesian VAR (Robertson and Tallman, 1999b). This leads to an observation that the VAR model shows an initial increase in parameter uncertainty that settles down to a long run distribution for each of the parameters. VAR models look for persistence (long-run or permanent movements) in the data (Sims et al., 1990) therefore it is possible to have inaccurate results if some of the variables exhibit high persistence (Stock, 2001). After an initial time period the parameter risk would be considered as having reached its maximum. In contrast, the BVAR model shows a significantly higher level of parameter risk than the VAR and over the same horizon will not reached a long run distribution for the parameters (Robertson and Tallman, 1999a). The BVAR checks if the effect of a shock is permanent or fades with time (Sims and Zha, 1998). Therefore, a VAR model is not suited for long term forecasts as it often produces erratic forecasts (Sims and Zha, 1998).

For this study, all the analysis will be performed using R, S+ and S+Finmetrics. In R , a specialized packages for econometric methods include vars (Pfaff, 2008) and MSBVAR (Brandt, 2011). The software is well developed with detailed manuals and implementation of econometric models is not difficult once the theory is understood.

4.2.4 Summary

In the second study, the parameters in the Heligman-Pollard model are modeled as a multivariate time series system in an age-period mortality model. The correlation between the Heligman and Pollard model parameters is captured in the econometric methods allowing interaction between changes in mortality for differing ages. A vector autoregression (VAR) model is used for the relationship between the past (lagged) values and current values of the Heligman-Pollard parameters. The classical unrestricted VAR model does not account for uncertainty in its coefficients. A Bayesian VAR model accounts for the uncertainty in the VAR coefficients allowing quantification of parameter risk. These models will be referred to as the HP-VAR and HP-BVAR models respectively. Models.

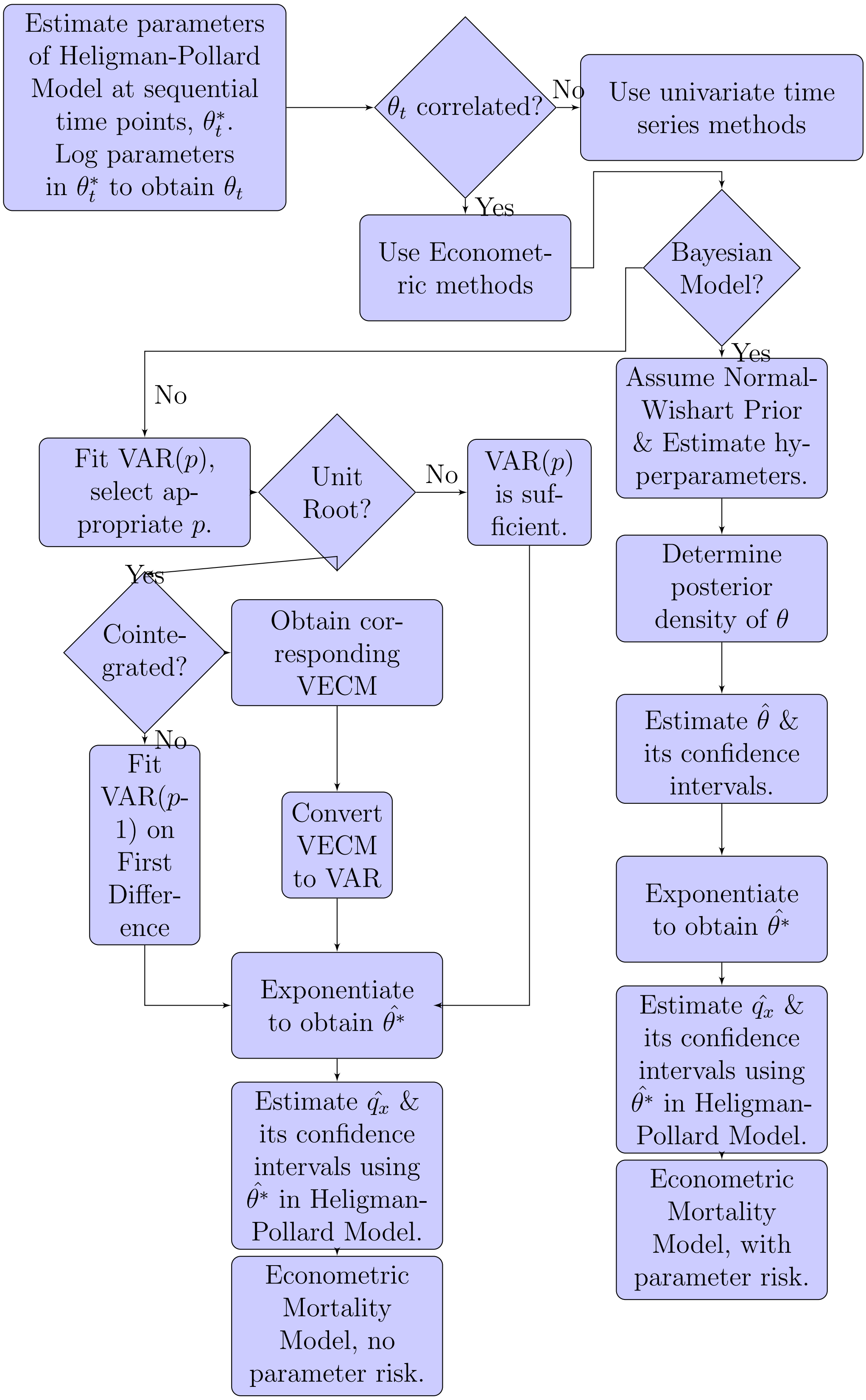


Figure 4.14: Flow chart illustrating the methodology of this thesis' 2nd Study: Modeling Mortality with a Bayesian Vector Autoregression.

4.3 Analysis 3: Application to Insurer Longevity Risk Based Capital Stress Margins

The purpose of APRA-specified margins is to either simplify the process, to overcome the problem of scant data and ensure consistency that would be lacking if individual insurers were allowed to set their own margin or to set margins which APRA considers should be the same across industry. APRA (2010b) outlines stress margins as specified by APRA.

4.3.1 Current Proposed Simplifications for Computing the Longevity Stress Margin

The APRA-specified longevity stress margin is currently proposed as "A (permanent) 25 per cent decrease in mortality rates for each age". This simplification has shortcomings in its structure and its magnitude. It is important to acknowledge that this stress is similar to the quantity proposed in Solvency II in CEIOPS (2007). The initial formulation of the longevity risk capital charge under Solvency II calibrated the permanent decline in mortality that was consistent with the 99.5% probability of having sufficient funds to meet liabilities as 25%. The Committee of European Insurance and Occupational Pensions Supervisors (CEIOPS) Consultation paper 49 attempted to justify why the permanent decrease was set to 25%. This percentage change in mortality was estimated using UK insurance companies' data (CEIOPS, 2007). Several articles including Börger (2010); Plat (2010); CEIOPS (2008) and in particular the European insurance and reinsurance federation (CEA) criticised the amount as too high and the approach as too simplistic. Eventually the stress was recalculated and reduced by 5% and currently stands at 20%.

The main shortcoming of the specified LSM after analysis in publications like Börger (2010) is its structure and an age-dependent stress is suggested as a more appropriate alternative. The structure of the specified LSM that all ages are subject to the same constant percentage decline means that for higher mortality rates the value of the shocks will be larger (eg 25% of 0.1 when compared to 25% of 0.01). According to Börger (2010), it would be more realistic to have smaller relative reductions for old ages because the elderly are subject to more causes of death and an improvement in mortality due to one cause would have a smaller impact on mortality due to all causes. However, Willets (2004) describes a phenomenon where "The ages showing the greatest (average annual) improvements are steadily moving upwards." Willets (2004) analysed England and Wales mortality from 1901 to 2001 and concluded that despite this phenomenon the effect of cause specific mortality improvements still reduces progressively with advancing age. This is in agreement with the observation in Blake et al. (2008) that as age increases there is greater uncertainty regarding longevity. Therefore, an age dependent stress with the percentage decline in mortality increasing with age is expected to be an appropriate alternative to the constant percentage decline across all ages.

The distribution of survival probabilities is obtained from a suitable stochastic mortality model which quantifies the level, trend and volatility of mortality improvements over time. A stochastic mortality model is required in order to calculate the quantiles of the distribution of survival probabilities. It is important that the stochastic mortality model portrays two important components of longevity risk (Börger, 2010; Plat, 2010). First, it must recognize that the next year's realised mortality rates may deviate from the expected rates. Secondly, it must anticipate that the realised mortality rates beyond the next year may also deviate from the expected rates. Mortality models that assume a fixed mortality trend fail to capture the second component (Börger, 2010) and consequently lead to over/undercapitalization.

Stressing the scenario of a downward shock to mortality rates requires consideration of how mortality improvements occur in reality. This research considers the population data from the Human Mortality Database [1] for Australian males from 1946-2007. Mortality improvements change with time and are not constant by age. The structure of a shock to mortality rates should reflect this.

Average mortality improvement for Australian Males by age is shown in figure 4.15. The black solid line represents the average annual mortality improvements between 1921 and 1950, the blue dot-dash line between 1951 and 1980 and the red dashed line between 1981 and 2007. Between 1921 and 1950 the greatest improvements were for children and adults aged 25 to 50, while from 1951 to 1980 although children had greater mortality improvements mortality at all ages except the very elderly does not vary by great amounts. Mortality improvements for the every elderly are however smaller in comparison. For the period between 1981 and 2007 average mortality improvements are highest for children and adults aged 50 to 85. The magnitude of the mortality improvements is greater in 1981-2007 than in 1951-1980 for most ages.

The implication of the persistence of the pattern of increased age at which greatest mortality improvements occur is that higher amounts of benefits will be paid out by pension funds and annuity providers to retirees because the number of surviving annuitants will increase and possible be greater than expected.

[1] www.mortality.org

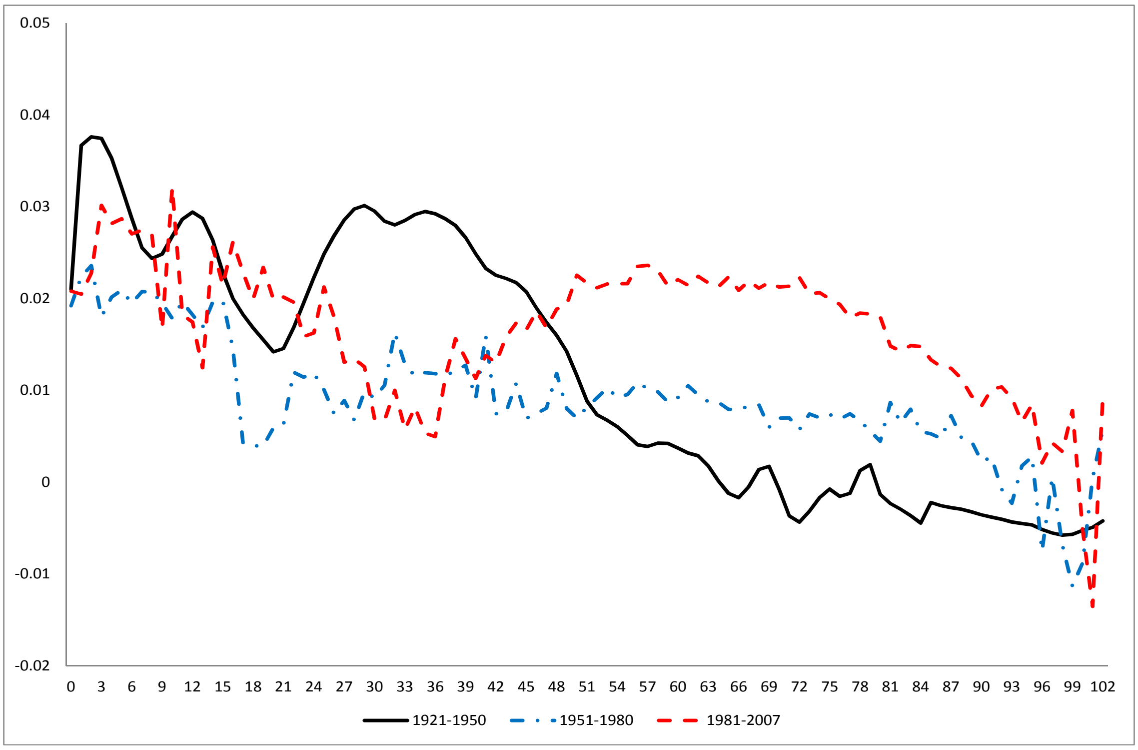


Figure 4.15: Average annual mortality improvements for Australian Males. The age at which the greatest mortality improvement is experienced is increasing with time. Mortality at very advanced ages (90+) changes less than at younger ages and the change is either an improvement or worsening of mortality as it fluctuates around zero.

From past experience as evidenced in figure 4.15 it is not realistic to assume that mortality will decline by the same constant percentage at all ages. The magnitude of mortality improvement differs by age and time period.

Assume that mortality improvements are indeed constant across all ages. Calibrating the magnitude, γ, of the percentage decrease is problematic because the effects of medical advancements and health policies on mortality rates are difficult to quantify in a standard way across all ages.

4.3.2 Method for Calibrating the LSM

CEIOPS (2009) mentions a number of different ways in which the longevity risk capital charge for Solvency II can be captured such as *a*) a reduction in base mortality rates *b*) using improvement factors or *c*) a combination of *a*) and *b*). A scenario based stress that tested the scenario of a permanent 25% (revised to 20% in CEIOPS (2010)) decrease in mortality rates is used to calculate the capital requirement of the longevity risk sub-module. This is a form of method *a*).

It has been suggested that the longevity stress margin should be calibrated differently, and that an age and duration dependent treatment of longevity should be considered. Several justifications are given by CEIOPS for retaining the permanent 25% decrease in mortality rates as the stress. First, it is simple. Secondly, when shocks were differentiated by duration, the shocks for different durations were small and not monotone. Thirdly, when shocks were differentiated by age portfolios of products subject to longevity risk were "generally heavily weighted in favour of older age groups". Finally, it stated that sufficient reliable data was not available and therefore granular calibration was not possible. This simplification assumes that both men and women experience the same mortality shocks.

In CEIOPS (2009) it is acknowledged that results of the stochastic model of future mortality improvements implied a lower stress. However, due to the significant degree of uncertainty in mortality modeling more weight is attached to the analysis of historic improvements. Without giving any detailed justification a lower stress is used in CEIOPS (2010).

Therefore, as a first step in analysing the suitability of the calibration of the Australian longevity stress margin, it is necessary to analyse the historic mortality improvements of the Australian population. In the previous section, it was shown that historical mortality improvements did not occur at a constant rate. Assuming that future mortality improvements will occur at a constant mortality rate will affect the amount of capital that will be sufficient to meet expected liabilities of pension funds and annuity providers.

The second step in analysing the suitability of the LSM is using a stochastic mortality model. Under the assumption that the LSM is a quantity such that the liabilities due to longevity risk are met 99.5% of the time in a one year horizon:

$$LSM^{VaR} = \arg\min_{x} \left\{ \mathbb{P}\left[(NAV_0 - NAV_1 * v) > x \right] \leq 0.005 \right\} \tag{4.15}$$

Under the APRA-specified simplification the LSM is:

$$LSM^{SHOCK} = (NAV_0 - NAV_1 * v | \text{Longevity Shock}) \tag{4.16}$$

The capital requirement, LSM^{SHOCK} is calculated as the change in net asset value (assets minus liabilities) following a permanent decrease in mortality rates.

For insurance companies and pension funds, longevity risk is quantified using the probability distribution of the present value of future payments in a portfolio of annuities. In a risk based framework, capital requirements for uncertain future liabilities are calculated using a quantile risk measure Dellinger (2006).

The value of the best estimate of liabilities (BEL) is defined as the expected present value of all future payments:

$$BEL_t = \sum_{j=0}^{\tau} E_t[L_{t+j}]P_t^{(j)} \tag{4.17}$$

where $E_t[]$ is the expectation given the information available at time t, L_{t+j} is the amount payable at time $t+j$ and $P_t^{(j)}$ denotes the price at time t of a zero coupon bond that matures at time $t+j$. APRA regards the zero coupon spot yield curve of Commonwealth Government Securities (CGS) as the best proxy for risk free rates for Australian-denominated liabilities APRA (2010b). The actuarial present value of a life annuity that pays \$1 at the end of each year discounting at the risk-free rate,r^f, is $a_x = \sum_{k=1}^{\infty} v_k^k p_x$ where $v = \frac{1}{1+r^f}$.

The life insurance prudential standards for the valuation of liabilities and the determination of regulatory capital require the discounting of future cash flows using risk-free discount rates APRA (2010b). APRA proposes that the Risk-Free Best Estimate of Liabilities (RFBEL) be determined as per the Best Estimate Liability calculated under LPS 1.04 (that is, Best Estimate Liability = Value of expected future benefit payments plus Value of expected future expenses) but with the gross investment yield and liability discount rate set equal to the risk-free discount rate. The longevity capital charge will be the sum of the Best estimate of longevity liabilities and the longevity stress margin.

Consider a portfolio of one cohort of immediate life annuities where the annuitants are all aged x_0 when the contract begins at time t_0. In this case the single premium calculated by the expected value of the life annuity is a reasonable measure of the cost of longevity for each annuitant in the portfolio.

N_t, $t = 0, 1, \ldots, \omega - x_0$ is a random number of annuitants alive at time t, with $N_0 = n_0$ being the initial size of the portfolio. Let $T^j_{x_0\tau}$ represent the remaining lifetime of the j^{th} annuitant aged x_0. $\Pi_t = \{j | T^j_{x_0\tau} > t\}$ defines the in-force portfolio at time t.

Denote the mathematical reserve at time t of a life annuity that pays \$1 for an individual annuitant as:

$$V_t = a_{x+t} \tag{4.18}$$

The change in the mathematical reserve from year t to $t+1$ is trivially due to interest earned on the reserve less the annual benefits paid. The value increases as a result of the mutuality effect (financial gains because future benefits for annuitants in the portfolio who died between t and $t+1$ are retained in the portfolio).

$$V_{t+1} - V_t = \text{Interest} + \text{Mutuality Effect - Annual Benefits Paid} \tag{4.19}$$

The initial technical provisions for annuitant j are $V_{t=0}^{j} = S$, where S is the single premium. For the portfolio, the reserve using the best-estimate assumptions is $V_t^{(\Pi)[BE]}$.

The total reserve for portfolio of annuities, V_t^{Π} is the sum of the Best Estimate of Liabilities, $V_t^{(\Pi)[BE]}$, and the risk margin, RM_t of adverse future scenarios Olivieri and Pitacco (2008).

$$V_t^{(\Pi)} = V_t^{(\Pi)[BE]} + RM_t \tag{4.20}$$

The assets available to meet the risks is $M_t = A_t - V_t^{(\Pi)[BE]}$ where A_t is the total amount of assets. In this case, $NAV_t = A_t - V_t^{(\Pi)[BE]}$.

For immediate life annuities, the Longevity stress margin, LSM, is $\Delta NAV_t = V_t^{(\Pi)[-\gamma]} - V_t^{(\Pi)[BE]}$ where $V_t^{(\Pi)[-\gamma]}$ is the expected value of future benefits calculated using mortality rates that are $\gamma\%$ below the best estimated mortality rates.

As explained in detail in Börger (2010) the longevity stress margin is based on the assumption that

$$LSM^{SHOCK} = (BEL_0|\text{Longevity Shock}) - BEL_0 = (BEL_0|Shock^{VaR}) - BEL_0 = LSM^{VaR} \tag{4.21}$$

where Shock $= -\gamma\%$ is the percentage decline in mortality rates that is constant across all ages and VaR = 99.5% VaR from forecasted mortality rates. Different magnitudes of γ will yield different values of LSM^{SHOCK}. The APRA-specified magnitude, $\gamma = 25\%$ is assumed to be equivalent to the longevity event which occurs with 0.05% probability. If the two resulting quantities, LSM^{SHOCK} and LSM^{VaR} are significantly different, then the simplification has not achieved its goal - offering a simple, standard formula for calculating the stress event while capturing the true underlying risk.

The Risk Margin at time t is:

$$RM_t = \sum_{h=0}^{Maturity} CoC * LSM_{t+h}(1+r_f)^h \tag{4.22}$$

From equation (4.22) the LSM needs to be computed to realistically reflect the uncertainty in mortality rates or else its errors will be magnified in the computation of the risk margin leading to over/under-capitalization.

The sensitivity of different products to changes in mortality rates should be considered when the capital requirements are being determined. A product that is less sensitive to changes in mortality rates will require less capital for meeting liabilities due to adverse longevity experiences. Coughlan et al. (2007) defines a measure of the sensitivity of an underlying exposure to changes in mortality called q-duration. It is analogous the interest rate duration. It is the change in value of liabilities, VL, due to a unit (percentage) change in mortality rates (Plat, 2010; Wang et al., 2010; Tsai et al., 2010). For a life annuity the q-duration, qD, is calculated as:

$$qD = -\frac{VL_{+} - VL_{-}}{2(VL)\Delta q} = \frac{VL_{-} - VL_{+}}{2(VL)\Delta q} \tag{4.23}$$

where Δq is the change in mortality, VL_{+} and VL_{-} are the liability values at mortality rate $q + \Delta q$ and $q - \Delta q$ respectively.

This quantity is an important measure of the significance of a longevity insurance product to changes in mortality. The current APRA specification treats all life annuities as one product (L3 - Annuity with Longevity Risk) and does not consider the duration of the product to be a significant factor.

In this thesis, the adequacy of the magnitude and structure of the APRA-specified simplification is investigated. The main difference between the investigation in this thesis and that in Börger (2010) is that the underlying model used is also developed in this thesis. Further, a different perspective of the behavior of uncertainty in mortality at older ages is taken in this thesis. Another difference is Börger (2010) suggests that a more appropriate alternative to the constant decline mortality stress is an age-dependent stress with smaller relative reductions for old ages. In light of the findings in Blake et al. (2008) this thesis takes a different approach that will be supported by the results of the model developed and considers that survivorship uncertainty increases with age such that an age-dependent stress with greater mortality reductions for old ages is tested as a more appropriate alternative.

4.3.3 Summary

The suitability of the magnitude and structure of the APRA-specified longevity stress margin is investigated in the third analysis. In particular, there is potential for over or under-capitalization due to a mismatching of the one in two-hundred year event and the prescribed shock. The adequacy of the specified longevity stress scenario is the main concern addressed in the third study in this thesis.

5

An Analysis of Cross-Country Mortality Trends using Econometric Techniques

5.1 Introduction

A sound understanding of historical mortality experience is fundamental to the development of a mortality model. In this chapter, past mortality rates are studied in order to understand how mortality behaves. This chapter investigates trends, including common trends through co-integration, and the factors driving the volatility of mortality using principal components analysis for a number of developed countries including Australia, Japan, Norway, UK and USA.

Similar studies have been done in publications including Willets (2004), Sherris and Gaille (2010b) and Booth et al. (2002a). Willets (2004) analysed mortality in the twentieth century and analysed patterns of mortality improvements for all cause and cause-specific mortality rates but the evaluation had a minimal amount of statistical analysis. In comparison, Booth et al. (2002a) analyse mortality trends and consider the number of principal components that are needed to explain the age-time interactions of Australian mortality and extend the Lee-Carter model. On the other hand, Andreev and Vaupel (2005) performs a limited analysis of the surface of period trends that does not look at the factors that drive them. Recently, Sherris and Gaille (2010b) does a good econometric analysis cause-specific mortality rates for common trends in all-cause mortality rates on a cross-country basis. In this chapter each of these studies is extended as the forces, observable and unobservable, that drive mortality trends, time trends and cohort trends of all-cause mortality rates are comprehensively analysed using statistical and econometric techniques on a broad multi-country basis.

The purpose of this analysis is to investigate if multiple principal components and factors are needed for modelling mortality rates, time trends and cohort trends across all these countries. Previous studies do not investigate the factors or principal components that drive time trends and cohort trends. Three main techniques are used: factor analysis, principal component analysis and cointegration analysis. Dimension reduction through factor analysis will reveal unobservable variables that explain the correlation between observable variables in contrast to dimension reduction through principal component analysis which reveals linear combinations of the observed variables that explain the variation in the data. Next, unit root tests on mortality trends by country are used to test if they are stochastic. These results are significant because variables with stochastic trends are potentially cointegrated and when PCA is performed on cointegrated variables the first principal component is the common stochastic trend Alexander (1999). The next analysis involves checking for the existence of cross-country common trends and cointegrating relationships in countries that have comparable living conditions. The implications of factors, principal components and common trends in mortality models are discussed based on the range of models developed in this chapter. This study is based on Sherris and Njenga (2009).

5.2 Background Information

It is common practice to analyse and model the logarithms of mortality rates. This serves to incorporate the high variation old age mortality rates (Shang et al., 2010). It also keeps the projected mortality rates positive. However, this thesis takes an additional approach and investigates the differences of the logarithms of the mortality rates, $\Delta_h \ln m_{x,t}$ and $\Delta_d \ln m_{x,t}$, defined in section 4.1.1. In figures 5.1, 5.2 and 5.3 the data is arranged with the youngest ages at the bottom (the oldest ages at the top) and is classified into 4 categories for each time series, each representing a quartile of the data. The pale yellow/cream areas represent the 25% (0%-25%) of the data with the smallest magnitude, the dark yellow areas represent the second quartile of the data (the 25% (25%-50%) of the data with the second smallest magnitude), the orange areas represent the third quartile of the data (the 25% (50%-75%) of the data with the second largest magnitude) and the red areas represent the fourth quartile of the data (the 25% (75%-100%) of the data with the largest magnitude). When viewed in black and white the darkness of the shading reduces as magnitude decreases as the darkest shading represents the quartile with greatest magnitude and the lightest shading represents the quartile with smallest magnitude.

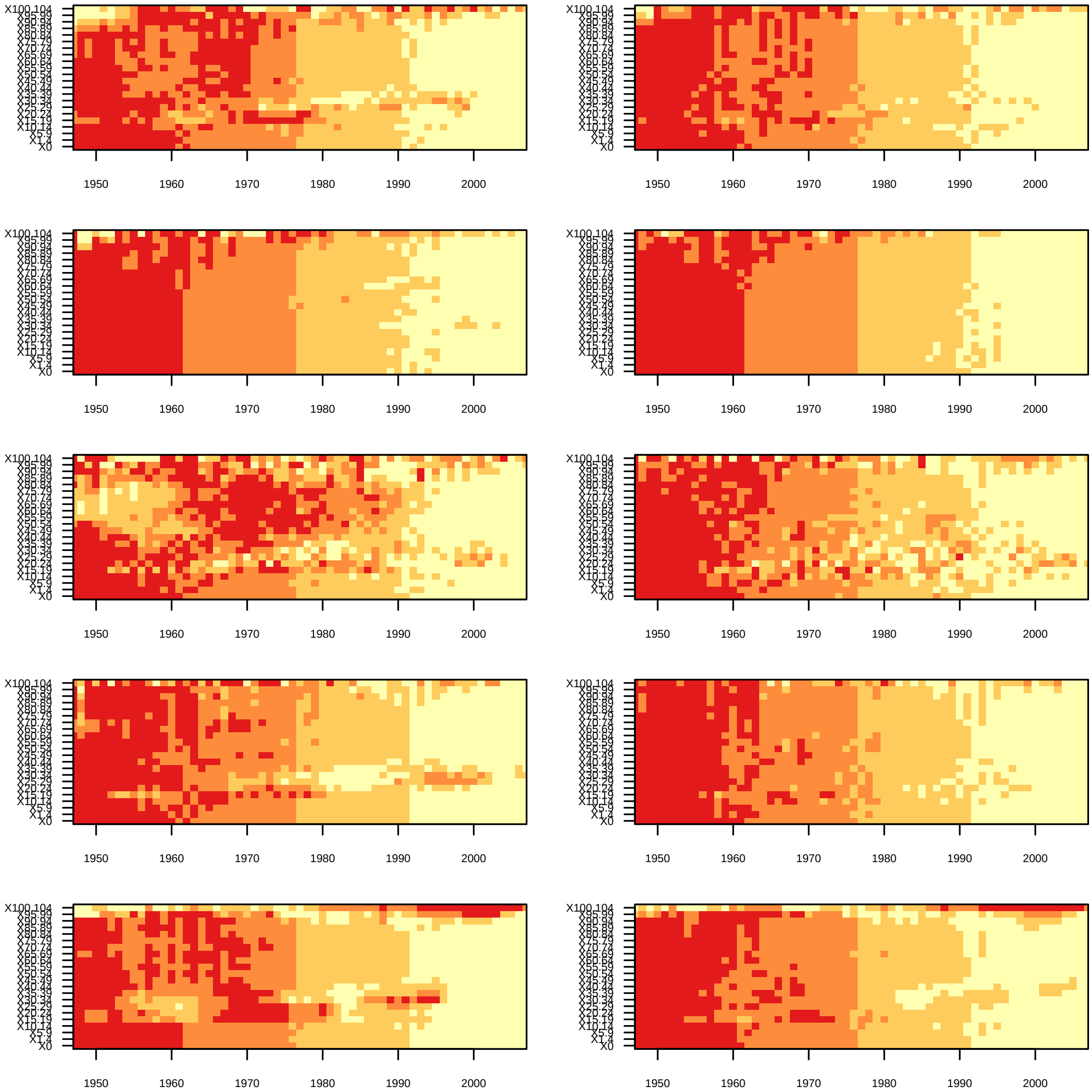


Figure 5.1: $\ln m_{x,t}$. mortality rates have been gradually declining with time at all ages. Countries from top to bottom are Australia, Japan, Norway, UK and USA. The left column is males while the right column is females.

Figure 5.1 shows that mortality rates have been gradually declining with time at all ages except the very elderly for most countries. The time series of mortality rates are shown in the four categories described above. Viewing the figure from left to right, the colour progression is red, orange, yellow then cream. For Japan, there are four distinctly coloured areas. This means that mortality rates for the Japanese declined in the same way for all ages. Other countries have some patches of different colours in some areas. For example, a red patch for age group x in a yellow area means that for age group x mortality rates were high (from the fourth quartile) in that period although for other age groups the mortality rates were relatively low (from the second quartile). The pattern for Norwegian Males between age 50 and

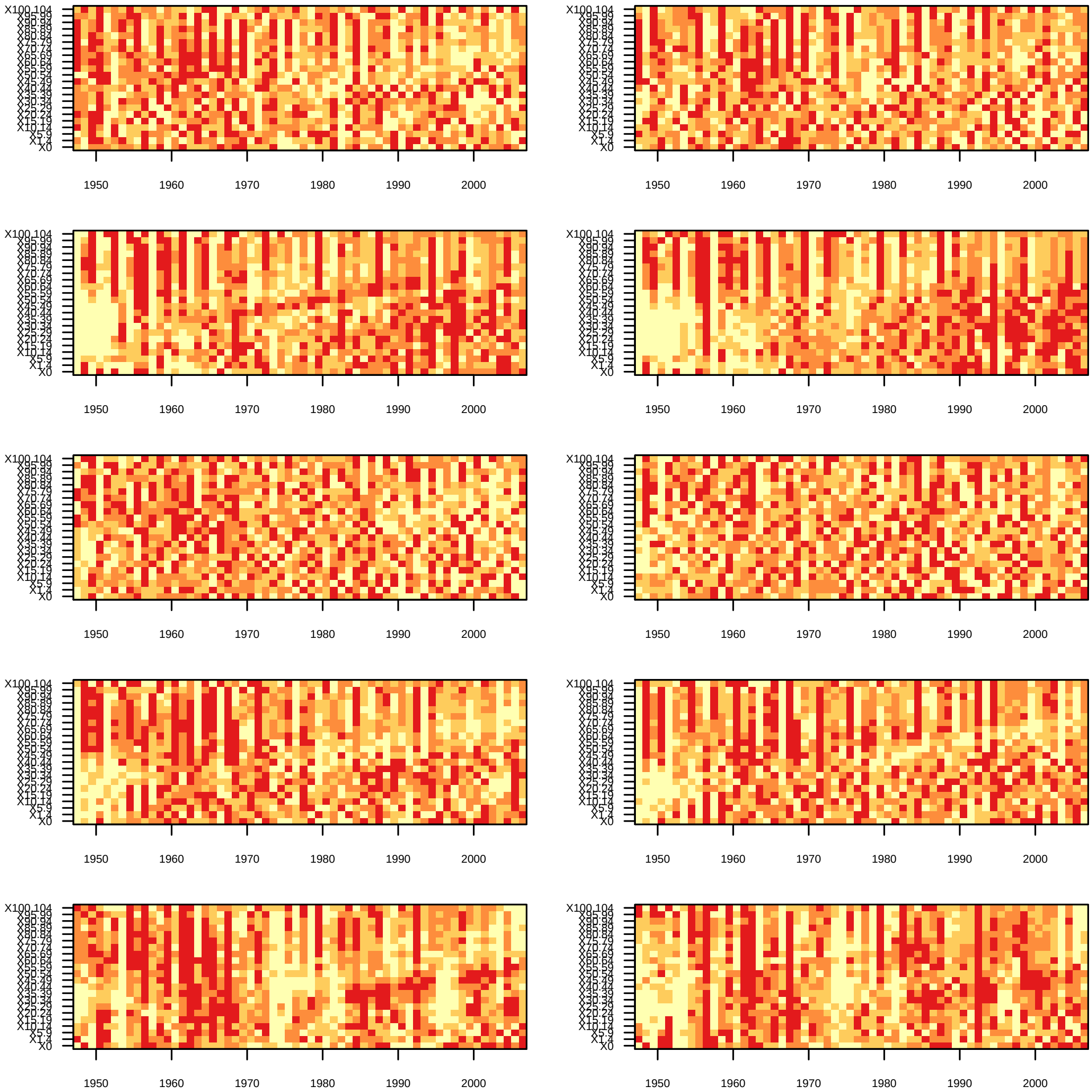


Figure 5.2: $\Delta_h \ln m_{x,t}$. Red areas mean that mortality either deteriorated or had the smallest improvement from one year to the next while cream areas mean that mortality improved from one year to the next. Countries from top to bottom are Australia, Japan, Norway, UK and USA. The left column is males while the right column is females.

80 is interesting because it shows that they experienced low mortality rates in the 1940s and 1950s (the predominantly yellow area). The highest mortality rates were in the 1960s and 1970s. However, the mortality rates after 1990 exhibit a similar behaviour to those of the other countries in the study. It is likely that the forces that drive Norwegian mortality rates will be different from those that drive mortality in the other countries. In general, mortality rates when examined from 1947 to 2007

declined for all the countries. Similar countries exhibit consistent mortality patterns like in Tuljapurkar et al. (2000).

Mortality improvements change with time and are not constant by age. The interpretation of figure 5.2 is similar to that of 5.1. However, while $\ln m_{x,t}$ is positive $\Delta_h \ln m_{x,t}$ can be negative. For example, if $\ln m_{x,t} - m_{x,t-1}$ is positive it means that mortality at t is higher than at $t-1$ and if it is negative it means that mortality at t is lower that at $t-1$. Other than that, the categories still represent the quartiles. Red areas mean that mortality deteriorated from one year to the next while cream areas mean that mortality improved from one year to the next. For most countries, there is an even mixture of colours. A vertical line of a single colour means that mortality rates for several age groups changed in the same way at that time. There are several instances where a red line is followed with a yellow or cream line. This is evidence of the observation that an increase in mortality improvement is usually followed by a decrease in mortality improvement (Willets, 2004; Andreev and Vaupel, 2006). Figure 5.3 shows the changes in mortality rates for a cohort. The bottom left corner represents those born in 1947. The diagonal going upwards to the right represents changes in mortality as those in a given age-group grow older. A yellow area shows that mortality was improving for cohorts at that time. A diagonal line of the same colour indicated the presence of a cohort effect. All the countries in the study except for Norway exhibit a cohort trend.

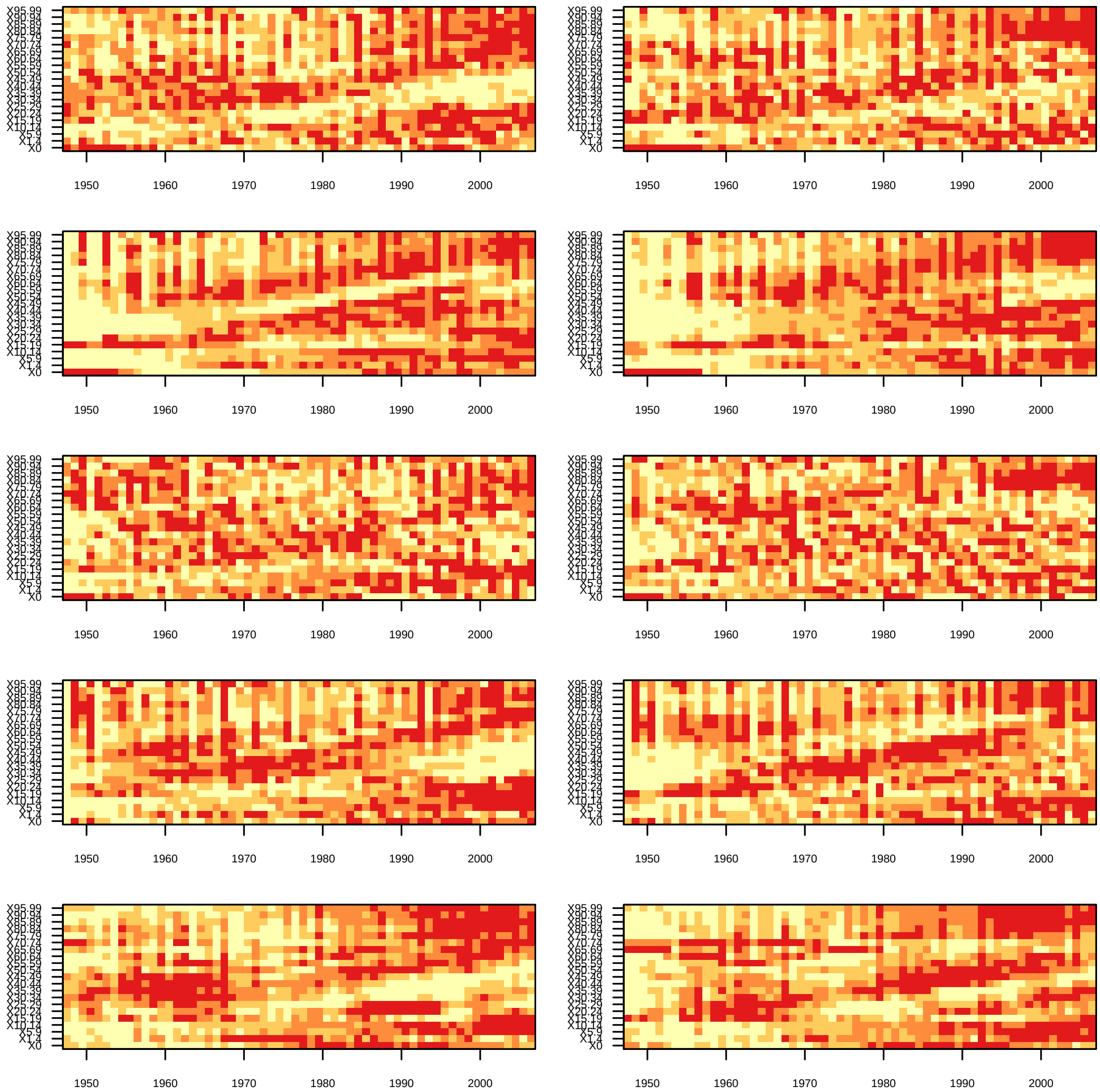


Figure 5.3: $\Delta_d \ln m_{x,t}$ Countries from top to bottom are Australia, Japan, Norway, UK and USA. The left column is males while the right column is females. A diagonal pattern in the same colour suggests a possible cohort effect.

5.3 Mortality trends, time trends and cohort trends viewed through the lenses of factor analysis and principal components analysis

Dimension reduction is often used in facial recognition software because it is useful for turning a blurry image into a clear one. In the same way, having obtained a visualization of mortality trends, time trends and cohort trends in figures 5.1, 5.2 and 5.3 respectively this section extracts and presents patterns that are not immediately recognizable in the data. The focus on the data is sharpened by looking through the lenses of dimension reduction techniques which provide tools that reduce the number of variables that describe the data by focusing on the variables that explain most of the variation. The lenses that are used to focus on this lower dimension space in this section are traditional statistical techniques. First, factor analysis is performed and then principal components analysis follows. Dimension reduction by factor analysis and principal components analysis is used to extract the age pattern of the data and thereby reduce the number of variables needed to explain the variation in the data (Booth and Tickle, 2008). Principal components analysis of mortality trends, for example, has been performed extensively in studies including Bell (1997); Bell and Monsell (1991); Bozik and Bell (1987); Shang et al. (2010); Hyndman and Booth (2008); Yang et al. (2010). To date time trends that explain the underlying pattern of mortality improvement over age and time have been analysed by Andreev and Vaupel (2005) on a basic level by smoothing the time trend surface using splines. That analysis does not give information about the factors that drive the variation in time trends. In addition, the analysis in Andreev and Vaupel (2005) makes inferences about countries of similar geographical location. This chapter of this thesis extends the research into time trends and cohort trends by analysing the factors that drive them. No other research has addressed this to date. Further, the analysis is done on a broad basis by considering five countries with a high Human Development Index (HDI) from different geographical locations. The results of this section will provide an understanding of the nature of trends in the data.

5.3.1 Factor Analysis:

It is important to reiterate the following two points from the literature view. Firstly, the definition of a factor as a variable that is not directly observable which describes the variations in several observed variables as given in the literature review; secondly, that factor analysis explains correlation between observable variables through (directly) unobservable factors. Mortality models that are based on deterministic trends for the levels of mortality rates make assumptions about the number of factors that should be included in the model. In this section, the number of factors to include in trend models for different countries is determined.

While performing a factor analysis, it is important to note that a k-factor model only makes sense if it has a positive number of degrees of freedom, $s \geq 0$, where s is determined from:

$$s = \frac{1}{2}(p-k)^2 + \frac{1}{2}(p+k) \tag{5.1}$$

This means that in the instance of a data set of mortality rates with ages $x = 0, \ldots, 99$ is used (100 variables), the maximum number of factors possible in a factor model is 86, while for age-groups, $x =$ "0", "$1-4$", "$5-9$", $\ldots$, "$100-104$" (22 variables) the maximum number of factors is 15.

Denoting A = Australia, J = Japan, N = Norway, UK = UK and US = USA, then for country $c = (A, J, N, UK, US)$ the factor model is:

$$x_c = \mu_c + \Lambda_c f_c + u_c \tag{5.2}$$

x_c are the trends for country c. The variance, Σ_c contains information about the common variation and the unique variation in the model (see equation (2.6)).

The factor loadings, $\Lambda = \{\lambda_{ij}\}$, and the common variation are related using $h_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2$, $i = 1, \ldots, k, \ldots, p$.

5.3.1.1 Factor Analysis of Mortality Trends and Volatility

The number of factors, k, that are significant[1] in a factor model on the mortality trends when the levels of the logarithms of the mortality rates for each of the five countries are analysed is given in table 5.1. There are a similar number of factors across the countries with a range between 6 and 10 factors except for Norway which has a significantly lower number of factors. Also, gender is important as the number of factors for males and females are not equal within a country (with the exception of the USA). It is interesting that for Australia, Japan, Norway and UK females require one less factor than males.

	Males	Females
Australia	7	6
Japan	8	7
Norway	4	3
UK	8	7
USA	8	8

Table 5.1: Number of factors necessary to explain the mortality trends of the countries in the study

Table 5.1 presented the number of factors and to include in a factor model of the mortality trends and most of the countries require a similar number of factors. The influence of the factors on different age groups can be studied through factor loadings, $\Lambda = \{\lambda_{ij}\}$. A visual illustration of the factor loadings is presented in

[1] p-value = 0.01. The hypothesis test is described in section 2.2.2

figure 5.4 showing only the first few factors for comparison across the countries in the study.

Figure 5.4 shows the factors for males and females (left and right columns respectively) after factor analysis. The first factor is similar for all genders and all the countries above. It starts off high for the very young and becomes low for the elderly. However, for Norway, UK and USA (males only) it also declines between ages 19 and 45. This mid-life decline is not pronounced for Australia and Japan. In the populations where the mid-life decline is very evident, the second factor exceeds the first factor in those age-groups.

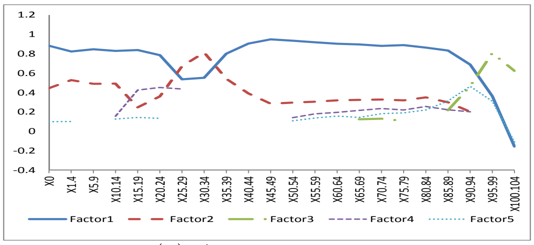
The second factor is less than the first factor for the youngest ages. With the exception of Japan, Australia (females only), UK (females only) and USA (females only) it has a hump between ages 19 and 40. For Japan, the first two factors move in opposite directions except for those older than 90 for males. For females, the second factor is similar for Australia, Japan and USA. The second factor for Japan is very similar for both genders. This is also the case for Norway and UK. Between ages 19 and 45 the influence of the first factor decreases and the influence of the second factor increases. The interplay between the first two factors is more pronounced when the curves of the factor loadings cross as they do for Norway and the UK in both males and females; and for Australia and USA males.

The similarity in the behaviour of the factors of Norway and the UK, two countries in the same geographical region, is noteworthy from figure 5.4. This suggests that geographical location contributes to the variability which is due to common factors. The population sizes are significantly different with UK having a population that is 13 times[1] as large as Norway's. However, the number of factors required in a factor model for Norway is much smaller than the number required for a factor model for the UK. Therefore, the conclusion is that although more unobservable factors drive the mortality trends for the UK than for Norway the first two trends are similar for both countries.

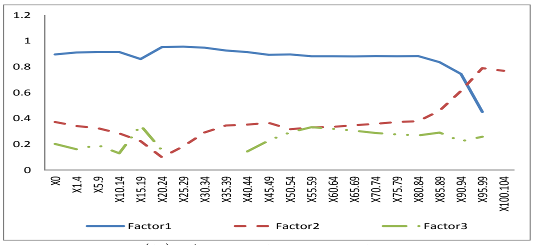
The first factor in both genders is similar in its behaviour in all the countries as explained above while the second factor is different for females. The first column of 5.4 represents males while the second represents females. The factors behave in a similar way with each country for both males and females but differ by country. Therefore, finding a suitable factor model that is generalizable for all these different countries is a challenge.

Next, factor analysis is used to identify the number of factors, k, driving the

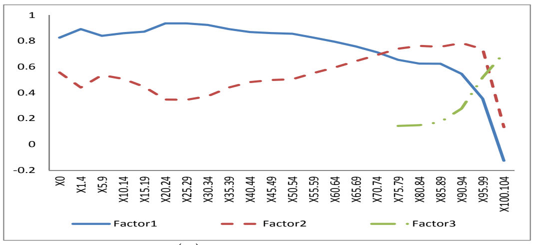
[1]From table 3.1 population size in thousands is 4,762 and 62,309 for Norway and UK respectively.



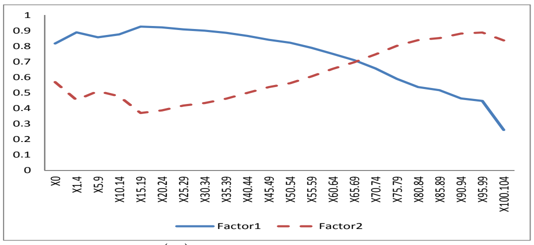
(a) Australia Males



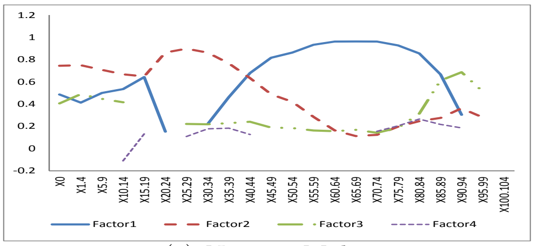
(b) Australia Females



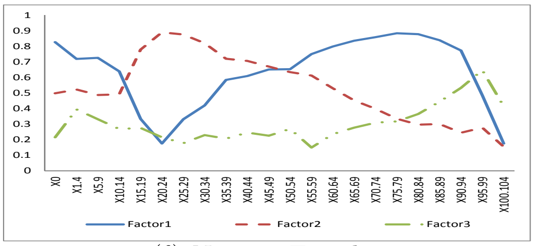
(c) Japan Males



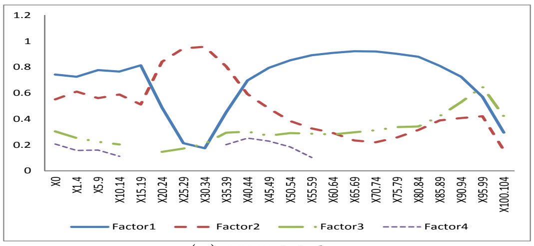
(d) Japan Females



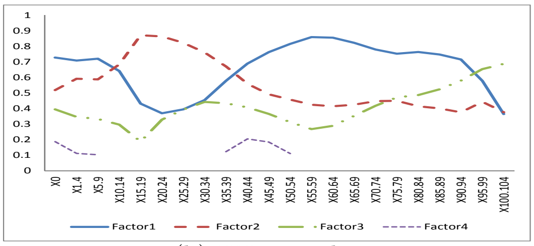
(e) Norway Males



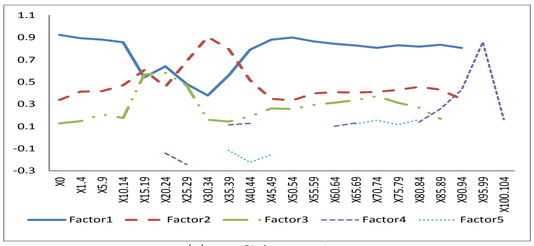
(f) Norway Females



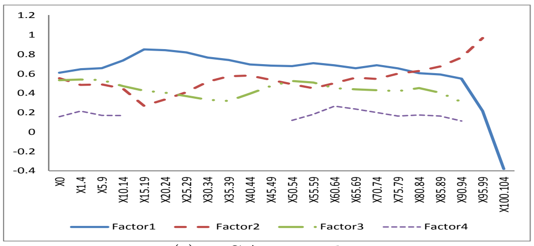
(g) UK Males



(h) UK Females



(i) USA Males



(j) USA Females

Figure 5.4: A visualisation of the loadings of the factors of mortality rates when analysed in levels. These plots show which variables are loaded most strongly on each factor. Some factors only affect specific age groups. Only factors that affect at least one third of the data points are plotted. The first factor is the blue (solid) line. The second factor is the red (broken) line. The third factor is the green (dash-dot-dot) line. The fourth factor is the purple (thinner smaller dotted line). The fifth factor is the blue (thin dotted) line.

volatility. This analysis is performed after removing the mortality trend through the drift term to obtain the time trends and cohort trend. From tables 5.2 and 5.3 the numbers of factors that are required to explain the variation in the horizontal differences (or time trends) and the diagonal differences (or cohort trends) respectively are given. The number of factors to explain age-period effects is fewer than those required to explain the age-period and cohort effects. For Norwegian data the cohort effect was not pronounced (see figure 5.3) and it is therefore not surprising that the number of factors required to explain the diagonal differences is not different from that required to explain the horizontal differences. The number of factors is much higher in the differences of the levels than in the levels themselves.

	Males	Females
Australia	5	4
Japan	4	4
Norway	4	2
UK	3	5
USA	5	4

Table 5.2: Number of factors necessary to explain the time trends $\Delta_h \ln m_{x,t}$

	Males	Females
Australia	7	4
Japan	10	6
Norway	3	2
UK	8	7
USA	9	10

Table 5.3: Number of factors necessary to explain the cohort trends $\Delta_d \ln m_{x,t}$

Certain factors would be expected to be common to certain age-groups, certain cohorts and the entire population. The communalities, $h_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2$, of each of the factor models measure how much of the variability is due to common factors. If the communality is close to one it means that the corresponding variable is explained the determined number of factors quite well.

The communalities are presented in tables 5.4, 5.5 and 5.6 for the factor analysis on levels, horizontal differences and diagonal differences of the logarithms of the mortality rates.

It is desirable to have communalities that are close to 1 for most ages. Variables with smaller communalities have more unique or variable specific variation. For the factor models on the levels of the logarithms of the mortality rates the communalities are close to one for all ages except 95+(see table 5.4). This means that factor models that include the number of factors computed in table 5.1 explain most of the variability which is due to common factors. For the levels of the mortality trends the amount of common variation explained in most of the attributes or age groups is close to 100% with a notable exception for the elderly age groups at the bottom of the table. The results are consistent for all countries except Norwegian Males aged 15-39 and Norwegian Females aged 5-19 and 30-34. Looking back at the raw data in

Age	Aus(M)	Aus(F)	Jpn(M)	Jpn(F)	Nor(M)	Nor(F)	UK(M)	UK(F)	USA(M)	USA(F)
0	0.99	0.99	1.00	1.00	0.96	0.98	1.00	1.00	1.00	1.00
1-4	0.99	0.98	1.00	1.00	0.97	0.94	0.99	0.99	1.00	1.00
5-9	0.99	0.98	1.00	1.00	0.96	**0.87**	1.00	0.99	1.00	1.00
10-14	0.98	0.97	1.00	1.00	0.92	**0.72**	0.99	0.98	1.00	1.00
15-19	0.98	0.97	0.99	1.00	**0.85**	**0.80**	0.97	0.98	0.99	0.98
20-24	0.97	0.98	1.00	1.00	**0.77**	0.90	0.99	1.00	0.99	1.00
25-29	0.94	0.96	1.00	1.00	**0.88**	0.91	0.98	0.99	1.00	1.00
30-34	0.96	0.99	1.00	1.00	**0.87**	**0.89**	0.99	0.99	0.98	1.00
35-39	0.98	0.99	1.00	1.00	**0.88**	0.90	0.98	1.00	1.00	1.00
40-44	0.99	0.99	1.00	1.00	0.93	0.93	1.00	1.00	1.00	1.00
45-49	1.00	0.99	1.00	1.00	0.94	0.92	1.00	1.00	1.00	1.00
50-54	1.00	0.99	0.99	1.00	0.95	0.91	1.00	1.00	1.00	1.00
55-59	1.00	1.00	1.00	1.00	0.97	0.96	1.00	1.00	1.00	1.00
60-64	1.00	1.00	1.00	1.00	0.97	0.97	1.00	1.00	1.00	1.00
65-69	1.00	1.00	1.00	1.00	0.98	0.98	1.00	1.00	1.00	1.00
70-74	1.00	1.00	1.00	1.00	0.98	0.99	1.00	1.00	1.00	1.00
75-79	1.00	1.00	1.00	1.00	0.97	0.99	1.00	1.00	1.00	1.00
80-84	0.99	1.00	1.00	1.00	0.97	0.99	1.00	1.00	1.00	1.00
85-89	0.99	1.00	1.00	1.00	0.93	0.99	1.00	1.00	1.00	1.00
90-94	0.98	0.99	1.00	1.00	**0.78**	0.95	0.99	1.00	1.00	1.00
95-99	0.90	0.93	0.96	0.99	**0.42**	**0.77**	0.95	0.97	**0.60**	0.94
100-104	**0.52**	**0.73**	**0.43**	**0.86**	**0.20**	**0.22**	**0.49**	**0.82**	0.97	**0.81**

Table 5.4: Communalities of the factors measuring the percentage of sample variance of each variable explained by the factors after FA of $\ln m_{x,t}$. Items in bold are variables where less than 90% of the sample variation is explained by k factors determined by FA. With the exception of Norway, low communalities are experienced in the age groups for those aged 95 and over.

figure 5.1, recall that Norwegian Males exhibited different patterns of relative levels of mortality with low mortality rates being experienced in 1940-1960 then increasing in from 1960-1990 and decreasing again after 1990. This pattern is different from that of the other countries in the study.

The communalities of the time trends are presented in table 5.5 and communalities with values over 70% are starred. The high communalities dominate the lower part of the table and particularly ages 65 to 99. This means that a high percentage of the common variation in the time trends of the elderly is explained using factor models with the number of factors indicate in table 5.2.

The communalities of the Japanese population's time trends stand out with over 70% of the sample variation in 16 out of 22 age groups for both males and females being explained by the factors. The percentage of common variation in time trends in the Norwegian population is smaller than the percentage explained in the other countries.

For the cohort trends of the five countries the highest communalities are experienced in the USA and the lowest communalities occur in Norway. The estimated factor models are suitable for use on the cohort trends of all countries in this analysis except Norway.

Age	Aus(M)	Aus(F)	Jpn(M)	Jpn(F)	Nor(M)	Nor(F)	UK(M)	UK(F)	USA(M)	USA(F)
0	0.37	0.36	0.73*	0.68	0.14	0.23	0.44	0.33	0.24	0.10
1-4	0.18	0.05	0.67	0.82*	0.30	0.13	0.30	0.61	0.39	0.35
5-9	0.17	0.20	0.60	0.67	0.45	0.09	0.25	0.76*	0.53	0.41
10-14	0.41	0.51	0.47	0.64	0.39	0.05	0.20	0.30	0.43	0.56
15-19	0.35	0.51	0.60	0.73*	0.24	0.15	0.50	0.64	0.83*	0.73*
20-24	0.20	0.57	0.81*	0.74*	0.11	0.06	0.65	0.51	0.76*	0.58
25-29	0.53	0.06	0.78*	0.76*	0.30	0.15	0.45	0.48	0.82*	0.79*
30-34	0.49	0.39	0.85*	0.78*	0.30	0.10	0.41	0.61	0.85*	0.67
35-39	0.17	0.41	0.82*	0.83*	0.33	0.06	0.38	0.42	0.85*	0.68
40-44	0.43	0.14	0.78*	0.74*	0.20	0.10	0.44	0.45	0.86*	0.59
45-49	0.38	0.26	0.64	0.61	0.53	0.03	0.47	0.49	0.65	0.62
50-54	0.34	0.32	0.71*	0.57	0.14	0.00	0.68	0.81*	0.61	0.61
55-59	0.58	0.65	0.74*	0.73*	0.22	0.26	0.63	0.77*	0.81*	0.58
60-64	0.69	0.57	0.83*	0.86*	0.34	0.19	0.88*	0.76*	0.77*	0.67
65-69	0.75*	0.48	0.88*	0.90**	0.46	0.16	0.90**	0.83*	0.80	0.72*
70-74	0.79*	0.66	0.96**	0.95**	0.55	0.57	0.93**	0.92**	0.81*	0.77*
75-79	0.75*	0.77*	0.98**	0.98**	0.58	0.50	0.94**	0.95**	0.85*	0.90**
80-84	0.85*	0.87*	0.91**	0.94**	0.77*	0.57	0.92**	0.96**	0.90**	0.92**
85-89	0.89*	0.93**	0.89*	0.88*	0.79*	0.69	0.94**	0.98**	0.90**	0.90**
90-94	0.78*	0.85*	0.89*	0.87*	0.58	0.57	0.88*	0.96**	0.84*	0.91**
95-99	0.58	0.72*	0.81*	0.77*	0.18	0.48	0.75*	0.83*	0.66	0.87*
100-104	0.63	0.53	0.42	0.23	0.16	0.22	0.23	0.22	0.40	0.48

Table 5.5: Communalities of the factors measuring the percentage of sample variance of each variable explained by the factors after FA of $\Delta_h \ln m_{x,t}$. * =70%-89% of the variation is explained; ** = 90% or more of the variation is explained.

Age	Aus(M)	Aus(F)	Jpn(M)	Jpn(F)	Nor(M)	Nor(F)	UK(M)	UK(F)	USA(M)	USA(F)
0	0.43	0.43	0.78	0.85	**0.15**	**0.08**	0.68	0.60	0.95	0.85
1-4	**0.27**	**0.28**	0.95	0.91	**0.28**	**0.28**	0.59	0.52	0.65	0.87
5-9	0.83	**0.34**	0.93	0.86	0.46	**0.03**	0.92	0.84	0.92	0.96
10-14	0.73	0.75	0.98	0.82	0.77	0.44	0.84	0.80	0.99	0.97
15-19	0.60	**0.22**	0.97	0.65	**0.27**	**0.03**	0.84	0.73	0.81	0.87
20-24	0.80	0.46	0.75	0.78	**0.00**	**0.06**	0.70	0.65	0.86	0.92
25-29	0.55	0.43	0.99	0.98	0.44	**0.08**	0.61	0.85	0.89	0.91
30-34	0.83	0.71	0.98	0.97	0.47	**0.01**	0.92	0.91	0.93	0.82
35-39	0.69	0.58	0.96	0.97	0.42	**0.13**	0.86	0.91	0.94	0.84
40-44	0.84	**0.39**	0.92	0.96	**0.17**	**0.12**	0.82	0.92	0.95	0.85
45-49	0.75	0.48	0.83	0.91	**0.08**	**0.14**	0.77	0.75	0.90	0.92
50-54	0.72	0.51	0.75	0.80	0.43	**0.00**	0.76	0.87	0.85	0.89
55-59	0.62	0.58	0.79	0.90	0.49	**0.33**	0.60	0.80	0.80	0.87
60-64	0.66	0.43	0.84	0.90	**0.24**	0.69	0.77	0.82	0.97	0.83
65-69	0.72	0.49	0.99	0.94	**0.39**	0.48	0.88	0.88	0.93	0.60
70-74	0.79	0.53	0.96	0.98	0.68	0.55	0.96	0.82	0.88	0.97
75-79	0.94	0.88	1.00	0.99	0.70	0.83	0.87	0.94	0.97	0.95
80-84	0.89	0.98	0.95	0.97	0.72	0.84	0.92	0.96	0.99	0.98
85-89	0.84	0.99	0.94	0.93	0.53	0.73	0.98	0.98	0.96	1.00
90-94	0.75	0.77	0.95	0.92	**0.21**	0.41	0.85	0.93	0.93	0.99
95-99	0.61	**0.37**	0.69	0.68	**0.05**	**0.05**	0.60	0.60	0.92	0.96

Table 5.6: Communalities of the factors measuring the percentage of sample variance of each variable explained by the factors after FA of $\Delta_d \ln m_{x,t}$. Items in bold are variables where 40% or less of the sample variation is explained by k factors determined by FA.

5.3.1.2 Factor Analysis on Australia Mortality Data

In this section, the results of factor analysis on Australian Mortality Data are presented. The aim of this analysis is to determine the number of factors that are necessary to describe the mortality curve. This analysis is performed on $m_{x,t}$ without any transformations. For males 7 factors[1] are sufficient to explain the variation in the mortality rates and similarly for females[2] 7 factors are also sufficient.

Age-Group	0	1-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
Males	1.00	0.99	0.98	0.98	0.95	0.96	0.93	0.98	0.96	0.99	0.99
Females	1.00	1.00	0.98	0.95	0.92	1.00	0.97	0.99	0.99	0.99	1.00

Age-Group	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100-104
Males	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	1.00	0.87	0.43
Females	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.92	0.69

Table 5.7: Communalities of the factors measuring the percentage of the sample variance of each variable explained by the factors after FA of $m_{x,t}$ for Australia.

A comparison of results of factor analysis (see figures 5.5 and 5.6) and the estimated parameters of the Heligman-Pollard Parameters (see figure 4.12) reveals an interesting result. The shape of the first factor (first row of figure 5.5) and the shape of the "A" parameter of the Heligman-Pollard model for Australian data is similar. Further to that, the loadings of the first factor show that it is significant for the very young. Similarly, the fourth factor and the "E" parameter and the third factor and the "D" parameter are comparable. This means it is reasonable to assume that the Heligman-Pollard model is a measure of some of the hidden factors that drive Australian mortality.

The communalities in table 5.7 are close to or equal to one for almost all age groups showing that the factor models explain most of the variation in Australian mortality.

[1]Males: The p-value for including 6 factors is 0.00071 with a χ^2-statistic of 168.35 on 114 degrees of freedom. The p-value for including 7 factors is 0.0308 with a χ^2-statistic of 125.79 on 98 degrees of freedom.

[2]Females: The p-value for including 6 factors is 0.000298 with a χ^2-statistic of 173.14 on 114 degrees of freedom. The p-value for including 7 factors is 0.0116 with a χ^2-statistic of 132.52 on 98 degrees of freedom.

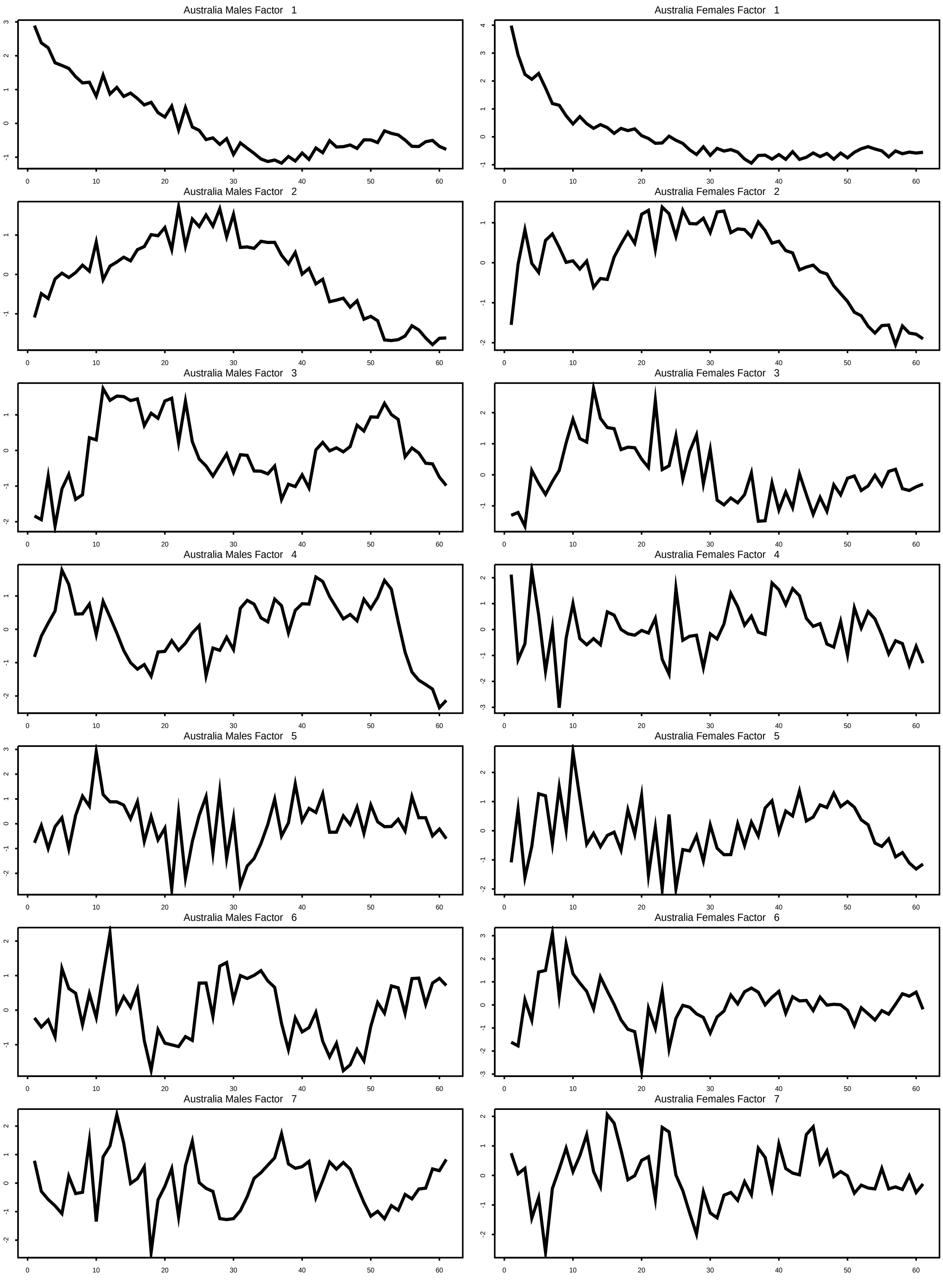


Figure 5.5: Seven Factors that drive Australian Mortality Rates, $m_{x,t}$.

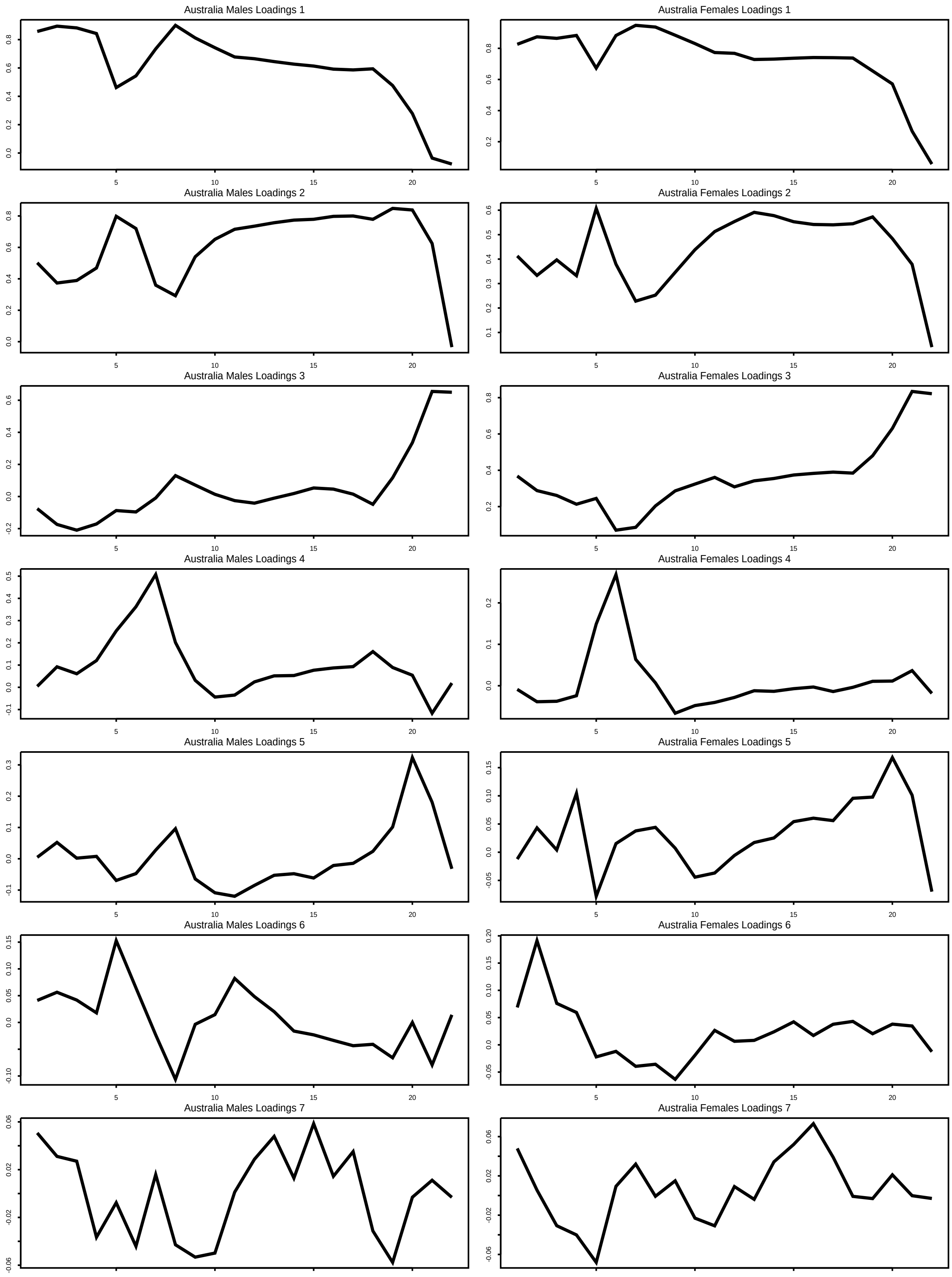


Figure 5.6: Loadings of the Seven Factors that drive Australian Mortality Rates, $m_{x,t}$.

5.3.2 Principal Components Analysis (PCA) on Mortality Trends and Volatility

In PCA, the variables themselves are of interest and a set of independent standardised linear combinations (SLC) of the variables that explain the variation in the data are sought. When only a small number of SLCs is retained, the dimensionality of the data is reduced and a small percentage of the variation in the data is explained. In this section, the number of principal components (SLCs) to retain in order to explain the variation in mortality trends and the volatility are investigated.

First, a PCA on the logarithm of mortality rates, $\ln m_{x,t}$, for the five countries is performed to determine the number of PCs necessary to drive mortality changes and the amount of variation explained by the selected number of PCs. The amount of variation in mortality trends that is explained by the first factor is important when choosing to use an existing mortality model such as, for example, the Lee-Carter model that assumes that one common factor drives mortality changes. The number of principal components to retain in this case is determined using the Cattell Scree test (Appendix A.1) which works well with strong factors where there is a single and clear break/elbow (for details see Cattell, 1966) and is one in all the cases. As the results are presented, it is important to emphasize that the Lee-Carter model is essentially a one-PC model. The first PC is k_t^{LC} in equation (2.12).

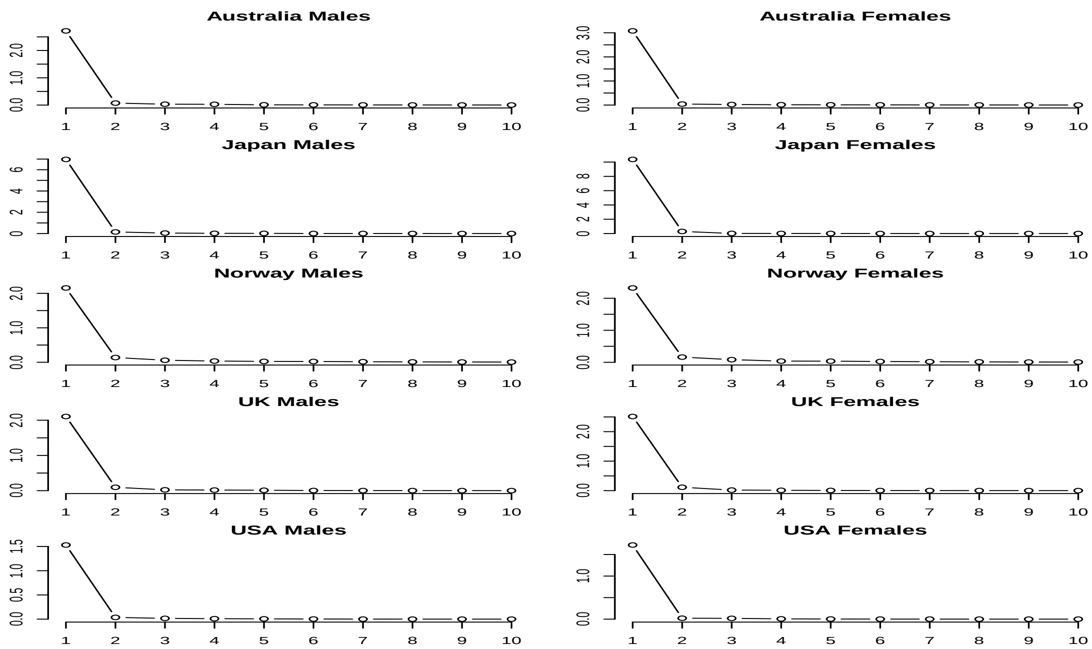


Figure 5.7: PCA Scree Plots. By Cattell's Scree Test One factor should be retained.

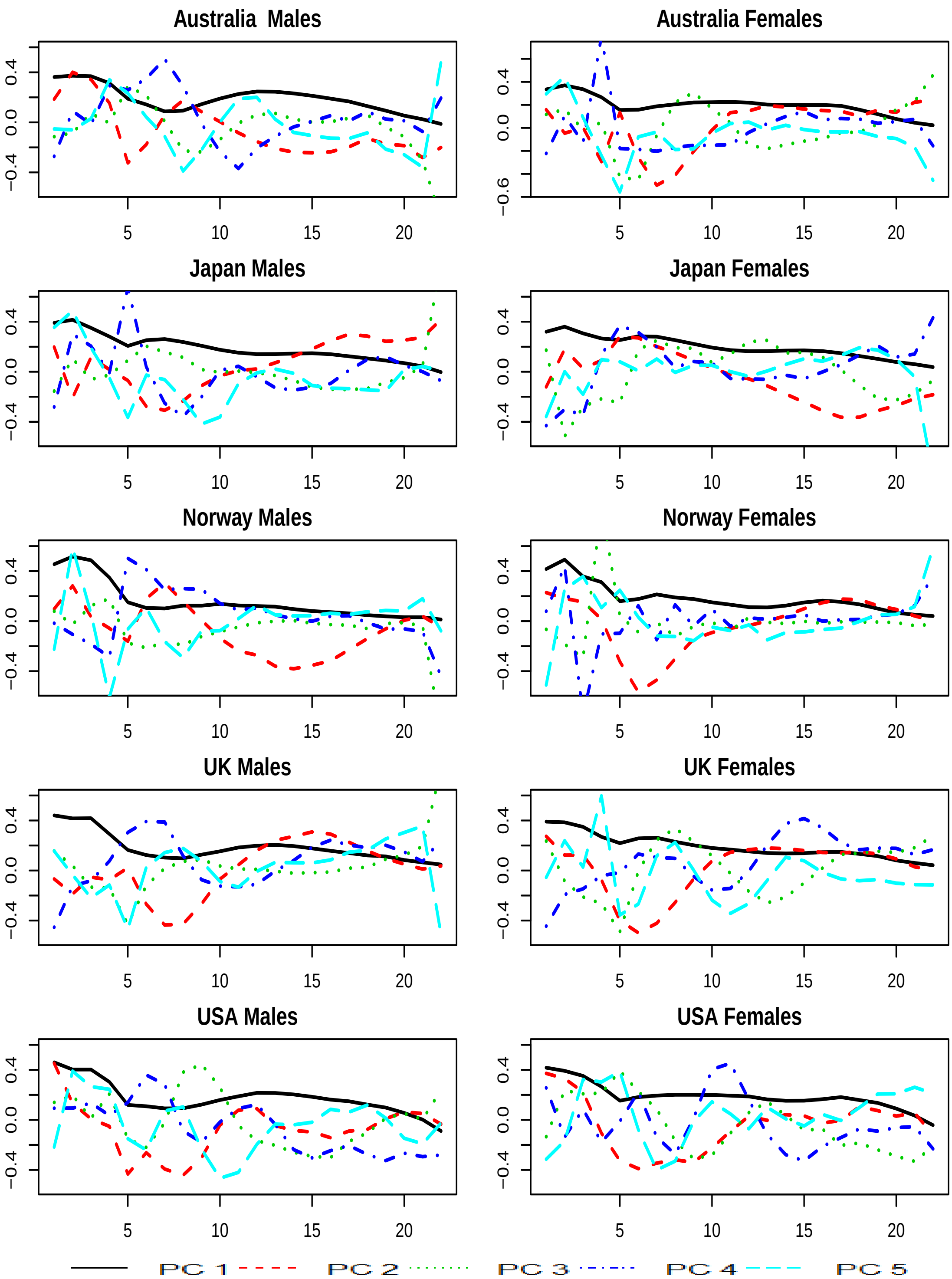


Figure 5.8: The loadings of the first five principal components of the mortality trends.

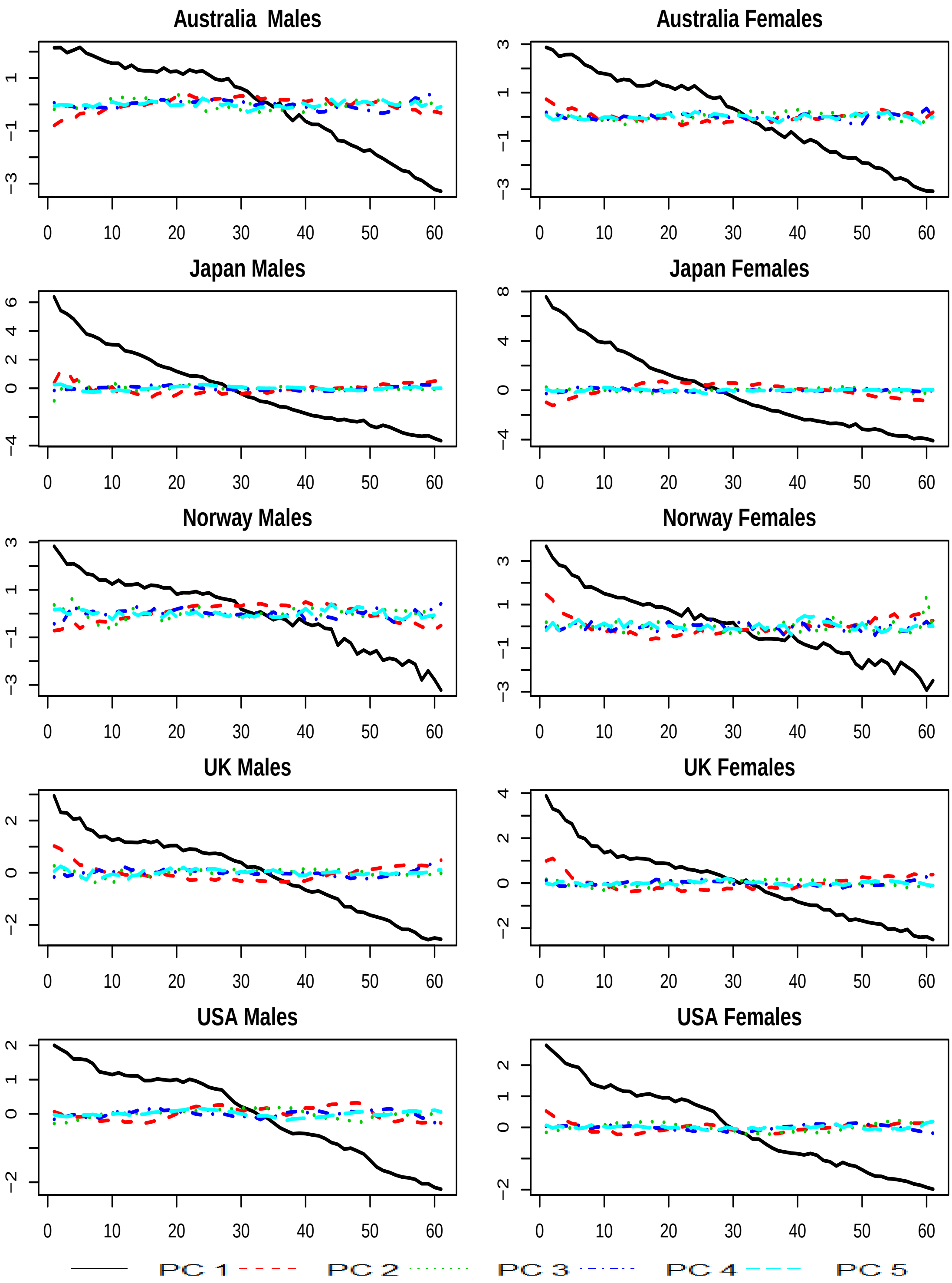


Figure 5.9: The first five principal components.

For all the countries, it is clear that the first principal component is the same as the Lee-Carter k_t^{LC} in equation (2.12) and the loadings are b_x^{LC}. A visual comparison to the estimated k_t^{LC} and b_x^{LC} when the Lee-Carter model is fitted (see figures 4.1-

4.10) confirms that the Lee-Carter model is based on one principal component.

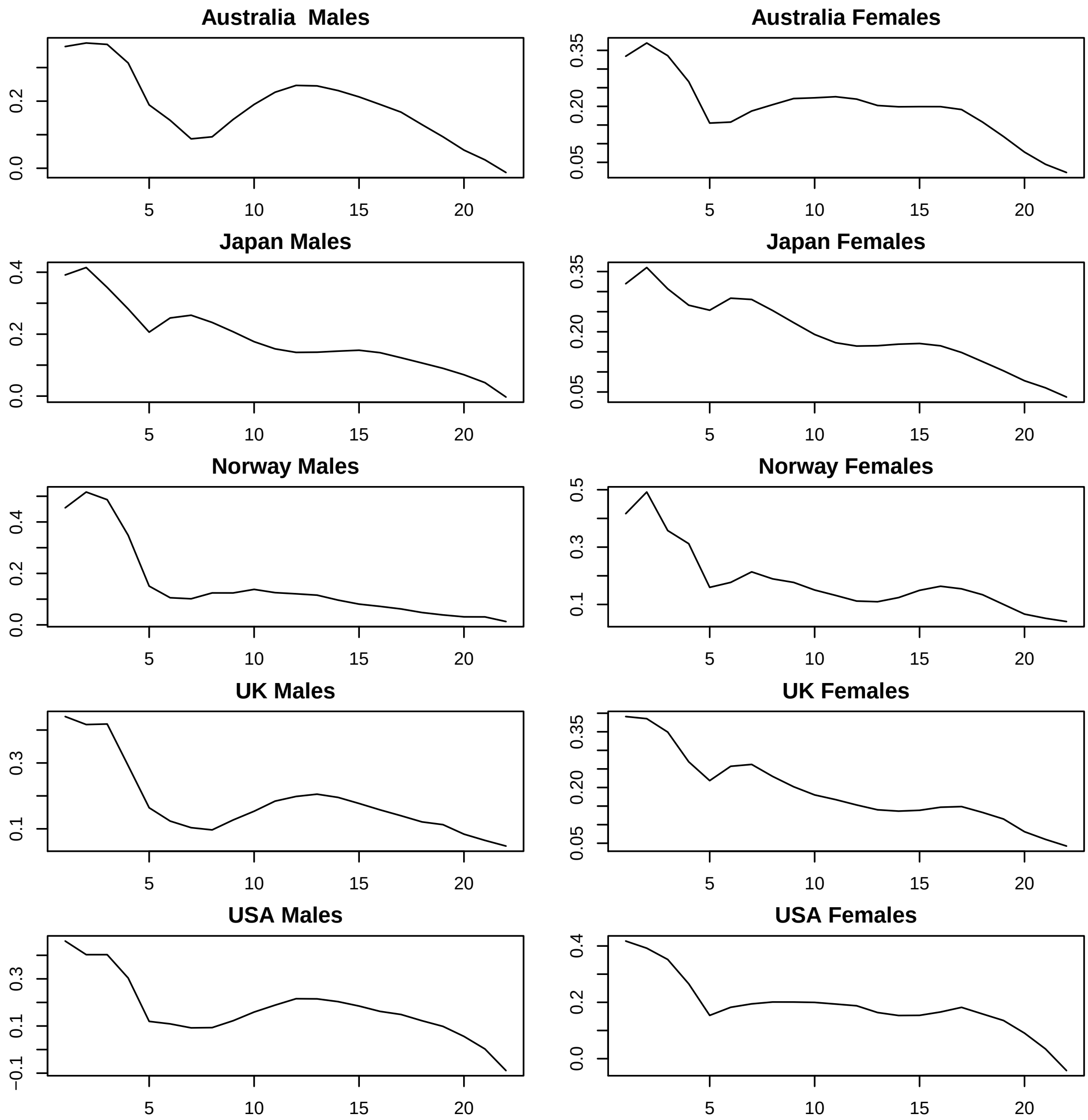


Figure 5.10: The Loadings of the 1st PC

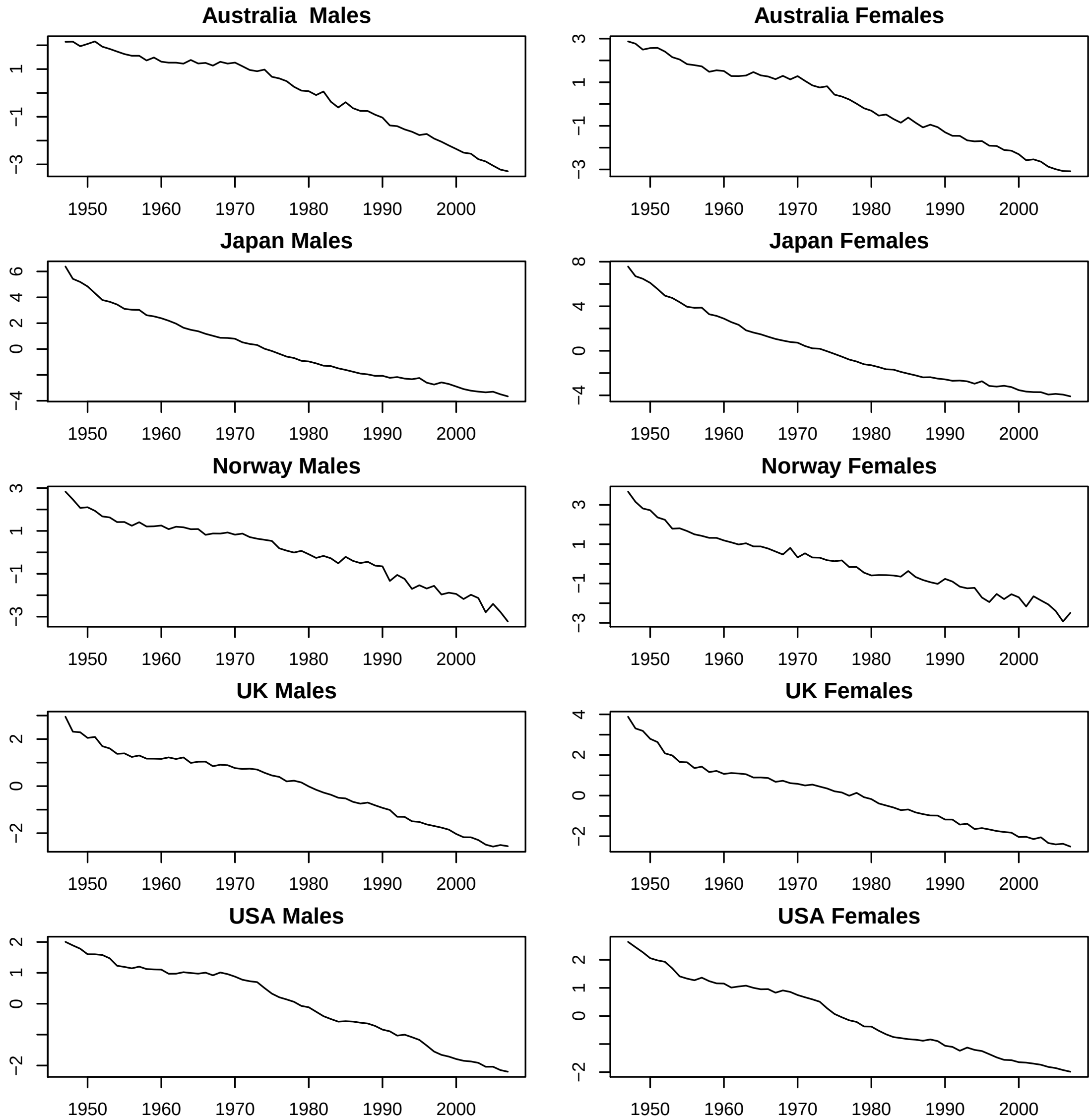


Figure 5.11: The 1st PC

The cumulative percentage of variation explained by the principal components is given in table 5.8. For all the countries in the study except for Norway, the first principal component explains at least 90% of the variation. As more principal components are added the cumulative percentage of variation increases. Adding extra principal components adds the number of parameters required by the model making it less parsimonious. In the case of the Lee-Carter model, including additional variation by adding one extra principal component leads to an additional $b_x k_t$ term and therefore the number of parameters increases by $N+T$, where N is the number of age groups and T is the number of observations.

When the percentage of variation explained by the first principal component is high the Lee-Carter model is likely to work well (Girosi and King, 2007). Based on this observation and the results in table 5.8 it is unlikely that the Lee-Carter model will work well when applied to Norwegian mortality trends.

		PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Australia	(M)	0.94	0.96	0.97	0.98	0.99	0.99	0.99	0.99
	(F)	0.96	0.97	0.98	0.98	0.99	0.99	0.99	1
Japan	(M)	0.96	0.99	0.99	0.99	1	1	1	1
	(F)	0.97	0.99	1	1	1	1	1	1
Norway	(M)	0.86	0.92	0.94	0.96	0.97	0.98	0.98	0.99
	(F)	0.85	0.9	0.93	0.95	0.96	0.97	0.98	0.98
UK	(M)	0.93	0.97	0.98	0.99	0.99	0.99	1	1
	(F)	0.94	0.98	0.99	0.99	0.99	1	1	1
USA	(M)	0.95	0.98	0.98	0.99	1	1	1	1
	(F)	0.97	0.98	0.99	0.99	1	1	1	1

Table 5.8: The cumulative percentage of variation in mortality trends, $\ln m_{x,t}$, explained by the first 8 Principal Components.

Another observation is that the first principal component is not approximately linear in cases such as Australia, Norway and UK. A quick check shows that the time series models that describe the evolution of the first principal component are different. The estimated autoregressive integrated moving average (ARIMA) models are summarised in table 5.9. The random walk with drift fits the USA population which is in line with the assumption in the Lee-Carter model.

	Males	Females
Australia	ARIMA(0,2,2)	ARIMA(0,1,1) with drift
Japan	ARIMA(2,2,2)	ARIMA(2,2,1)
Norway	ARIMA(0,1,3) with drift	ARIMA(0,1,1) with drift
UK	ARIMA(1,1,1) with drift	ARIMA(1,2,1)
USA	ARIMA(0,1,0) with drift	ARIMA(0,1,0) with drift

Table 5.9: The ARIMA model that best estimates the first principal component.

Having analysed the principal components that explained the variation in mortality trends, the next step is to analyse the volatility in the mortality trends as represented in the time trends, $\Delta_h \ln m_{x,t}$, and cohort trends, $\Delta_d \ln m_{x,t}$.

The number of principal components required to explain the changes in the time trends at different times is not easily determined from the scree plots in 5.12 but since the percentage of variation explained is given in table 5.10 the variance explained criterion is used. More than one principal component is required the variation in time trends unlike the mortality trends where one principal component was sufficient to explain over 90% of the variation. The number of principal components to explain the variation in the time trends varies by country.

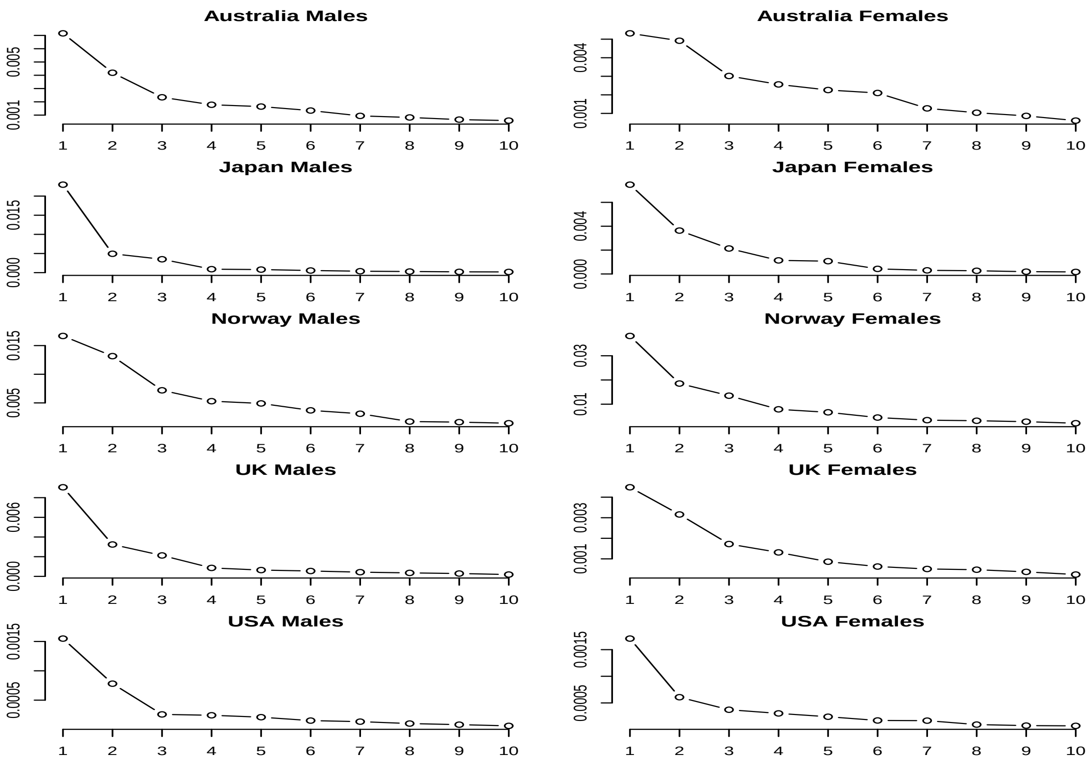


Figure 5.12: PCA on $\Delta_h \ln m_{x,t}$ Scree Plots. By Cattell's Scree Test More than One factor should be retained.

The loadings of the first five principal components are shown in figure 5.13.

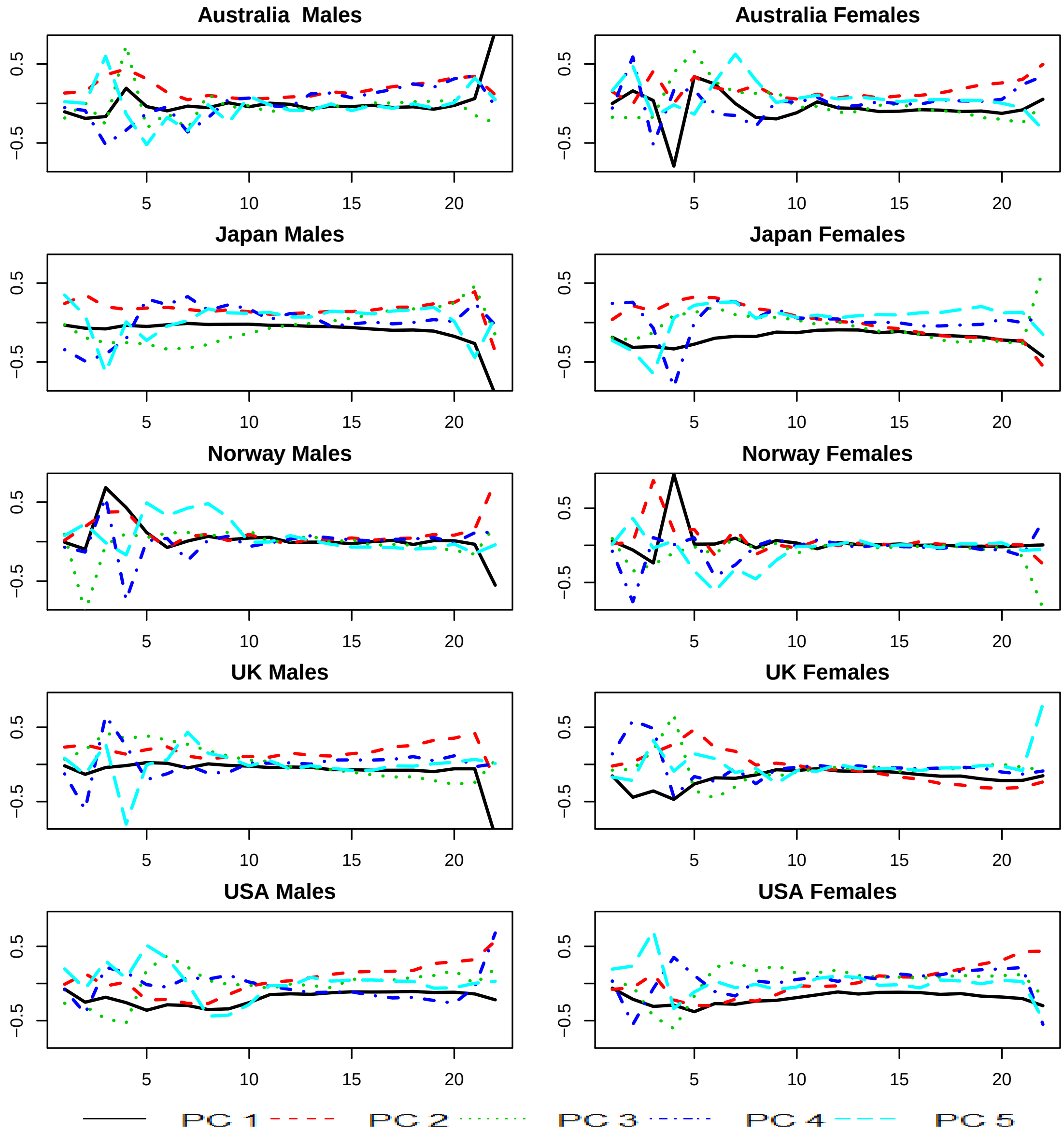


Figure 5.13: The loadings of the first five principal components of $\Delta_h \ln m_{x,t}$.

The interpretation of the principal components in this section is guided by the recommendations in Rao (1964) which gives an interesting discussion on how principal components are used and interpreted in applied research. It is necessary to isolate which principal components have an intrinsic mortality significance, which principal components represent the trend with time and which principal components measure random errors. The trend is only reflected in the first principal component (Rao, 1964). The first principal component of the period trends fluctuates around zero for all countries and both genders. Therefore, the is no trend in the changes in mortality rates with time.

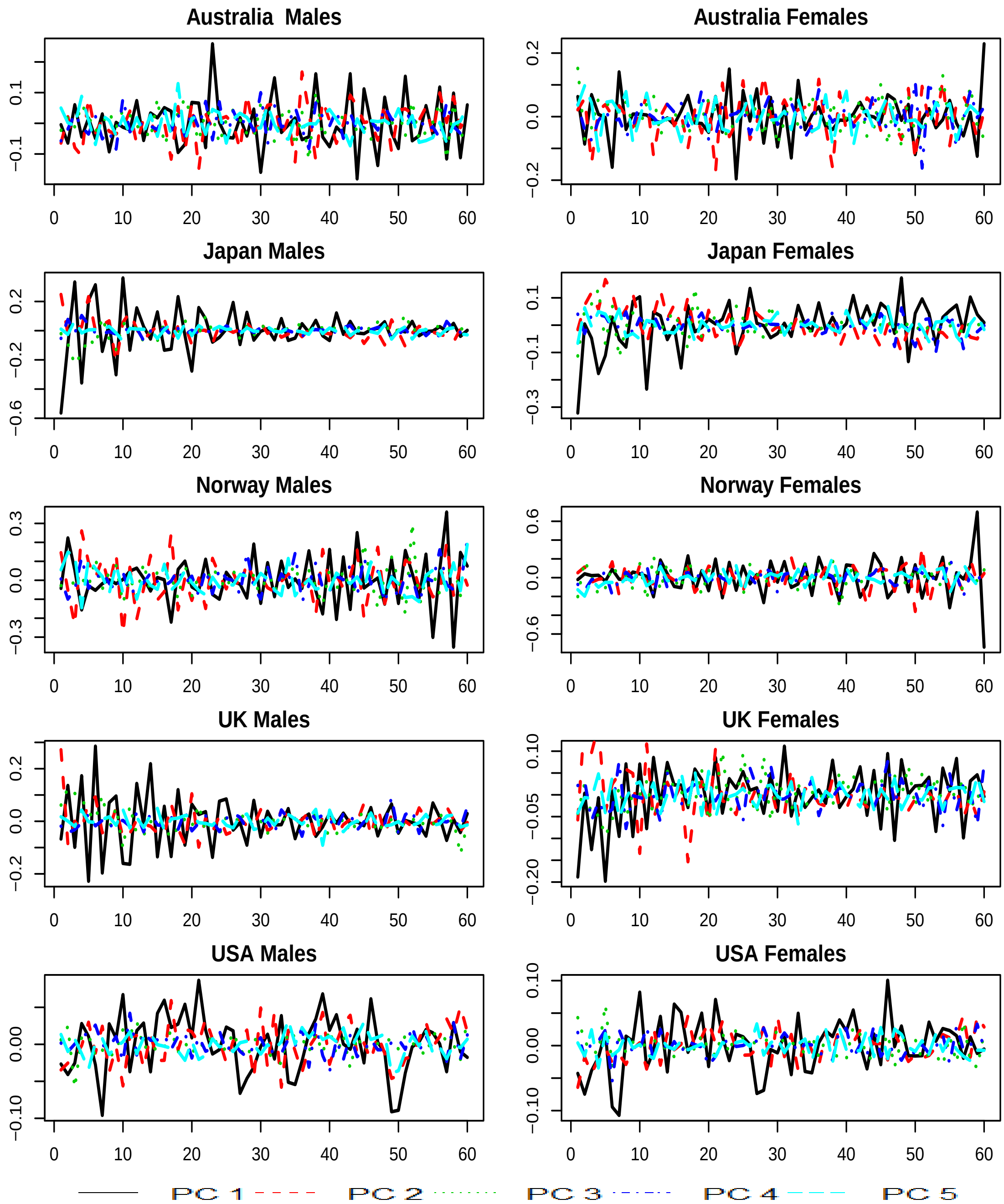


Figure 5.14: The first five principal components of $\Delta_h \ln m_{x,t}$.

The scree plots of the principal component analysis on the cohort trends are shown in figure 5.15. Since for countries such as Norway and the UK it is not easy to use Cattell's test directly, the Variance Explained criterion is used based on table 5.11. The number of principal components required to explain 90% of the variation is summarised in table 5.13.

		PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14
Australia	(M)	0.31	0.49	0.59	0.67	0.74	0.79	0.83	0.87	0.9	0.92	0.94	0.96	0.97	0.98
	(F)	0.21	0.4	0.51	0.6	0.69	0.77	0.82	0.86	0.9	0.92	0.94	0.96	0.97	0.98
Japan	(M)	0.63	0.75	0.87	0.9	0.93	0.95	0.96	0.97	0.98	0.98	0.98	0.99	0.99	0.99
	(F)	0.38	0.62	0.75	0.82	0.88	0.9	0.92	0.94	0.95	0.96	0.97	0.98	0.98	0.99
Norway	(M)	0.26	0.47	0.58	0.67	0.74	0.8	0.85	0.88	0.9	0.93	0.95	0.96	0.97	0.98
	(F)	0.36	0.54	0.66	0.74	0.8	0.84	0.87	0.9	0.93	0.95	0.96	0.97	0.98	0.98
UK	(M)	0.53	0.67	0.78	0.83	0.86	0.89	0.92	0.94	0.95	0.96	0.97	0.98	0.98	0.99
	(F)	0.29	0.52	0.65	0.74	0.8	0.85	0.89	0.92	0.94	0.95	0.97	0.98	0.98	0.99
USA	(M)	0.41	0.61	0.68	0.75	0.8	0.84	0.88	0.9	0.92	0.94	0.95	0.97	0.97	0.98
	(F)	0.43	0.57	0.66	0.74	0.8	0.84	0.88	0.9	0.92	0.94	0.95	0.96	0.97	0.98

Table 5.10: The percentage of variation in $\Delta_h \ln m_{x,t}$ explained by the Principal Components.

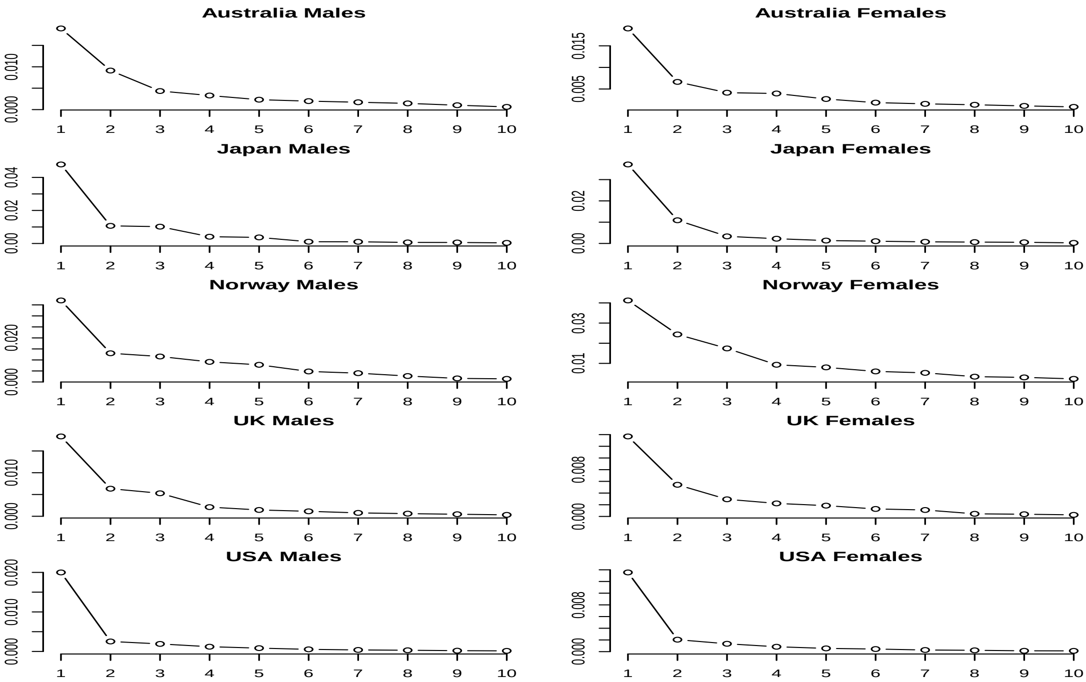


Figure 5.15: PCA on $\Delta_d \ln m_{x,t}$ Scree Plots. By Cattell's Scree Test More than One factor should be retained.

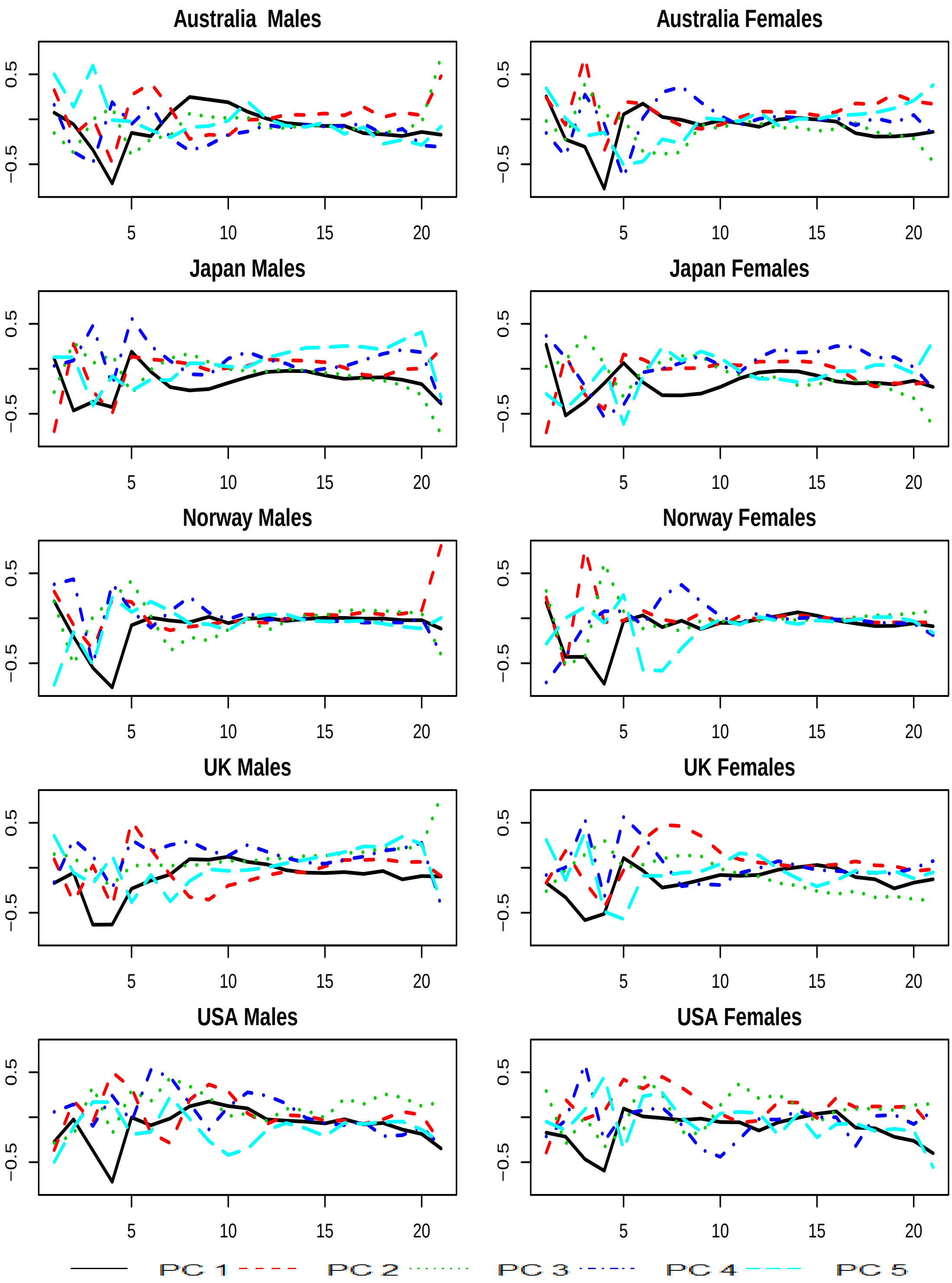


Figure 5.16: The loadings of first five principal components of $\Delta_d \ln m_{x,t}$.

The first principal component is an upward trend (see figure 5.17) and this first component has a decreasing effect as age increases from 0 - 25 (see figure 5.16) and then it rises then peaks at approximately age 35 for all countries in the study except Japan. At the oldest ages first principal component has a similar general effect in all five countries being initially fairly constant and then decreasing as age increases. Towards the left hand side of each of the graphs in figure 5.16 common feature for the different countries is that for the middle ages (50-75) the effects are fairly constant but at different intensities. At the higher ages (above age 75) there is a lot more variability in the way each principal component affects a specific age's mortality change. The second principal component for Japan and the third principal component for USA at the oldest ages both increase as age increases. This may be the same principal component but its order differs for each country. For some countries, such as Japan, the PC's are more variable, while for the other countries such as Norway (Males) they appear relatively constant between age 75 and age 85.

From figure 5.17 the first principal component further implies that for the countries in the study that mortality is improving at a decreasing rate. The first principal component increases and then plateaus. The plateau begins at different times for different countries. For Japan, for example, it begins to plateau at t=30 (1977) while for Australia (males) it begins to plateau just before t=50 (1997).

		PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14
Australia	(M)	0.39	0.59	0.68	0.75	0.8	0.84	0.88	0.91	0.94	0.95	0.96	0.97	0.98	0.99
	(F)	0.41	0.56	0.65	0.74	0.8	0.84	0.88	0.91	0.93	0.95	0.96	0.97	0.98	0.98
Japan	(M)	0.56	0.7	0.84	0.89	0.93	0.95	0.96	0.97	0.98	0.98	0.99	0.99	0.99	0.99
	(F)	0.6	0.81	0.87	0.9	0.92	0.94	0.96	0.97	0.98	0.99	0.99	0.99	0.99	1
Norway	(M)	0.37	0.51	0.63	0.72	0.8	0.85	0.89	0.92	0.94	0.95	0.97	0.98	0.98	0.99
	(F)	0.33	0.53	0.66	0.73	0.8	0.84	0.88	0.91	0.93	0.95	0.96	0.97	0.98	0.99
UK	(M)	0.48	0.66	0.79	0.83	0.87	0.9	0.93	0.94	0.96	0.97	0.97	0.98	0.99	0.99
	(F)	0.46	0.61	0.7	0.78	0.84	0.89	0.93	0.94	0.96	0.97	0.97	0.98	0.99	0.99
USA	(M)	0.71	0.79	0.86	0.91	0.93	0.95	0.97	0.97	0.98	0.99	0.99	0.99	0.99	1
	(F)	0.68	0.78	0.85	0.89	0.92	0.94	0.95	0.96	0.97	0.98	0.98	0.99	0.99	0.99

Table 5.11: The percentage of variation in $\Delta_d \ln m_{x,t}$ explained by the Principal Components.

	Males	Females
Australia	9	9
Japan	4	6
Norway	9	8
UK	7	8
USA	9	9

Table 5.12: Number of PCs necessary to explain 90% of the cumulative variation in the time trends $\Delta_h \ln m_{x,t}$

	Males	Females
Australia	8	8
Japan	5	4
Norway	8	8
UK	6	7
USA	4	5

Table 5.13: Number of PCs necessary to explain 90% of the cumulative variation in the time trends $\Delta_d \ln m_{x,t}$

It is expected that the variation in mortality rates through time and age require more principal components to explain the changes than the variation in mortality

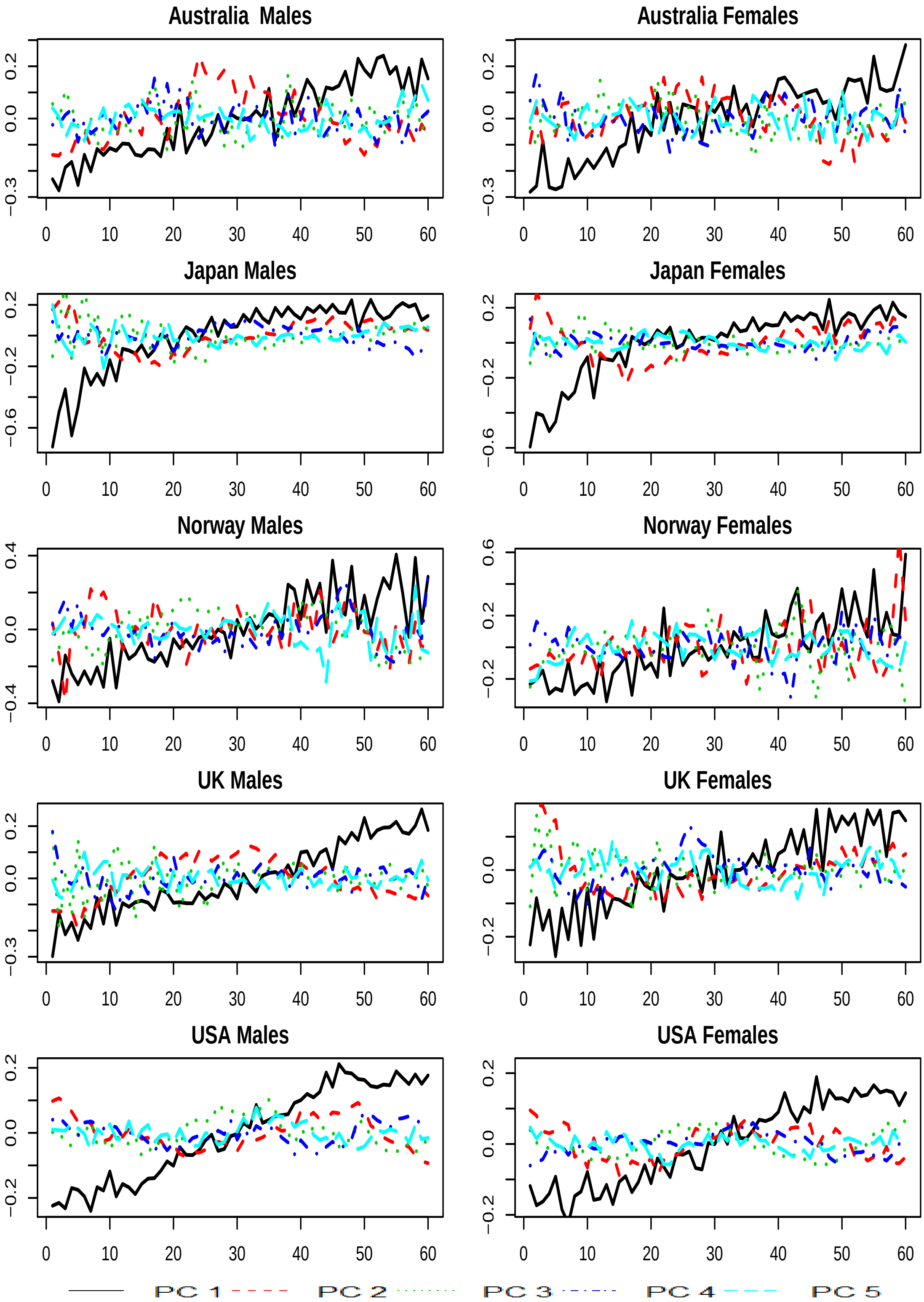


Figure 5.17: The first five principal components of $\Delta_d \ln m_{x,t}$.

rates through time alone. However, the principal components do not capture this as well as the factors. Therefore, from this analysis the conclusion is that mortality is driven by several unobservable variables.

5.3.3 Summary on PCA and FA

A basic VAR time series model of age-specific death rates is written as a system of equations for ages from 0 to N including a drift $\mu(\cdot)$.

$$\begin{bmatrix} m(0,t) \\ m(1,t) \\ \vdots \\ m(N,t) \end{bmatrix} = \begin{bmatrix} \mu_0 \\ \mu_1 \\ \vdots \\ \mu_N \end{bmatrix} + \tilde{\Phi} \begin{bmatrix} m(0,t-1) \\ m(1,t-1) \\ \vdots \\ m(N,t-1) \end{bmatrix} + \begin{bmatrix} \epsilon(0,t) \\ \epsilon(1,t) \\ \vdots \\ \epsilon(N,t) \end{bmatrix} \qquad \tilde{\Phi} = \begin{bmatrix} \Pi_0 & 0 & 0 & \dots \\ 0 & \Pi_1 & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Pi_N \end{bmatrix} \tag{5.3}$$

where Π_i is the coefficient of the lagged mortality rate time series for age i.

FA and PCA were carried out for the time trends, based on errors for horizontal differences of the mortality rates, from the simple VAR model for the age based mortality data. This determined the number of factors and principal components that affect mortality over time for the various ages. FA and PCA were performed for each country on the time differences of the levels of the mortality data assuming a first order stochastic trend. FA and PCA were also carried out for the cohort trends, based on errors for diagonal differences of the mortality rates.

The conclusion is that models based on deterministic trends for the level of mortality rates should include a larger number of factors driving mortality changes. There are also a similar number of factors across the countries suggesting the possibility of common factors across countries. This similar number of factors holds for both time trends (horizontal differences) and cohort trends (diagonal differences). However, a greater number of factors are required to explain the variation in the cohort trends than the time trends. This is not surprising since the cohort trends include the time trends along with an age effect. The number of factors is much higher in the differences of the levels than in the levels themselves.

5.4 Econometric Modelling:

In this section it is of interest to analyse the correlation and dependence through time between the variables in the data sets of mortality rates. VAR models dynamically model the interactions between the variables. It is necessary to test the time series of the mortality trends for unit roots. The presence of unit roots signals the potential for common trends to exist if the variables with unit roots are cointegrated. For the within (multicountry) analysis, mortality rates that are cointegrated mean that econometric models can be used to estimate the long run relationships between the age-groups (countries) and how changes in age-group (country) mortality rates respond to departures from the long run equilibrium between the mortality rates.

5.4.1 Unit Root tests:

For the multi-country analysis, unit root and stationarity tests were performed on the standardised mortality rate time series for the period from 1963-2007. The period from 1963-2004 is considered because there is a very drastic improvement in mortality for Japan between 1947 and 1962 which may distort the effects of common trends.

		ADF		PP Test	
		Males	Females	Males	Females
Australia	Test Stat.	-1.854	-0.5727	-3.249	-1.51
	P-value	0.6608	0.9755	0.08846	0.811
Japan	Test Stat.	-1.706	-3.441	-2.008	-2.871
	P-value	0.7314	0.05957	0.5811	0.1813
Norway	Test Stat.	-1.547	-3.442	-1.577	-3.44
	P-value	0.7974	0.05879	0.7861	0.05904
UK	Test Stat.	-2.338	-1.624	-3.804	-3.837
	P-value	0.4052	0.7669	0.02565	0.02367
US	Test Stat.	-0.7122	-0.9385	-0.9774	-0.9846
	P-value	0.9658	0.9421	0.9368	0.9358

Table 5.14: Unit Root tests on Standardised Mortality Rates. Critical values for the ADF and PP tests are found in Banerjee (1993).

The test statistic is greater than the critical value for the country mortality series but is smaller than the critical value for the differenced mortality series. The country unit root tests for both males and females are compared to the results of KPSS stationarity tests. These confirm the conclusion that differences of the mortality rates at the country level are stationary and not the levels. These tests confirm that the series are integrated of order one for both the males and the females.

The analysis demonstrates that mortality rates for fixed ages across time, based on the historical population data for the countries in this study, are mostly difference stationary with stochastic trends. Time trends in mortality rates for fixed ages across time should be modelled as difference stationary where they have unit roots. Shocks across time are permanent for these ages and the volatility of mortality rates increases across time. Thus uncertainty about future mortality will increase with increased forecast horizons in contrast to trend stationary models that assume a long run stationary level of volatility.

Treating mortality at each age as a non-stationary variable, cointegration allows the determination of the ages that have experienced similar persistence of random shocks. Mortality rates have stochastic trends through time and the major shocks to mortality rates accumulate in the series. These non-stationary time series require a transformation such as differencing to obtain stationarity. The underlying long-run information of a non-stationary time series is removed if the data is detrended or differenced.

Cointegration analysis models the long-run relationship and retains statistical information. The following section employs cointegration analysis to study if there exist long-run relationships between the standardised mortality rates for different countries. Cointegration tests within the country require large amount of data or analysis of a subset of the age-groups.

5.4.2 VAR Models and Cointegration Tests

Population mortality rates across countries are expected to contain common stochastic trends and to be cointegrated based on the previous analysis. The standardised mortality rates for males and females differ and are analysed separately based on the standardized age specific mortality rates for Australia, Japan, Norway, UK and USA from 1963-2007. This analysis is carried out for the standardized country mortality rates to estimate long run equilibrium common stochastic trends by estimating a VAR model, and if required, a VECM model that incorporates those long run stochastic trends. This allows data to be combined across countries leading to more efficient estimates of future mortality rates and providing information about the relationship between the different countries mortality improvement and longevity risk.

Denoting A = Australia, J = Japan, N = Norway, UK = UK and US = USA, then a VAR(1) model for this system would be specified as:

$$\mathbf{m}_t = A_0 + A_1\mathbf{m}_{t-1} + u_t$$

where

$$\mathbf{m}_t = \begin{pmatrix} m_{A,t} \\ m_{J,t} \\ m_{N,t} \\ m_{UK,t} \\ m_{US,t} \end{pmatrix}; \qquad \Delta\mathbf{m}_t = A_0 + \Pi\mathbf{m}_{t-1} + u_t$$

$$\Pi = \alpha\beta' = -(I - A_1)$$

$$\beta = (\beta_A, \beta_J, \beta_N, \beta_{UK}, \beta_{US}) \tag{5.4}$$

Π measures the long-run parameters. Cointegration implies that Π is of reduced rank with $rk(\Pi) = r$ where r is the number of cointegration relationships that exist between the variables or the cointegration rank. The α parameters are the loading matrix while β contains the coefficients of the long run relationships such that $\beta'\mathbf{m}_{t-1}$ gives the cointegrating relations.

The unit root tests indicated there may be cointegration of the mortality rates for the five countries. For the time period from 1963 to 2007 a VAR model was estimated including a constant and a trend as the deterministic regressors. Diagnostic tests are performed to assess the assumptions of the VAR model. There should be no serial correlation and no heteroscedasticity in the residuals and the residuals should be normally distributed.

The VAR model was estimated using the levels of the mortality rates and also using the logarithms of the levels of the mortality rates. The optimal lag length, p, for an unrestricted VAR model to analyse the cointegration is determined using a range of information criteria including the AIC and the final prediction error. It is also important to have a parsimonious model with as few lags as possible.

For the multi-country standardised mortality rates, the BIC finds that a VAR(1) is suitable to model the time series. Further, diagnostic tests found that a VAR(1) model is appropriate for both males and females based on the logarithms of the standardized rates (see table 5.15).

		Males	Females
		VAR(1)	
Log-Likelihood	Level	1719.575	1811.72
	Log-level	622.489	603.47
Portmanteau Test	Level	1.88E-05	1.46E-02
	Log-level	0.06426	0.3374
Jarque-Bera Test	Level	0.001825	0.5918
	Log-level	0.08784	0.5036
		VAR(2)	
Log-Likelihood	Level	1722.358	1810.904
	Log-level	633.973	611.52
Portmanteau Test	Level	7.75E-06	0.001399
	Log-level	0.003689	0.02007
Jarque-Bera Test	Level	0.05117	0.563
	Log-level	0.00348	0.807

Table 5.15: Diagnostic results for VAR(p)

For this model the normality assumption for the residuals is not rejected and the model captures all significant serial correlations. The Log-likelihood for the VAR(2) models are higher than those of the VAR(1) but the difference is negligible especially considering that an additional 25 ($n \times n$, when $n = 5$) parameters will be required for the additional lag. However, based on the test statistics, the residuals of the VAR(1) in levels were serially correlated which makes inference for parameters and goodness of fit unreliable based on the assumption of no serial correlation. The residuals of the VAR(1) of the logarithms are not serially correlated.

The suitable VAR(p) for males and females are given in 5.5 and 5.6 respectively.

$$
\begin{aligned}
\ln m_{At} &= 0.0406 - 0.0071t + 0.2118 \ln m_{At-1} - 0.2124 \ln m_{Jt-1} \\
&\quad + 0.0573 \ln m_{Nt-1} + 0.2935 \ln m_{UKt-1} + 0.6563 \ln m_{USt-1} \\
\ln m_{Jt} &= -2.8727 - 0.0149t - 0.0646 \ln m_{At-1} + 0.5365 \ln m_{Jt-1} \\
&\quad - 0.0934 \ln m_{Nt-1} - 0.3354 \ln m_{UKt-1} + 0.3491 \ln m_{USt-1} \\
\ln m_{Nt} &= -3.5011 - 0.0153t + 0.2445 \ln m_{At-1} - 0.3586 \ln m_{Jt-1} \\
&\quad + 0.7836 \ln m_{Nt-1} - 0.3669 \ln m_{UKt-1} - 0.0490 \ln m_{USt-1} \qquad (5.5) \\
\ln m_{UKt} &= -2.8256 - 0.0114t + 0.4077 \ln m_{At-1} - 0.1778 \ln m_{Jt-1} \\
&\quad + 0.0014 \ln m_{Nt-1} + 0.2662 \ln m_{UKt-1} - 0.1084 \ln m_{USt-1} \\
\ln m_{USt} &= -2.4519 - 0.0114t + 0.0460 \ln m_{At-1} - 0.0841 \ln m_{Jt-1} \\
&\quad + 0.1007 \ln m_{Nt-1} - 0.4218 \ln m_{UKt-1} + 0.8198 \ln m_{USt-1}
\end{aligned}
$$

$$
\begin{aligned}
\ln m_{At} &= -2.5254 - 0.0197t + 0.3594 \ln m_{At-1} - 0.3343 \ln m_{Jt-1} \\
&\quad - 0.0376 \ln m_{Nt-1} + 0.0114 \ln m_{UKt-1} + 0.4956 \ln m_{USt-1} \\
\ln m_{Jt} &= -2.0194 - 0.0174t - 0.1264 \ln m_{At-1} + 0.3997 \ln m_{Jt-1} \\
&\quad - 0.1078 \ln m_{Nt-1} - 0.1194 \ln m_{UKt-1} + 0.5571 \ln m_{USt-1} \\
\ln m_{Nt} &= -4.2941 - 0.0212t + 0.0387 \ln m_{At-1} - 0.4094 \ln m_{Jt-1} \\
&\quad + 0.5934 \ln m_{Nt-1} - 0.3004 \ln m_{UKt-1} + 0.2444 \ln m_{USt-1} \qquad (5.6) \\
\ln m_{UKt} &= -3.8224 - 0.0139t + 0.2637 \ln m_{At-1} - 0.1042 \ln m_{Jt-1} \\
&\quad - 0.1184 \ln m_{Nt-1} + 0.2682 \ln m_{UKt-1} - 0.0573 \ln m_{USt-1} \\
\ln m_{USt} &= -2.5600 - 0.0116t - 0.0878 \ln m_{At-1} - 0.1135 \ln m_{Jt-1} \\
&\quad - 0.0034 \ln m_{Nt-1} - 0.3189 \ln m_{UKt-1} + 1.0191 \ln m_{USt-1}
\end{aligned}
$$

However, the time series are non-stationary. This means that the VAR(1) model cannot be used directly. The unit root tests confirm that the levels of mortality rates are all non-stationary and are integrated of order one, I(1).[1] Therefore is it

[1] For both the males and the females, the ADF and PP test p-values do not reject the null hypothesis of the existence of a unit root for the country mortality time series but they all reject

necessary to test for potential cointegration relationships indicating the existence of common trends.

5.4.3 Cointegration Tests

The Π matrices for males and females are given in equations (5.7) and (5.8). The VAR(1) models estimated have $\Pi = -(I - A_1)$ with determinant different from zero so that $r = k = 5$ and there are no cointegration relations.

$$\Pi_{Males} = \begin{bmatrix} 0.2118 & -0.2124 & 0.0573 & 0.2935 & 0.6563 \\ -0.0646 & 0.5365 & -0.0934 & -0.3354 & 0.3491 \\ 0.2445 & -0.3586 & 0.7836 & -0.3668 & -0.04898 \\ 0.4077 & -0.1778 & 0.0014 & 0.2662 & -0.1084 \\ 0.0460 & -0.0841 & 0.1006 & -0.4218 & 0.8198 \end{bmatrix} \tag{5.7}$$

$$\Pi_{Females} = \begin{bmatrix} 0.3594 & -0.3343 & -0.0376 & 0.0114 & 0.4956 \\ -0.1264 & 0.3997 & -0.1078 & -0.1194 & 0.5571 \\ 0.0387 & -0.4094 & 0.5934 & -0.3004 & 0.2444 \\ 0.2637 & -0.1042 & -0.1184 & 0.2682 & -0.0573 \\ -0.0878 & -0.1135 & -0.0034 & -0.3189 & 1.0191 \end{bmatrix} \tag{5.8}$$

From Triacca (2002) if the elements of m_t are not cointegrated then Πm_{t-1} is $I(1)$. This means that u_t are $I(1)$. From the unit root tests it was found that $\Delta \ln m_t$ are stationary so that Π must be the null matrix and hence, $\Delta \ln m_t = u_t$. It is sufficient to model the differences of $\ln m_t$ for these countries in a VAR model. There are no common stochastic country trends based on these results.

5.5 Discussion

It is observed that mortality improvements are not constant. They vary by country, by age group and with time. The analysis of mortality rates of different countries with comparable standards of living in this chapter gives a deeper understanding of the nature of trends in mortality by reducing the number of variables that explain mortality trends and changes to a smaller number of unobservable variables and orthogonal linear combination of observable variables using factor analysis and principal components analysis respectively. The following conclusions are drawn that are important when assumptions for modelling the randomness in aggregate mortality rates over time are formulated.

There are several hidden underlying variables that drive mortality and volatility in mortality as shown in the factor analysis in 5.3.1. Principal components analysis

this null hypothesis for the first differences.

in 5.3.2 showed that a small number of linear combinations (principal components) of age groups can explain most of the variation in mortality trends and mortality volatility. In both cases, dimension reduction is used to explain the variation of the data using a smaller number of variables and the factors or principal components both bring out patterns that portray a clear picture of features that are not directly discernible from the data.

Mortality trends in most of the countries analysed are explained by a similar number of factors and factors behave in a similar way for different countries in some instances. People in countries with high HDI's have comparable standards of living including access to high quality medical care and education that both help the population to stay healthy. They are also more likely to have similar causes of death such as cancer and obesity related diseases rather than accidental causes and opportunistic and easily preventable diseases. Mortality improvements in European countries, for example, were attributed to a decline in deaths from heart disease (Willets, 2004).

Geographical location may also contribute to similar factors, as seen in the similarity between the factors for Norway and the UK. However, only two observations from the same region are used in this analysis and therefore no conclusive answer can be given. Based on Andreev and Vaupel (2005) it is very possible that geographical location can lead to similar mortality patterns. In their study they found that geographically close countries had similar patters of mortality improvements.

Gender effect is less than country effect for mortality trends. Comparing the factors and their loadings on the basis of both gender and country using the visualisation in figure 5.4 the response of a factor on both genders is similar but the response differs for different countries.

Communalities of the factor models for the different countries measure how much of the variability is due to common factors and with the exception of Norway the percentage of common variation explained in most of the attributes or age groups is close to 100%. In the cases where a high percentage of the common variation is explained by the, K factors, the model of the mortality trends is the linear regression model estimated as:

$$\widehat{\ln m_{x,t}} = \mu_{x,t} + \sum_{j=1}^{K} \lambda_{xt,j} f_j \tag{5.9}$$

For time trends and cohort trends that represent the volatility in mortality trends factor analysis gives in-depth understanding on the unobservable forces that drive the randomness in mortality. The number of factors needed to explain the variation in time trends and cohort trends is much higher than the number required for the levels of mortality themselves. Intuitively, this is because the cohort trends and period trends carry additional information about the changes in mortality rates through time and by age or through time alone respectively.

More factors are required to model cohort trends than time trends. However, more principal components are required to model time trends than cohort trends. This is counter intuitive. This suggests that mortality changes are driven by hidden, unobservable, latent variables that generally cannot be computed as linear combinations of the original variables.

The percentage of common variation explained by the factors in the factor model is less for the time trends than the cohort trends. Information regarding the changes in mortality is gained by adding the year of birth which is a source of variation. This is at the cost of requiring additional factors since, as explained above, more factors are required to model cohort trends than time trends.

A 7 factor model describes the mortality curve of the Australian population with the percentage of common variation in each age group that is explained by the factors being close to 100% in all age groups under 99 for both males and females. The Heligman-Pollard model is an 8 factor parametric model (Sherris and Gaille, 2010a). However, the difference between the 7 factor model and the Heligman-Pollard model is that the former is a linear regression while the former is non-linear.

In the principal component analysis in this chapter, just as as in other principal components analyses such as Lee and Carter (1992) the first principal component is interpreted to be the trend in the data. The first principal component explains most of the variation in $\ln m_{x,t}$ for all countries but the proportion of variation explained for Norway is significantly less. The age groups used in this analysis are the same as those used in Lee and Carter (1992) and the method of principal components analysis is based on SVD; therefore the findings of the PCA in this analysis can easily be compared to those of the Lee-Carter model.

The first PC in the analysis of mortality trends is essentially k_t^{LC} in the Lee-Carter model and is not always modelled by a random walk with drift. Of the countries analysed only the USA has k_t^{LC} as a random walk with drift. A key problem of relying on the first principal component of a data set with non-stationary values is that the variation that the principal component detects is due to the variation in a similar direction. Studies that present solutions to this problem include Hatzopoulos and Haberman (2011) and O'Hare and French (2011). The approaches in both studies are quite different as the first recognises the non-stationarity of the data then weights the identified linear combinations (that are the principal components) with sparse vectors such that some of the loadings are weighted as zero improving the recognition of variance patterns. The second aims to generalise the model by extracting a larger number of components using dynamic PCA which caters for the dynamic structure of mortality data. The form of PCA conducted in this thesis would be classified as static PCA in O'Hare and French (2011).

When PCA is performed on the time trends the first PC fluctuates around zero. In some countries such as USA the intensity of the fluctuations are fairly consistent as time goes by while for other countries such as Japan and UK, and particularly for males in both these countries, there is a lot more volatility in the first 20 time periods than in the last 20. In Norway the volatility in the first principal component increases with time. A large number of principal components was required to explain 90% of the variation in the data with an average of about 8 principal components required. A section of O'Hare and French (2011) performs PCA on the first difference of the logarithms of the mortality rates of countries including three of the countries in analysed in this study - Australia, UK and USA. For USA males O'Hare and French (2011) finds that in order to explain 72% of the variation in ten (static) principal components are required as compared to 4 principal components in this study. The difference in the number of principal components is attributable partly to the different periods of time 1950-2000 but more to the age range (single ages 20-89).

The cohort trends follow an individual aged x at time t to age $x+1$ at time $t+1$ and show that mortality improvements increasing at a decreasing rate. There were periods of high increase earlier but in recent times the increase is not so much. Mortality improvements affect the very young and the very old to a great extent. Japan has improvements over more ages than the other countries. In relation to the findings in Willets (2004), for data on the UK, that average annual mortality improvements in the 1990s were lower than those in the 1980s, the PCA on cohort trends showed that the mortality trends improvements were approximately constant after 2000 (see table in figure 5.18). The first principal component of the cohort trends showed that there was a slowing in the rate of improvement of mortality in recent years. Principal components analysis has fixed loadings of the principal components and does not show an important observation from Willets (2004) regarding the ageing of mortality improvement where the ages with the greatest mortality improvements are becoming older ages as time goes by. For Australian males this is shown in figure 4.15.

The findings from factor analysis and principal components analysis reduce re-

Age group	1960s	1970s	1980s	1990s
25-29	1.3%	0.1%	0.4%	−1.0%
30-34	1.5%	1.5%	−0.6%	−0.9%
35-39	1.5%	1.0%	0.2%	1.0%
40-44	−0.2%	2.2%	2.2%	0.6%
45-49	−0.1%	1.5%	2.4%	1.1%
50-54	0.0%	0.6%	3.2%	2.5%
55-59	0.9%	1.1%	3.1%	2.4%
60-64	0.6%	0.9%	1.7%	3.2%
65-69	−0.0%	1.4%	1.8%	3.1%
70-74	−0.0%	1.1%	1.5%	1.9%
75-79	0.5%	0.4%	1.5%	2.0%
80-84	1.5%	−0.1%	1.4%	1.4%
85 and over	−0.2%	0.7%	1.2%	0.5%

Figure 5.18: Average annual rate of mortality improvement from Willets (2004) for UK males by age group and decade. These results give a general picture of the period trend and cohort trend on a broad time horizon. Period trends (by going across the rows) neither increase nor decrease while cohort trends (by following an age group as it ages e.g. those aged 25-29 in the 1960s will be part of those aged 35-39 in the 1970s). The mortality of those aged 25-29 and 30-34 in the 1960s improves on average by 1.3% and 1.5%, then when they are aged 35-39 and 40-44 in the 1970s it improves are by 1.0% and 2.2% respectively and in 1980s when they are aged 45-49 and 50-54 it improves by 2.4% and 3.2% but in the 1990s when they are aged 55-59 and 60-64 the mortality improvements remain at 2.4% and 3.2%. Mortality improvements increased at a decreasing rate.

liance on expert opinion[1] and allow the data to dictate how many factors should be used in a mortality model by searching for the number of hidden variables that drive mortality. This approach is generalizable is an much as the suitable k-factor model is easily determined without having to test several curves to find the one that gives the best fit, for example in the process of developing the 2-factor model in Cairns et al. (2006b).

Mortality trends are mostly I(1) while the time trends and period trends are I(0). This is not surprising because mortality rates improve by accumulating the effects of shocks such as medical improvements. Surprisingly, there are no common stochastic country trends based on these results. A similar study of mortality trends but using cause-specific data is Sherris and Gaille (2010b). Although the countries used are not identical to those in this study, the findings are similar to those in this thesis and Sherris and Njenga (2009). The logarithms of cause-specific mortality trends were mostly non-stationary with stochastic trends and also found to be best modelled using a VAR(1). The causes of death within the countries were found to be cointegrated. The cross-country relationships were not studied. Due to the existence of similar cointegration relations within countries with similar mortality

[1] Wong-Fupuy and Haberman (2004) partly credit the underestimation of mortality rates to excessive reliance on expert opinion

experiences including Australia, UK and USA it was reasonable to expect to find common stochastic country trends in the cross-country data analysed in this thesis.

However, there is an expected finding regarding PCA that is influenced by the non-stationarity of the data. An analysis of the implications of PCA of non-stationary time series has been published in Lansangan and Barrios (2009) and it was found that when non-stationary data drift simultaneously in the same direction there can be empirical correlations. This leads to similar variance patterns. As a result, the first principal component will combine all the variables into a single component. This ties in with the observation in Alexander (1999) that if the variables are cointegrated then the first principal component is the common stochastic trend. The analysis of the mortality trends showed that one principal component was sufficient to explain most of the variation in the data. This finding could certainly be attributed to the non-stationarity of the data. One solution to this problem is to use a technique called Sparse PCA (Lansangan and Barrios, 2009) as done for mortality trend analysis in Hatzopoulos and Haberman (2011).

These findings have various implications for the process of mortality model development. First, More than one factor needed to model mortality trends and volatility for all the countries analysed. Secondly, because the behaviour across the different countries is not always similar, it is important to analyse the features of a mortality data set and use a mortality model that is suitable. For example, if a relatively small amount of variation in the data set is explained by the first principal component then it is not suited for use with Lee-Carter model. A third finding is that some factors and principal components that impact changes to mortality rates are common to certain age-groups, certain cohorts and the entire population. Further, some factors and principal components are similar across countries, especially those in the same geographical region. Therefore, it is useful to consider how mortality is changing internationally in countries with comparable living standards.

5.6 Conclusion

The analysis in this chapter provides a basis for the development of country and age-based longevity risk models that capture trends including common trends across age, sex and country. It also provides a basis for assessing longevity risk in a consistent modelling framework. Stochastic common trends can be included in the Lee-Carter model consistent with a difference stationary model but the Lee-Carter model does not include a sufficient number of random principal components to drive mortality changes.

Mortality rates are found to have stochastic trends for almost all ages and across all the countries in the study. This means that trends in the historical rates are stochastic and shocks are permanent. Volatility increases through time as shocks accumulate in the series. Multiple factors or principal components are driving mortality changes. The number of factors driving changes in the mortality rates is similar across countries.

This analysis also demonstrated how the standardized mortality rates across countries have stochastic trends based on the historical data. These stochastic trends were not common to all the countries in this analysis. Modelling the differences of the standardized mortality rates is sufficient to capture trends for each country at the aggregate level. This has implications for international diversification of longevity risk since there are no common long run relationships between countries mortality improvement and potential for risk diversification across countries.

6

Modelling Australian Mortality using a Bayesian Vector Autoregression Model

Introduction

In chapter 5 mortality risk models have been developed to capture trends and common factors driving mortality improvement. That analysis indicated the need for at least 7 random factors to model Australian mortality and that some ages are potentially cointegrated. In this chapter the application of VECM and VAR models to Australian male and female age specific mortality rates is developed.

There have been limited studies applying multivariate econometric modelling techniques to mortality data allowing for cointegration and non-stationarity for a range of ages. Early application of time series to mortality data appears in McNown and Rogers (1989). In order to reduce the number of random factors driving mortality changes over time a parameterized mortality model is cross sectionally estimated at a series of points in time and the evolution of the parameters is modelled as a VAR/VECM system. This not only reduces the dimension of the random variability but allows for smoothing across ages and improved forecasting performance of the model.

Although time series techniques have been applied to model parameters in various parametric mortality models, there has been limited analysis of parameter risk using Bayesian techniques. Previous studies such as McNown and Rogers (1989) that attempted to develop a dynamic parametric mortality model by modelling the evolution of parameters of a parametric model, such as the Heligman-Pollard model were plagued by a myriad of shortcomings including being unable to generate forecasts that are consistent. VAR models allow for dependence between the parameters of the Heligman-Pollard model. They are flexible and reflect trends in the data well, giving better forecasts of the parameters. However, VAR models are prone to over-parameterization.

This study additionally uses a Bayesian Vector Autoregressive (BVAR) model to fit and predict the parameters of the Heligman-Pollard model as estimated for the Australian population. Therefore, this chapter innovatively develops a dynamic parametric model by incorporating a Bayesian Vector Autoregression (BVAR) model that provides a compromise between over-parameterization (VAR models) and under-parameterization (univariate models). BVAR models are shown to significantly improve the forecast accuracy of VAR models for mortality rates based on Australian data. The Bayesian model allows for parameter uncertainty which is a significant component of total mortality risk. The resulting forecasts readily incorporate parameter uncertainty. Parameter risk is quantified and shown to be a significant component of total mortality uncertainty. This chapter is based on Sherris and Njenga (2009) and Sherris and Njenga (2011).

6.1 Background

With the passage of time mortality is generally higher for the youngest and the oldest (see figures 4.13(a) and 4.13(b)) and has a pattern similar to that shown in figure 2.1. The structure of mortality age patterns in Australia changed in the 1970s. This was described in section 4.1.1.1. Parametric mortality models for mortality rates capture the trends and volatility of large body of data using a small number of parameters. In addition, econometric models provide a more general framework for modelling mortality trends and volatility. This study combines parametric mortality models and econometric models to develop a model of Australian mortality.

6.2 Correlation of Heligman-Pollard parameters and Unit Root Tests

Consider the Heligman-Pollard parameters estimated in section 4.2.2.1. The parameters of the Heligman-Pollard model are found to be highly correlated which is consistent with Hartmann (1987). The correlation matrix of the parameters are shown in Table 6.1 and 6.2 for males and females respectively.

The parameters with significant correlation have p-values <0.001. With the exception of D for males, denoted as D_m, there is significant correlation between parameters in different terms of the Heligman-Pollard model (Equation (2.16)). Denoting the parameters for males with a subscript m and the parameters for females

	A	B	C	D	E	F	G	H
A	*****	-0.128	0.572	0.204	0.268	-0.697	0.885	-0.824
B	0.322	*****	0.605	-0.299	-0.5	0.472	-0.374	0.368
C	<0.001	<0.001	*****	0.02	-0.014	-0.245	0.423	-0.416
D	0.111	0.018	0.876	*****	0.739	-0.663	0.372	-0.312
E	0.035	<0.001	0.911	<0.001	*****	-0.769	0.608	-0.614
F	<0.001	<0.001	0.055	<0.001	<0.001	*****	-0.874	0.84
G	<0.001	0.003	0.001	0.003	<0.001	<0.001	*****	-0.968
H	<0.001	0.003	0.001	0.013	<0.001	<0.001	<0.001	*****

Table 6.1: Correlation and Significance Males: upper diagonal part contains correlation coefficient estimates; lower diagonal part contains corresponding p-values. Parameters with significant correlation have p-values <0.001.

	A	B	C	D	E	F	G	H
A	*****	0.104	0.567	0.512	-0.133	0.358	0.813	-0.401
B	0.42	*****	0.785	-0.084	-0.152	0.048	-0.056	0.07
C	<0.001	<0.001	*****	0.145	-0.022	0.079	0.462	-0.297
D	<0.001	0.517	0.262	*****	-0.127	0.714	0.138	0.204
E	0.304	0.237	0.863	0.325	*****	-0.583	0.265	-0.517
F	0.004	0.709	0.543	<0.001	<0.001	*****	-0.122	0.475
G	<0.001	0.665	<0.001	0.285	0.037	0.343	*****	-0.794
H	0.001	0.588	0.019	0.112	<0.001	<0.001	<0.001	*****

Table 6.2: Correlation and Significance Females: upper diagonal part contains correlation coefficient estimates; lower diagonal part contains corresponding p-values. Parameters with significant correlation have p-values <0.001.

with a subscript f it is inferred from tables 6.1 and 6.2 that A_m is significantly correlated with C_m, F_m, G_m and H_m; B_m with C_m, E_m and F_m; D_m with E_m and F_m; E_m with F_m,G_m and H_m; F_m with G_m and H_m; and G_m with H_m. For females the number of significant correlations are fewer but there is still significant correlation between the parameters in the different terms in equation (2.16) since A_f is significantly correlated with C_f, D_f and G_f; C_f with B_f and G_f; and F_f with H_f. This shows how old age mortality is correlated with mortality at younger ages because each term in equation (2.16) represents mortality at each of three age ranges. Recall that parameters A-C measure mortality of the very young, D-F measure mortality of the middle ages and G-H measure mortality of the elderly.

6.3 Cointegration of the Heligman-Pollard Parameters

Unit root tests for the parameters indicate they are at most I(1) with a constant and a trend. Additionally, testing for difference stationarity confirms that all the time series of the Heligman-Pollard Parameters are I(1) except for the female B parameter. The parameters are modelled as a stochastic system using a VAR(p). In order to improve the model fit and to ensure the parameters are positive the logarithms of the parameters are modelled.

A VAR(p) model is fitted to the logarithms of the time series of the Heligman-Pollard parameters for females and males using a model with an unrestricted constant as there is the presence of a constant and the presence of a trend in many of the variables in the model. These are denoted by VAR(p)$_f$ and VAR(p)$_m$ for females and males respectively. $\theta_{t,f} = (A_{t,f}, B_{t,f}, C_{t,f}, D_{t,f}, E_{t,f}, F_{t,f}, G_{t,f}, H_{t,f})$ and $\theta_{t,m} = (A_{t,m}, B_{t,m}, C_{t,m}, D_{t,m}, E_{t,m}, F_{t,m}, G_{t,m}, H_{t,m})$ are the variables included in VAR$(p)_f$ and VAR$(p)_m$ which model the VAR(p) for females and males.

The number of lags p that minimize the criteria in Equation (4.3) is 2. The diagnostic tests verify that a VAR(2) for the logarithms of the parameters is adequate for both males and females.

Since the time series of the variables (time series of the Heligman-Pollard model, $\theta_{t,m}$ and $\theta_{t,f}$) in the VAR(p) have unit roots then it is also necessary to fit a VECM. Since a VAR(2) is suited to model the time series of the parameters, the corresponding VECM is of lag $p-1 = 2-1 = 1$.

$$\Delta\theta = (\Omega_1 + \Omega_2 - I)\theta_{t-1} - \Omega_2 \Delta\theta_{t-1} + \epsilon_t \tag{6.1}$$

The next step involves testing for the existence of a stationary linear combination of the time series in the VECM. This simply requires the rank of the long-run impact matrix $\Pi = (\Omega_1 + \Omega_2 - I)$. Since the sample size is small[1] the trace test statistics are used to make inference on the number of cointegrating relationships present.

The results of the cointegration tests are shown in tables 6.3 and 6.4. In each case the LR_{test} rejects the null that $r_0 = 0$ at 1% significance and rejects the null that $r_0 = 1$ at 5% significance. The conclusion is that cointegration exists between the parameters of the Heligman-Pollard model. The tests show that at 99% confidence a VECM with one cointegration relation is required for both males and females. The

[1]The eigenvalue test is often regarded as superior but for it to be reliable it requires a large sample size (c.300 observations)

	Eigenvalue	Trace Stat	95% CV	99% CV
H(0)++	0.63	193.35	170.80	182.51
H(1)+	0.58	145.86	136.61	146.99
H(2)	0.47	104.52	104.94	114.36
H(3)	0.42	73.82	77.74	85.78
H(4)	0.36	47.30	54.64	61.24
H(5)	0.28	25.62	34.55	40.49
H(6)	0.17	9.65	18.17	23.46
H(7)	0.01	0.62	3.74	6.40

Table 6.3: Female Cointegration Tests. At 99% confidence one cointegration relation are required (However, at 95% confidence two cointegration relations are required).

	Eigenvalue	Trace Stat	95% CV	99% CV
H(0)++**	0.79	221.13	170.80	182.51
H(1)+	0.61	146.81	136.61	146.99
H(2)	0.48	101.38	104.94	114.36
H(3)	0.40	69.91	77.74	85.78
H(4)	0.33	45.05	54.64	61.24
H(5)	0.26	25.49	34.55	40.49
H(6)	0.20	10.88	18.17	23.46
H(7)	0.00	0.06	3.74	6.40

Table 6.4: Male Cointegration Tests. At 99% confidence one cointegration relation is required.

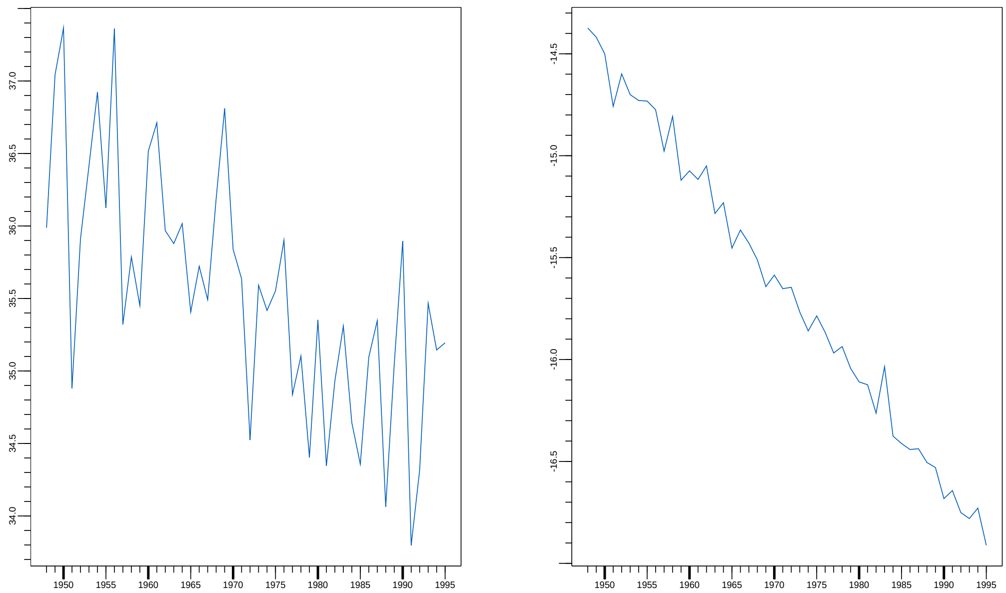
Trace tests significant at the 5% level are flagged by ' +'.
Trace tests significant at the 1% level are flagged by '++'.
Max Eigenvalue tests significant at the 5% level are flagged by ' *'.
Max Eigenvalue tests significant at the 1% level are flagged by '**'.

parameters of the Heligman-Pollard Model when modelled as a system are affected by at least one common stochastic trend. The common trends for males and females are presented in figure 6.1. The male common trend is a steady decline while the female common trend is an erratic decline.

	$\ln A$	$\ln B$	$\ln C$	$\ln D$	$\ln E$	$\ln F$	$\ln G$	$\ln H$
Females	1	-0.03	-0.87	-0.05	-0.57	-0.63	0.59	-24.09
Males	1	3.01	-12.74	-1.84	1.29	6.34	3.96	313.79

Table 6.5: Cointegrating Vectors β_f and β_m of the first cointegration relationship normalized on $\ln A$.

Table 6.5 shows the normalized cointegration vector for females and males respec-



(a) The Female Cointegration Relationship (b) The Male Cointegration Relationship.

Figure 6.1: The Cointegration relationships for males and females behave differently. The male cointegration relationship is a steady decline while the female cointegration relationship is erratic but has a general downward trend.

tively, that indicates that there exist linear combinations of the variables that are stationary (these are presented in figures 6.2 and 6.3). In particular, it is noteworthy that when parameter A_f increases, all the parameters except G_f also increase. However, when parameter A_m increases, only C_m and D_m increase.

For forecasting the VECM is transformed back to a VAR.

A VAR(2) model is determined to be the appropriate model for the logarithms of the parameters of the Heligman-Pollard model.

The existence of unit roots and cointegrated relations in the system of mortality parameters has also been confirmed.

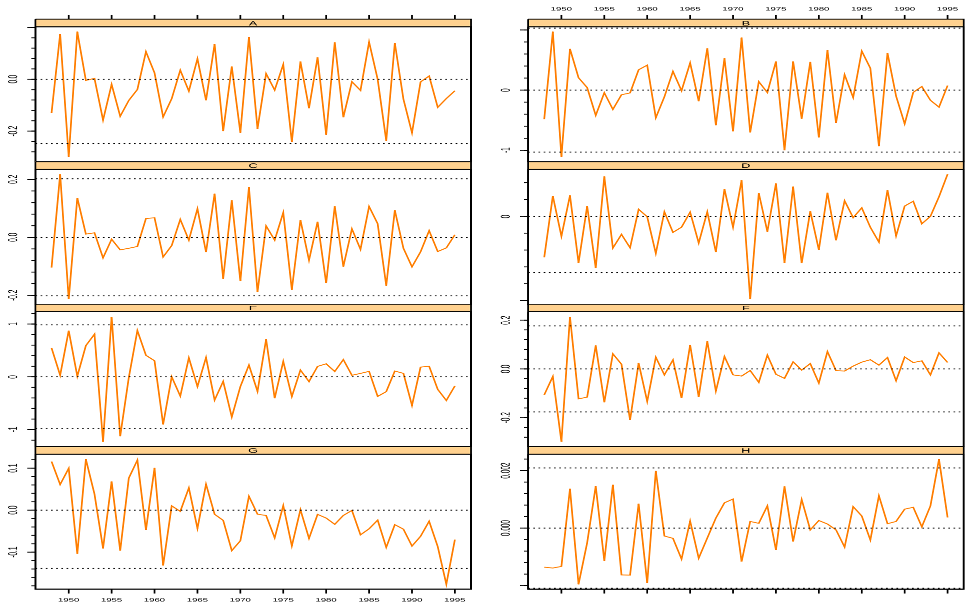


Figure 6.2: Stationary Linear Combinations of the logarithms of Heligman-Pollard parameters for Females

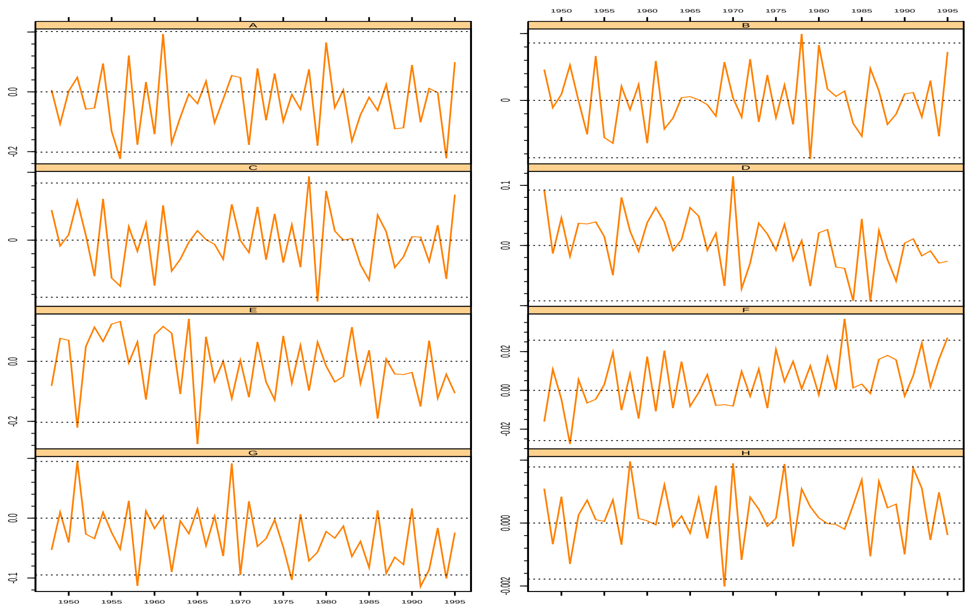


Figure 6.3: Stationary Linear Combinations of the logarithms of Heligman-Pollard parameters for Males

6.4 HP-VAR and HP-BVAR Model

The results of this study so far have yielded information necessary for estimating both VAR and BVAR models for the parameters of the Heligman-Pollard model. The estimated VAR(2) models for males and females are shown in figures 6.4 and 6.5 respectively. In these figures the parameters of the VAR(2) models are estimated by OLS and have no uncertainty in them. The VAR(2) models are also sensitive to the existence of unit roots and cointegrated relationships and must be extended to account for them.

The Bayesian VAR model used is based on a Sims-Zha (Normal-Inverse Wishart) prior. The choice of prior, as explained in the methodology, has several advantages such as improved forecasts because it has provisions for unit roots and cointegration in systems that are based on non-stationary variables; varying degrees of strength in the beliefs of the prior hyper parameters can be specified; and minimal computation effort. Importantly, for projection purposes, the chosen prior ensures that an analytic posterior will be obtained.

In modelling a B-VAR, unit roots and cointegration are captured by the priors through the hyperparameters. There is no need for explicit modelling to be included. Since there are no values available from prior mortality studies a range of hyperparameters were considered by comparing the forecasted variables to the observed "future" variables (data for 1996-2007). The best model performance was found to be for the set ($\lambda_0 = 0.9$, $\lambda_1 = 0.1$, $\lambda_2 = 1$, $\lambda_3 = 1$, $\lambda_4 = 0.05$, $\mu_5 = 5$, $\mu_6 = 5$).[1] The interpretations of the selected hyperparameters are as follows:

1. $\lambda_0 = 0.9$ - the model strays from a random walk.
2. $\lambda_1 = 0.1$ - the diagonal elements of the coefficient matrix for the first lag tend to one and all other elements tend to zero.
3. $\lambda_3 = 1$ - the parameters contained in higher lags vary less around their conditional mean of zero.
4. $\lambda_4 = 0.05$ - constants tend to zero.
5. $\mu_5 = 5$ - increased likelihood that the model can be expressed in first differences.

[1]Over 7000 possible combinations of hyperparameters were considered. Other combinations with small RMSE when compared to the observed variables were also considered including ($\lambda_0 = 0.9$, $\lambda_1 = 0.4$, $\lambda_2 = 1$, $\lambda_3 = 0.5$, $\lambda_4 = 0.2$, $\mu_5 = 6$, $\mu_6 = 5$) were also considered and the resulting errors were in the same range.

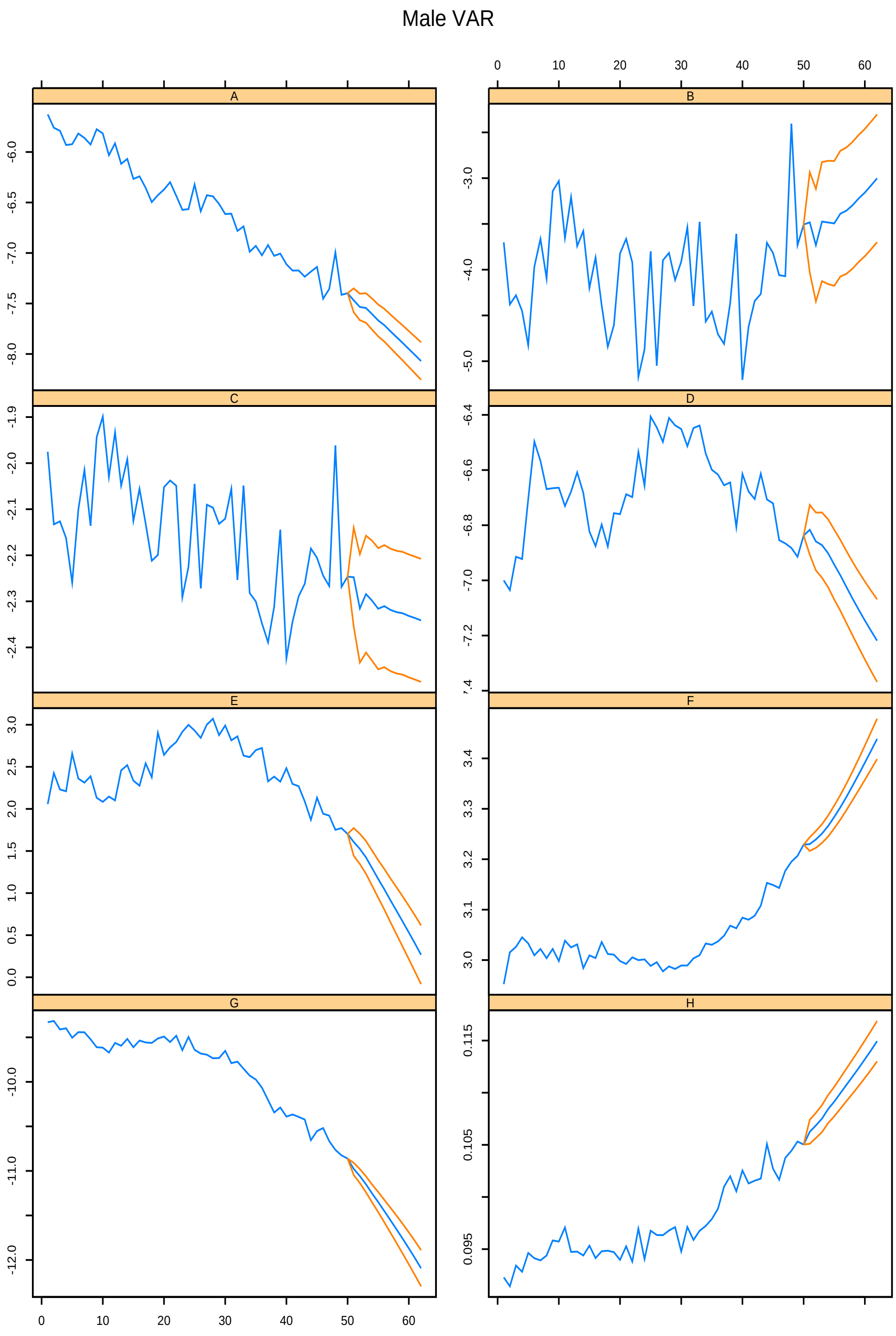


Figure 6.4: Fitted Values (1946-1995) and Projected Values (1996-2007) of $\ln \theta_{t,f}$ using VAR(2) Males

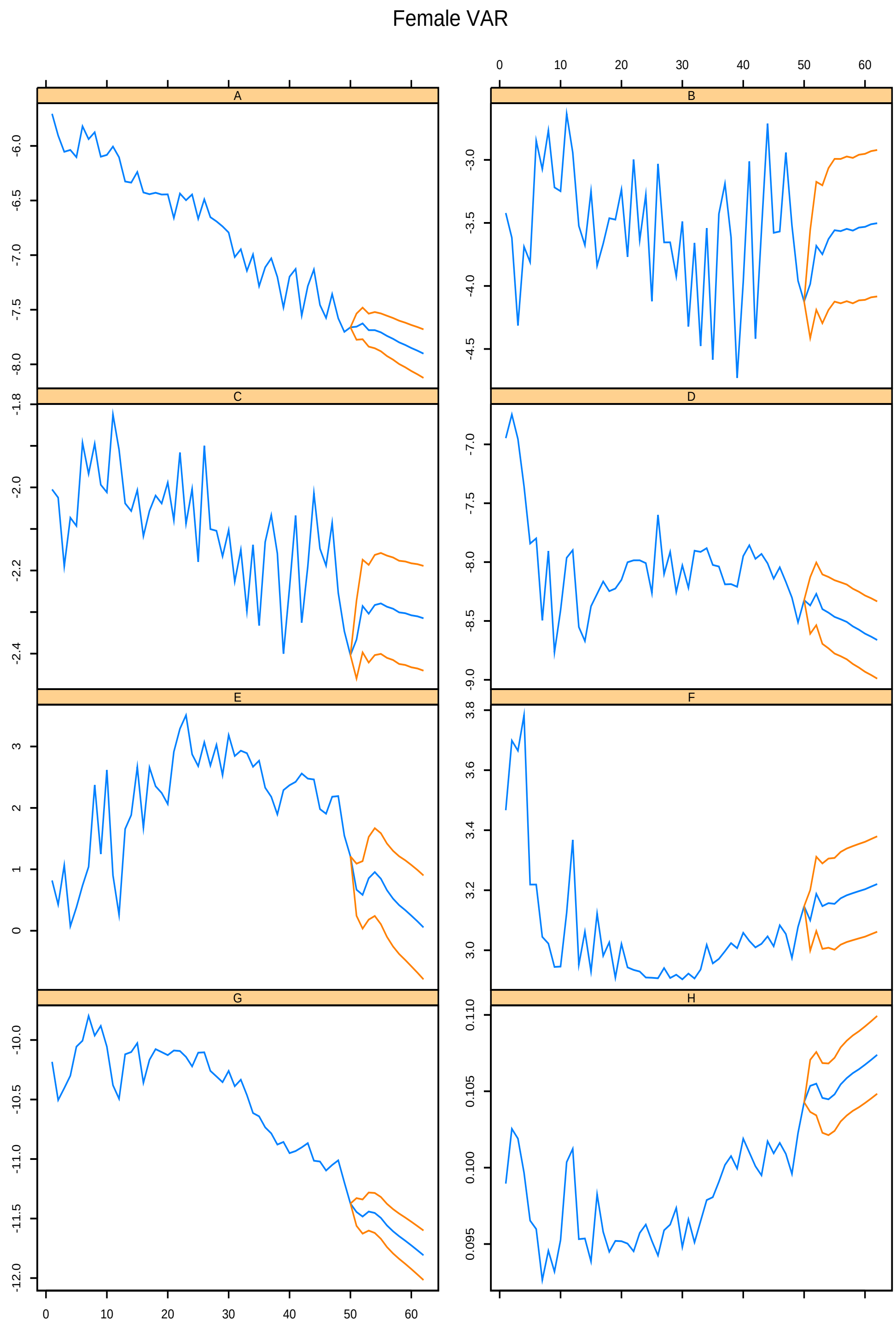


Figure 6.5: Fitted Values (1946-1995) and Projected Values (1996-2007) of $\ln \theta_{t,f}$ using VAR(2) Females

6. $\mu_6 = 5$ - likelihood of co-integration is given more weight.

The BVAR(2) is written as an eight variable version of (2.47) with prior means specified as an eight variable version of equation 2.48. The parameters of the BVAR model will be non-zero only if the variable contains information.

The predictions from fitting 50 years (1946-1995) of data are compared to the out-of-sample 12 years (1996-2007) of data that are available for the VAR model for females are displayed in Figure 6.5. The equivalent results for the BVAR(2) for females are shown in Figure 6.6. The VAR model shows an initial increase in parameter uncertainty that settles down to a long run distribution for each of the parameters. After an initial time period the parameter risk would be considered as having reached its maximum.

The results for the BVAR model show a significantly higher level of parameter risk and over the same horizon has not reached a long run distribution for the parameters. It is clear that parameter uncertainty is very significant and prediction intervals that take into account parameter risk are significantly higher than the case where this risk is not included. Figure 6.7 shows the equivalent BVAR(2) results for males confirming similar results as for females.

An advantage of the BVAR model is that parameters are estimated with greater accuracy as well as reflecting the true uncertainty in the predictive distribution. Table 6.6 shows this. Comparing the estimated parameters of the VAR model and the estimated parameters of the BVAR model reveals that the constant, c, is negligible in all the parameters. The values in columns 2-9 of table 6.6 are Ω_1 in equation (2.36). The elements along the diagonals have the greatest value in the BVAR model. This represents the strong belief that the variables follow a random walk. These were incorporated by setting the standard deviation around ω^i_{ij} as $\lambda^h_1 = 0.1$. Since the old age mortality is of interest, consider the equations for $\ln G$ and $\ln H$. The value of $\ln G$ depends mainly on its own past values and the past values of $\ln H$ while the value of $\ln H$ depends mainly on its own past values.

This is intuitively a reflection of $\ln H$'s interpretation as the relative rate of change of mortality rates. From the analysis in Chapter 5, the first principal component of $\Delta_d \ln m_{x,t}$ as shown in figure 5.17 has an evolution that is similar but not identical to that of $\ln H$. The magnitude of the coefficient of $\ln H$ in both VAR and BVAR is significant. The magnitude of the coefficients $\ln H.lag1$ in the VAR model is very large except in the equation for $\ln H$ itself. This is because VAR models look for long-term permanent movements (Sims et al., 1990) such as those in $\ln H$. BVAR checks if the effect of a shock is permanent or fades with time (Sims and Zha, 1998).

The effect of the shock, measured in the coefficient, due to movements in $\ln H$ on the variables other than $\ln H$ should not be significantly greater than the shock due to movements in the variable itself. For example, consider equation $\ln A$ in VAR_f. The coefficient of the lagged value of $\ln A$ at lag 1 is $\ln A.lag1 = 0.6859$ for females and the coefficient of the lagged value of $\ln H$ at lag 1 is $\ln H.lag1 = 96.2176$. The shock due to movements in $\ln H$ is much stronger. In comparison, the equation $\ln A$ in BVAR_f has $\ln A.lag1 = 0.9584$ and $\ln H.lag1 = -0.2109$. All the variables in the BVAR still depend on the past values of $\ln H$ but the shock due to movements in $\ln H$ is reduced for all variables. Therefore, it is expected that the VAR model will produce erratic forecasts (Sims and Zha, 1998) because it assumes that the shock is long term and permanent. The inability of the VAR model to check if the shocks persist or fade with time further means that VAR model parameter estimates often lack intuitive interpretation.

The parameters of the BVAR model result in a more parsimonious model. As mentioned in section 2.3.4 a key shortcoming of the VAR model is that it is over-parameterized. From table 6.6 some of the non-zero parameters of the VAR model are zero parameters of the BVAR model especially in the vector of constants, c and the coefficients at the second lag. This difference is because BVAR creates a barrier that prevents its parameters from being non-zero unless they contain information unlike VAR which estimates its parameters by use of ordinary least squares. As a result, the BVAR model is more parsimonious than the VAR model.

For longevity risk the uncertainty at the older ages is of most interest. Figure 6.8 shows the observed and predicted parameter values for old age mortality using both the VAR and BVAR for males and females. The observed values are indicated by the green line, the VAR predictions by the red line and the B-VAR predictions by the blue line. $\ln(H)$ is the relative rate of old age mortality increase. This should reflect the decline in the rate of mortality increase after the mid-1990s. The plots show how the relative rate of change in mortality, given by $\ln(H)$, is over-estimated by the VAR whereas the Bayesian-VAR captures the reducing rate of change of old age mortality.

VAR models produced less accurate estimates when forecasted. The confidence bands for the estimated parameters are shown by the thin broken and thin dotted lines. The thick red solid line is the observed value of the parameter estimated for that year, the thick black broken line is the BVAR (with parameter risk) estimate of the parameter for that year and the thick blue dot-dashed line is the VAR (without parameter risk) estimate of the parameter for that year. The observed parameter value (thick red line) should fall within the confidence bands of the density of the

VAR_f	c	lnA.lag1	lnB.lag1	lnC.lag1	lnD.lag1	lnE.lag1	lnF.lag1	lnG.lag1	lnH.lag1	lnA.lag2	lnB.lag2	lnC.lag2	lnD.lag2	lnE.lag2	lnF.lag2	lnG.lag2	lnH.lag2
lnA	2.7849	0.6859	-0.3366	0.8078	0.0069	-0.078	-0.0689	1.5703	96.2176	-0.0862	-0.1072	0.6698	-0.0316	-0.0626	0.2077	-1.4953	-132.0613
lnB	12.1821	0.8679	-0.8244	3.5019	0.2456	-0.2632	-0.3132	6.7306	390.6899	-0.7038	0.0177	1.2511	-0.0061	-0.1985	-0.064	-9.5697	-727.9425
lnC	1.3804	-0.057	-0.1265	0.7734	0.0169	-0.0278	-0.1019	1.6686	110.6891	0.0607	-0.1126	0.5251	-0.009	-0.0463	-0.0152	-1.851	-140.2451
lnD	-4.2349	-0.8283	0.0361	-0.1261	0.2589	0.0399	-0.0211	1.5385	144.1147	0.8145	-0.2725	0.1966	0.2621	0.0231	-0.0112	-1.2781	-121.3759
lnE	22.7082	-2.1337	-0.5828	4.6311	-1.0031	0.3188	0.1177	-0.1632	-35.5825	-0.1471	-1.0076	3.6633	1.4199	-0.2537	-3.1969	3.1735	238.6835
lnF	0.8106	0.1335	0.0162	-0.3199	0.2225	0.0036	-0.0328	0.164	27.5631	0.6323	-0.114	0.116	-0.1028	0.0183	0.352	-1.6556	-120.0886
lnG	-0.4342	-0.2967	-0.1608	1.1058	-0.1327	-0.0004	0.2405	1.5272	66.857	-0.2849	-0.0139	0.166	0.1297	-0.0561	-0.3044	0.7602	58.9427
lnH	-0.0061	0.0043	0.0021	-0.0152	0.0013	0.0002	-0.0033	-0.0168	-0.5172	0.004	-0.0004	-0.0005	-0.0021	0.0008	0.0061	-0.0014	-0.2494

$BVAR_f$	c	lnA.lag1	lnB.lag1	lnC.lag1	lnD.lag1	lnE.lag1	lnF.lag1	lnG.lag1	lnH.lag1	lnA.lag2	lnB.lag2	lnC.lag2	lnD.lag2	lnE.lag2	lnF.lag2	lnG.lag2	lnH.lag2
lnA	-0.0002	0.9584	-0.028	-0.0954	-0.019	-0.0019	-0.056	0.0284	-0.2109	0.042	0.0252	0.0948	0.0197	-0.0107	0.0556	-0.028	0.1785
lnB	-0.0005	-0.1335	0.8887	-0.3707	-0.0761	0.0149	-0.1783	0.0856	2.9977	0.1343	0.0993	0.3668	0.0777	-0.0246	0.1763	-0.0845	-3.107
lnC	-0.0001	-0.0259	-0.0213	0.9241	-0.0158	0.0038	-0.0403	0.0193	0.4725	0.0262	0.019	0.0752	0.0161	-0.0071	0.0399	-0.0191	-0.4957
lnD	-0.0004	-0.0534	-0.0321	-0.1225	0.9227	-0.0035	-0.0468	0.0177	1.0263	0.0551	0.033	0.1257	0.078	0.0236	0.0437	-0.0168	-1.1237
lnE	0.0008	-0.0916	-0.0132	-0.0578	-0.224	0.8901	0.1687	0.0661	-7.0796	0.0896	0.0118	0.0541	0.2202	0.0571	-0.1663	-0.0672	7.1986
lnF	-0.0002	-0.0081	-0.0056	-0.0177	0.0201	0.0158	0.9193	-0.0045	1.0732	0.0086	0.0067	0.0193	-0.0195	-0.0038	0.0794	0.0048	-1.1083
lnG	0.0002	-0.0027	-0.0007	0.0063	-0.0083	-0.0197	0.0538	0.9942	0.6524	0.0024	-0.0006	-0.0076	0.008	-0.0045	-0.052	0.0054	-0.6035
lnH	0	0	0	-0.0001	0.0002	0.0003	-0.0008	0.0001	0.9852	0	0	0.0001	-0.0002	0	0.0008	-0.0001	0.0141

VAR_m	c	lnA.lag1	lnB.lag1	lnC.lag1	lnD.lag1	lnE.lag1	lnF.lag1	lnG.lag1	lnH.lag1	lnA.lag2	lnB.lag2	lnC.lag2	lnD.lag2	lnE.lag2	lnF.lag2	lnG.lag2	lnH.lag2
lnA	-1.2646	0.5625	0.1854	-1.4227	-0.1851	-0.0484	2.1264	0.3902	23.7682	0.2857	-0.342	1.4633	0.5037	-0.0986	-0.1037	0.443	19.4773
lnB	3.4734	0.6575	1.9911	-10.3217	1.0621	-0.7431	5.554	0.5912	1.355	-0.6266	-2.0902	11.7959	0.7617	-0.2019	-1.8982	1.437	191.3919
lnC	1.6076	0.0616	0.3006	-1.4716	0.261	-0.1563	0.7405	0.3376	15.2666	-0.0963	-0.4365	2.4097	0.0391	-0.0801	-0.5485	0.2911	42.8929
lnD	-7.683	0.0963	0.1277	-0.7094	0.3414	0.139	1.3471	0.4867	21.1418	-0.0149	-0.0162	0.0519	0.1871	0.1442	0.0097	0.0012	21.0224
lnE	0.378	-0.3702	-0.1176	0.8857	-0.3909	0.4049	-0.8032	0.0253	23.7933	-0.1667	-0.0158	0.2298	0.1764	0.2594	1.6093	2.0437	133.1608
lnF	1.1445	0.0077	0.015	-0.0759	0.0102	-0.0518	0.2897	-0.0173	-4.3073	-0.006	-0.0275	0.1645	-0.0417	0.0268	0.1338	-0.2185	-13.2135
lnG	-0.7553	-0.1365	0.0437	-0.105	-0.0031	0.0582	-0.5758	1.3181	71.2391	0.1335	-0.1405	0.7253	-0.0777	-0.0907	0.26	0.2476	7.9704
lnH	0.0095	0.0014	-0.0014	0.0067	-0.0008	-0.0001	0.0017	-0.0148	-0.7626	-0.0009	0.0022	-0.0123	0.0005	0.001	0.0007	0.0032	0.316

$BVAR_m$	c	lnA.lag1	lnB.lag1	lnC.lag1	lnD.lag1	lnE.lag1	lnF.lag1	lnG.lag1	lnH.lag1	lnA.lag2	lnB.lag2	lnC.lag2	lnD.lag2	lnE.lag2	lnF.lag2	lnG.lag2	lnH.lag2
lnA	-0.0002	0.9402	-0.014	-0.0751	-0.0489	0.0111	0.2096	-0.0109	-0.3581	0.0609	0.0122	0.0756	0.0493	-0.013	-0.2111	0.0115	0.2968
lnB	-0.0006	-0.1722	0.92	-0.4452	0.0779	0.0855	0.8295	-0.123	1.3494	0.1728	0.0689	0.4398	-0.0763	-0.0888	-0.8329	0.124	-1.4427
lnC	-0.0001	-0.0306	-0.0147	0.9164	0.0246	0.0203	0.146	-0.0209	0.2922	0.0309	0.0126	0.0828	-0.0242	-0.021	-0.1467	0.0211	-0.3153
lnD	0	0	0.0009	0.0026	0.9638	0.0084	0.0536	0.0495	-2.9827	0.0003	-0.0018	-0.0029	0.036	-0.0077	-0.0535	-0.0494	2.9749
lnE	0.0001	-0.01	0.0041	0.0326	-0.093	0.9262	-0.2281	0.0302	-1.8992	0.0108	-0.0037	-0.0314	0.0926	0.0727	0.2284	-0.0301	1.8935
lnF	0	0.0009	0.0008	0.0034	-0.0049	-0.0072	0.9585	-0.0005	0.0633	-0.0012	-0.0007	-0.0036	0.0049	0.007	0.0416	0.0004	-0.0564
lnG	0	0.0017	-0.0034	-0.019	0.0328	0.0132	-0.0597	0.9557	4.4882	-0.0005	0.004	0.0206	-0.0329	-0.0145	0.0593	0.0446	-4.5146
lnH	0	-0.0001	0.0001	0.0003	-0.0006	-0.0001	0.0004	0.0011	0.8953	0.0001	-0.0001	-0.0003	0.0006	0.0001	-0.0004	-0.0012	0.1049

Table 6.6: Coefficient Matrices of VAR_f and $BVAR_f$;VAR_m and $BVAR_m$ showing that some of the coefficients in the BVAR are zero while in the VAR they are non-zero. A shortcoming of VAR models is some coefficients are estimated to be non-zero by chance.

distribution of the parameter, as estimated by the BVAR, 95% of time. The closer the observed value of the parameter for that year is to the BVAR estimate, the better the BVAR estimate. As the time from the last observed value used in the model, $T = 1995$, increases, the difference between the VAR estimate and the observed value increases. The plots show clearly that the BVAR projections outperform the VAR projections.

Figures 6.9 and 6.10 show the predictive distributions from the BVAR model for the male older age mortality parameters in the Heligman-Pollard model for 1996-2007 . The actual parameter values are also shown. The VAR estimates are also shown for comparison. The plots show how the BVAR model provides more accurate predictions and also quantify uncertainty in parameter estimates consistent with the actual mortality experience.

The relative performance of the two models is compared using the Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n^p - y_n)^2} \tag{6.2}$$

where for N observations, y_n is the observed value and y_n^p is the predicted value of observation n.

Table 6.7 shows the RMSE for the parameters for both males and females. The RMSE for the estimates from the B-VAR are smaller demonstrating the improved fit.

	VAR	B-VAR
G		
Female	2.39E-11	1.09E-11
Male	5.37E-11	3.25E-12
H		
Female	5.98E-05	2.96E-05
Male	8.00E-05	1.42E-05

Table 6.7: RMSE of predicted values 1996-2007 for parameters G and H using the VAR model (no parameter risk) and BVAR model (with parameter risk).

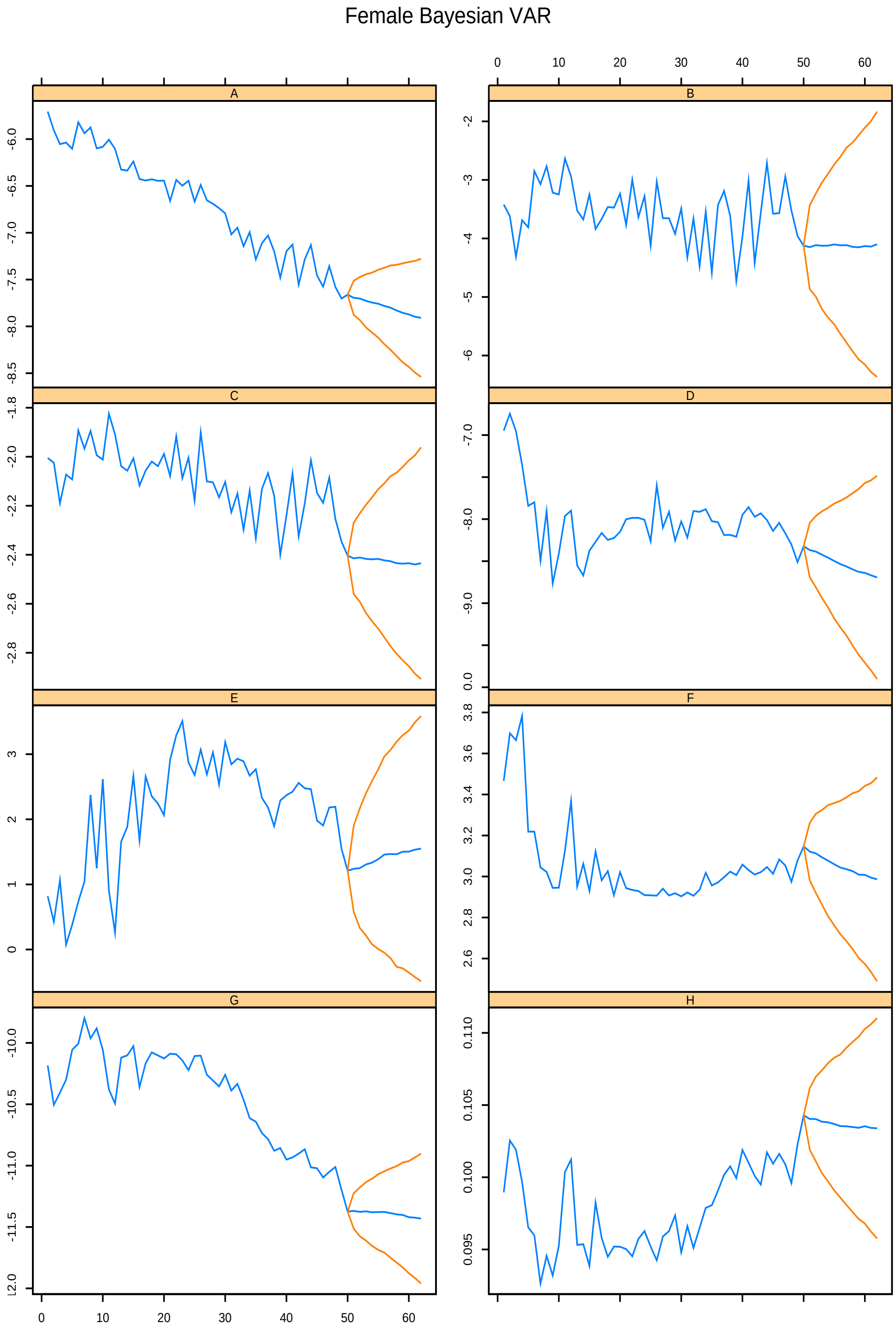


Figure 6.6: Fitted Values (1946-1995) and Projected Values (1996-2007) of $\ln \theta_{t,f}$ using BVAR(2) Females

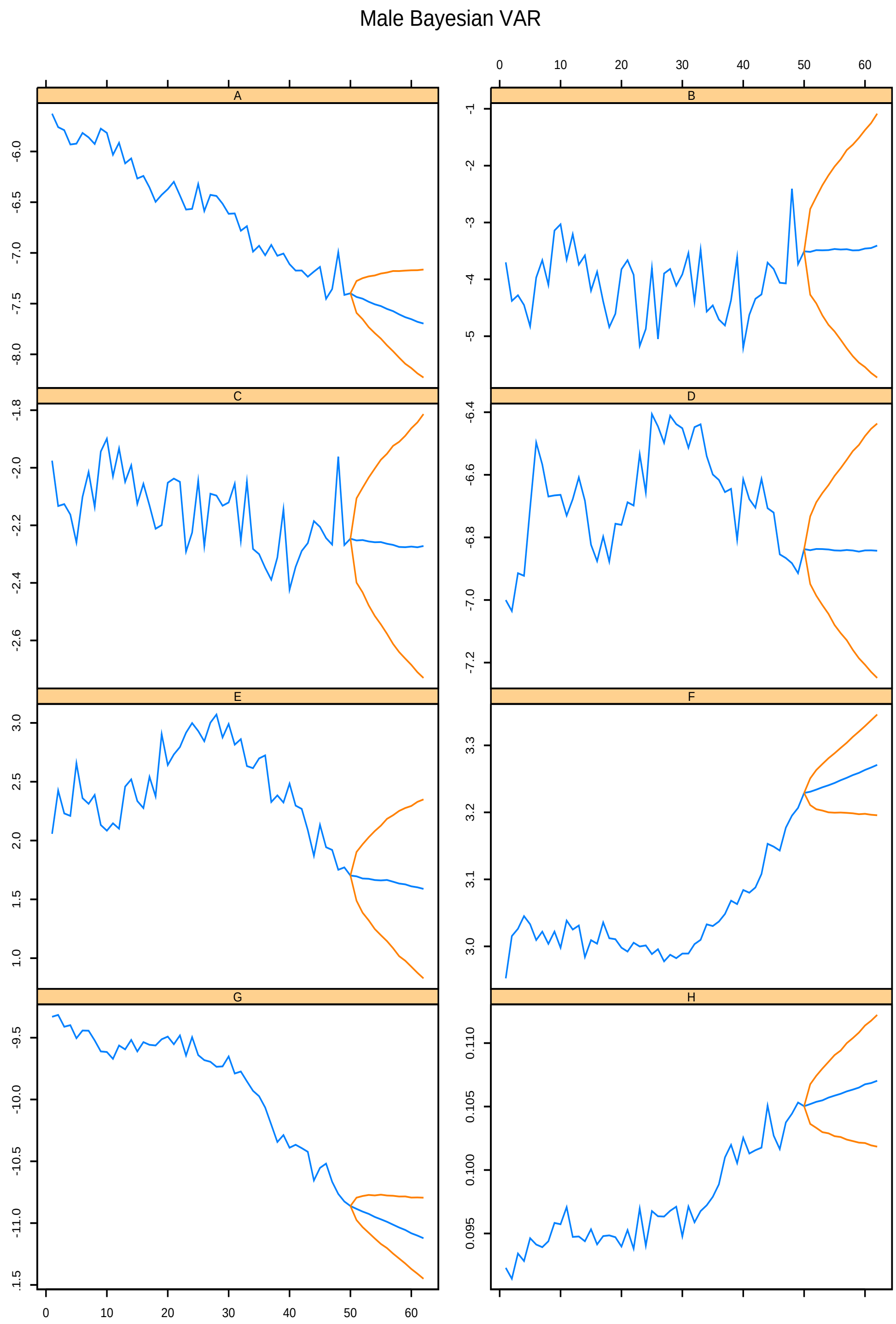
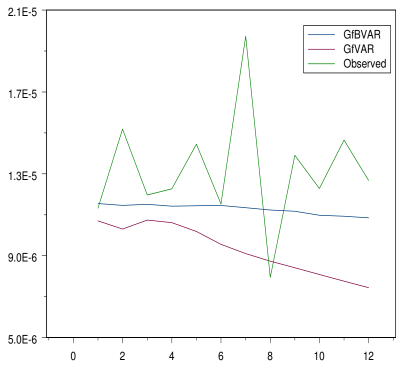
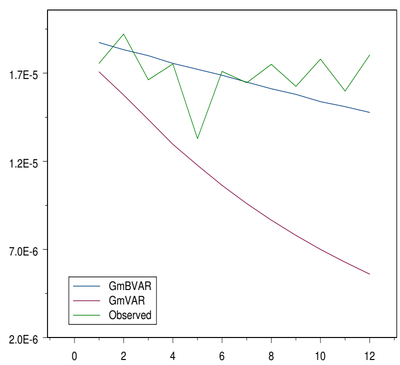


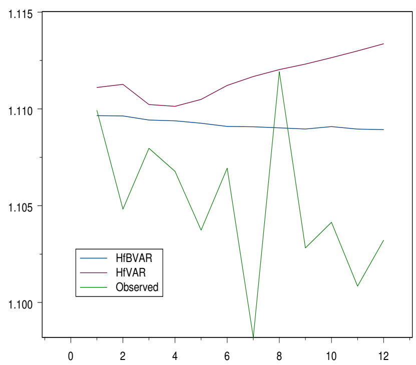
Figure 6.7: Fitted Values (1946-1995) and Projected Values (1996-2007) of $\ln \theta_{t,m}$ using BVAR(2) Males



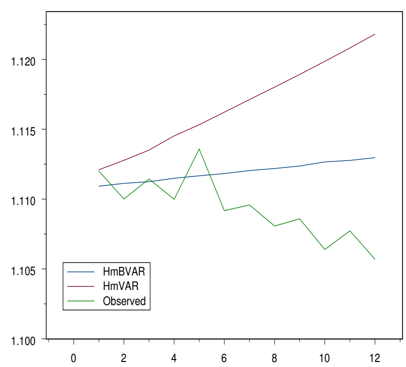
(a) $\ln(G)$ Females



(b) $\ln(G)$ Males



(c) $\ln(H)$ Females



(d) $\ln(H)$ Males

Figure 6.8: Estimated Old Age Parameters and the realised values from 1996-2007

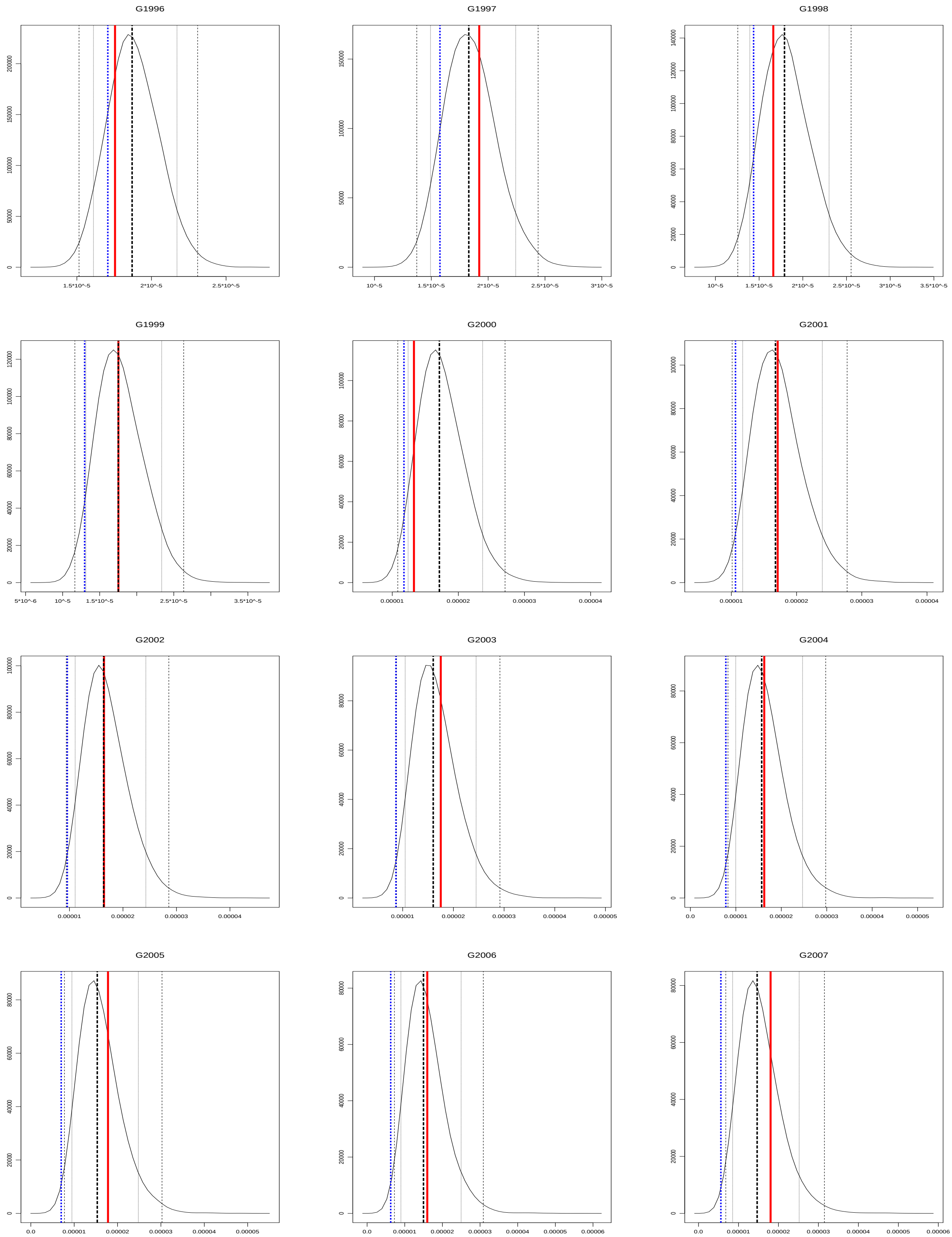


Figure 6.9: Parameter Risk in G (Males) for the predicted values for 1996-2007. Red (Thick Solid Line)=Observed, Black (Thick Broken Line)= BVAR Estimate, Blue (Thick Dot-Dashed Line)=VAR Estimate. The 95% and 99% confidence intervals are indicated by the thin dotted and thin broken line respectively.

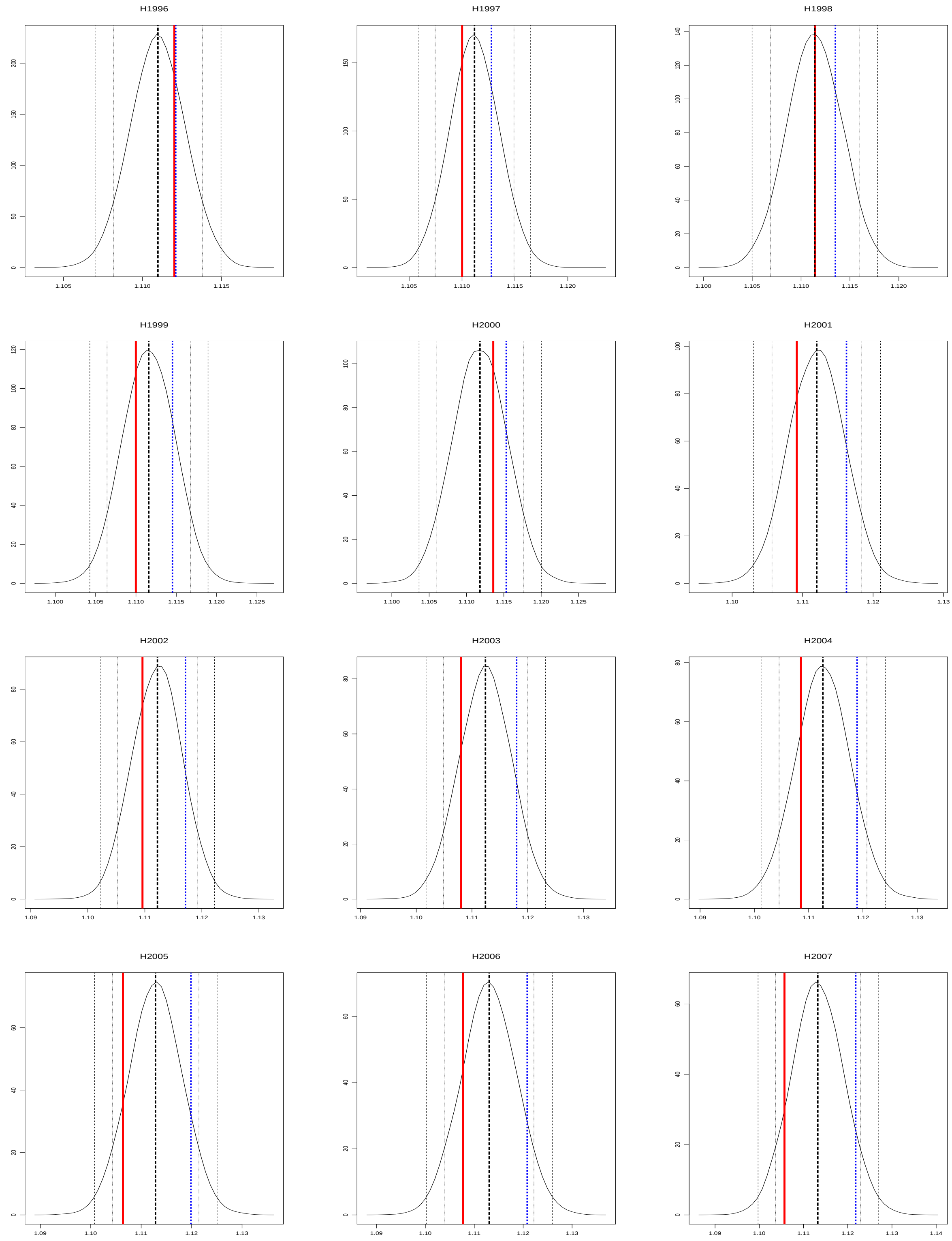


Figure 6.10: Parameter Risk in H (Males) for the predicted values for 1996-2007. Red (Thick Solid Line)=Observed, Black (Thick Broken Line)= BVAR Estimate, Blue (Thick Dot-Dashed Line)=VAR Estimate. The 95% and 99% confidence intervals are indicated by the thin dotted and thin broken line respectively.

6.4.1 Mortality Rate Estimates

Figures 6.11 and 6.12 show the confidence intervals for $\ln q_x$. The solid black line and the dotted black line are the BVAR estimate and 99% confidence intervals respectively while the blue broken line and the blue dotted line are the VAR estimate and 99% confidence intervals. The plots confirm the superior performance of the BVAR model. They also demonstrate the increased uncertainty in prediction intervals for mortality rates that arises from quantifying parameter risk in a Bayesian model framework.

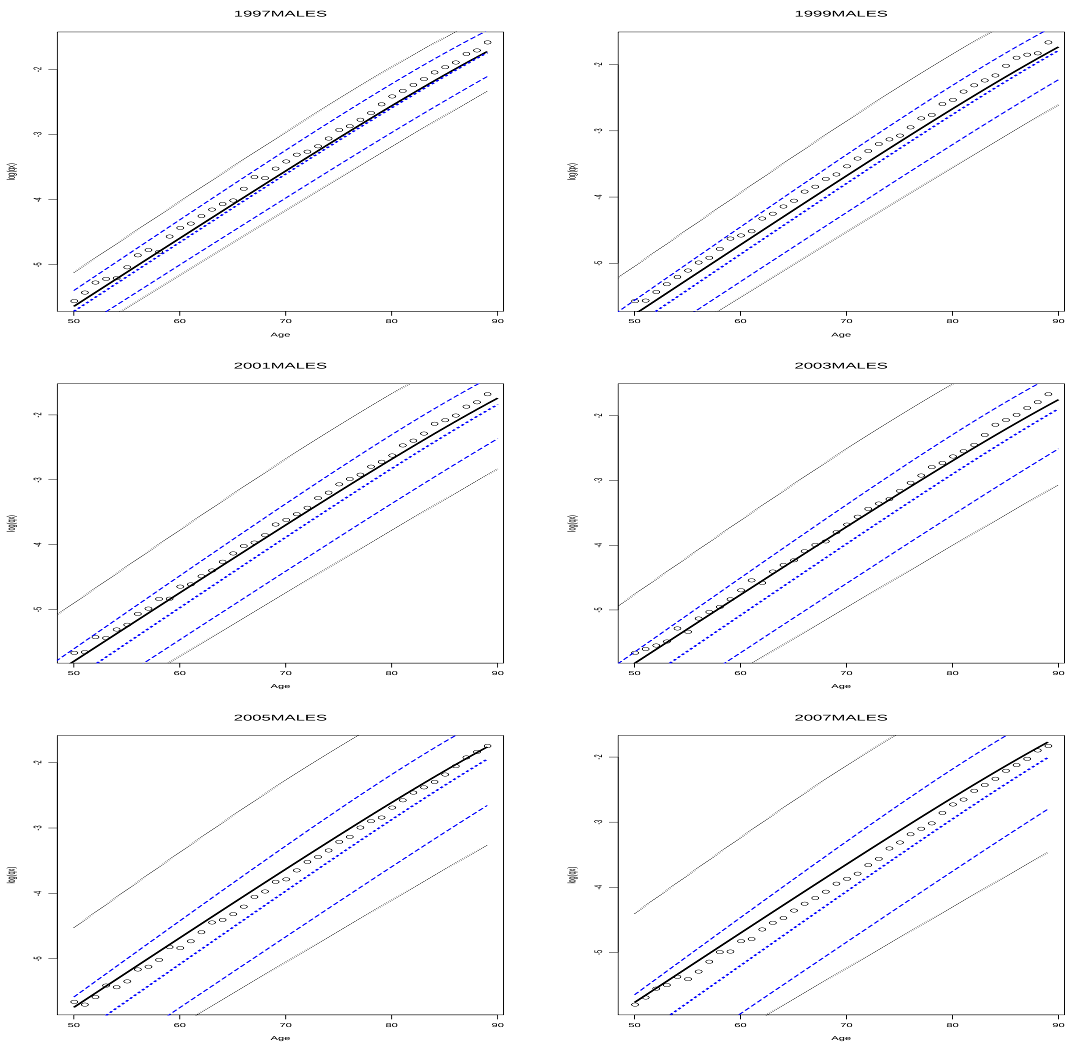


Figure 6.11: Old age Predictions of Male $\ln q_x$ by Year using HP-VAR and HP-BVAR and the observed $\ln q_x$ (small circles). The blue lines are for the HP-VAR model while the Black lines are for the HP-BVAR model. The HP-BVAR model has wider confidence intervals that reflect the parameter uncertainty.

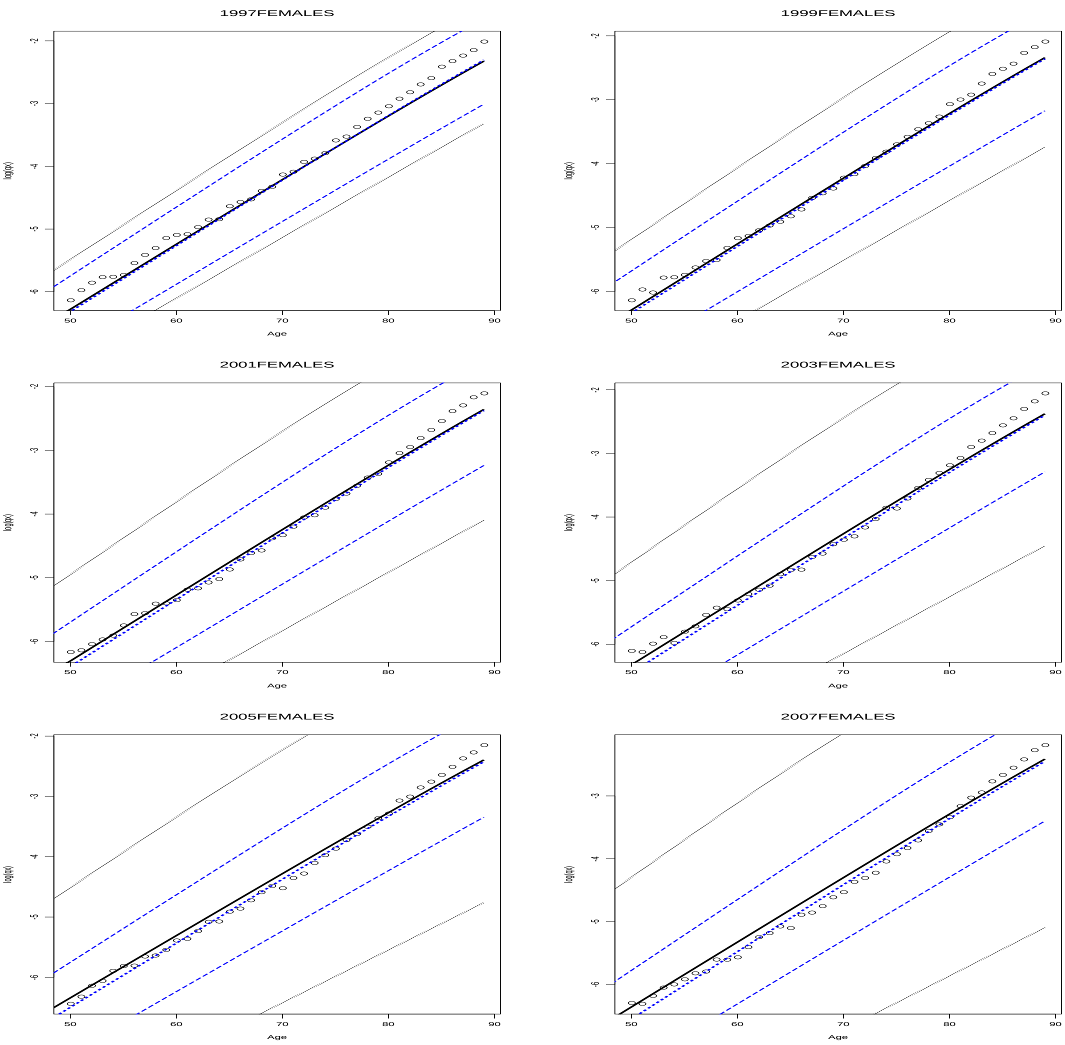


Figure 6.12: Old age Predictions of Female $\ln q_x$ by Year using HP-VAR and HP-BVAR and the observed $\ln q_x$ (small circles). The blue lines are for the HP-VAR model while the Black lines are for the HP-BVAR model. The HP-BVAR model has wider confidence intervals that reflect the parameter uncertainty.

6.5 Discussion

The results of this study have shown that it is possible to extend existing models to capture the effects of common trends on the mortality of a population using various econometric techniques. This study used VAR models and BVAR models to extend the static parametric Heligman-Pollard model and transform it into a dynamic parametric mortality model.

The parameters of the Heligman-Pollard model are correlated Hartmann (1987) and this often presents a problem for modelling of the evolution of the time series of the parameters. According to Booth and Tickle (2008) the ideal scenario would be to forecast the parameters simultaneously such that the interdependencies in the parameters are accounted for. The VAR and BVAR models fulfil this ideal scenario.

Several of the parameters of the Heligman-Pollard model are trending and unit root tests confirm that some of the Heligman-Pollard parameters have a unit root. Further analysis reveals that cointegration relationships also exist between these parameters. The presence of the unit roots and cointegration relationships must be accounted for when the evolution of the Heligman-Pollard parameters is modeled through time. A desirable feature of the BVAR model is its capacity to incorporate the unit roots and cointegration. This reduces the number of steps that are necessary to estimate the model.

A significant advantage of using the BVAR over the VAR is that the projected mortality forecasts in the HP-BVAR produce wider confidence intervals than the HP-VAR. This is important with regards to capturing parameter uncertainty. Bayesian models allow a quantification of parameter risk in a predictive distribution. For mortality modelling, the full age range of mortality rates is often captured by a parametric model at any given time in order to provide a parsimonious fit to the data. To forecast future mortality, the parameters are fitted by a time series model and used for projecting expected rates and the volatility of the rates. There has been limited analysis and quantification of parameter uncertainty in mortality models.

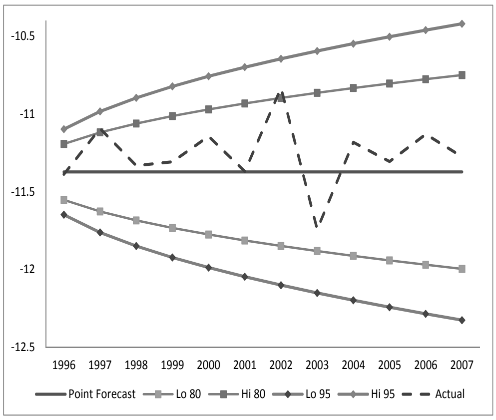
Another key benefit of using the HP-BVAR approach is that a trade-off between univariate methods and the unrestricted VAR is achieved. In McNown and Rogers (1989) a parameterized time series is modelled as described in Thompson et al. (1989) using the univariate time series techniques of Box and Jenkins (1976). A unit root in the time series of the parameters required differencing the time series to achieve stationarity (McNown and Rogers, 1989). The parameters for an ARIMA models were then extrapolated to obtain a series of Heligman-Pollard curves with time varying parameters. Because of the assumption of independence of the Heligman-Pollard parameters in the process given in Rogers (1986) and McNown and Rogers (1989) the forecasts are not necessarily accurate and will be inconsistent (Lee, 1992). Lee (1992) also notes the absence of confidence intervals which is attributed to problems that arise from the independence assumption.

Using a Bayesian Vector Autoregression model incorporates more uncertainty regarding projected mortality. Keeping in mind that the Bayesian-VAR is a compromise between VAR and univariate techniques, the projections using the HP-BVAR are based on an “adequate” number of parameters - as a result of the trade-off between over-parameterization and under-parameterization. In a VAR model the coefficients of the lagged variables in each equation in (2.39) are correlated as they are estimated using ordinary least squares - hence the over-parameterization. In contrast, in the univariate model only the lagged values of the variable of interest are used to model a variable and therefore it is assumed that there is no correlation between the coefficients of the lags of the variable of interest and the coefficients of the lags of the other parameters of the Heligman-Pollard model (e.g. to model $\ln G$, only the lags of $\ln G$ are used in the univariate model) - hence the under-parameterization.

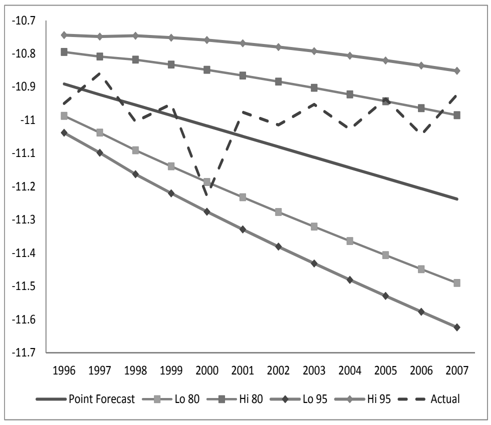
Consider figure 6.13 for example. For females, the univariate projections of $\ln G$ and $\ln H$ are constant. The implications of such projections will include high forecast error over medium to long term horizons because the projected mortality rates will remain constant as time from the date of projection increases. Recall the projected values of $\ln G$ and $\ln H$ using the HP-VAR AND HP-BVAR shown in figure 6.8. The HP-BVAR female projections have a slight downward trend while for the HP-VAR projections there is a downward trend for $\ln G$ and an upward trend for $\ln H$.

A reduced number of model parameters are expected to lead to a better out-of-sample fit. The performance of the HP-VAR and HP-BVAR models measured against the performance of other models commonly used such as the Lee-Carter (denoted as LC in table 6.8) model is presented in this section. The performance of these models is also compared to the univariate method based on McNown and Rogers (1989) (denoted as MR in table 6.8). In table 6.8, the models are arranged in order of increasing number of model parameters with the univariate MR on the left and the Lee-Carter (LC) on the right. It is important to consider the different number of parameters in each model because as the number of parameters in a model increases the fit of forecasted values is expected to be better. The mean absolute percentage error (MAPE)[1] at ages 0 - 89 for the four models for males and females

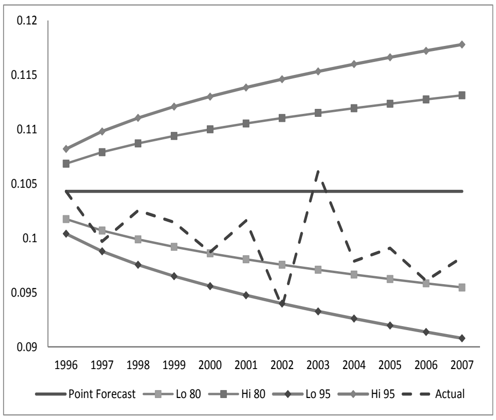
[1] $MAPE = \frac{1}{N}\sum_{n=1}^{N}\left|\frac{q_n^p - q_n}{q_n}\right|$ where q_n^p is the forecasted value and q_n is the actual observed



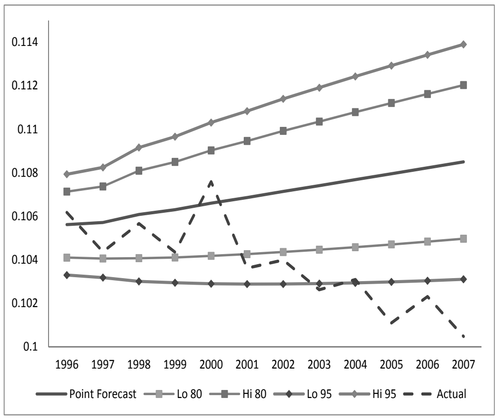
(a) ln(G) Females - ARIMA(0,1,0)



(b) ln(G) Males - ARIMA(1,1,0) with drift



(c) ln(H) Females - ARIMA(0,1,0)



(d) ln(H) Males - ARIMA(1,1,0) with drift

Figure 6.13: Old Age Parameters using Univariate Methods.

are as follows:

The models which capture the interdependencies between the parameters of the Heligman-Pollard model perform better than the univariate models for Australian males. The MR model was described to forecast USA mortality with reasonable accuracy (McNown and Rogers, 1989) and table 6.8 shows that the accuracy of the forecasts is improved by capturing the relationship between the parameters as suggested in McNown and Rogers (1989). For males, the BVAR model that captures the interdependencies in parameters of the Heligman-Pollard model with a smaller number of model parameters performs mostly better than the VAR model which is over parameterized. The models capturing interdependencies between the parameters of the Heligman-Pollard model outperform the univariate MR models. The Lee-Carter model gives better estimates than all the other models but it has

value.

	Males				Females			
	MR	BVAR	VAR	LC	MR	BVAR	VAR	LC
1996	27.03	**23.09**	23.23	9.15	24.93	23.39	**23.33**	9.77
1997	28.48	**22.79**	23.70	9.51	**22.36**	24.95	25.42	10.24
1998	25.79	**20.74**	22.57	9.04	**24.97**	29.18	30.29	10.50
1999	26.76	**20.99**	24.55	12.58	**24.91**	29.55	29.36	16.65
2000	26.52	**21.19**	25.37	13.34	**23.90**	28.64	28.99	14.02
2001	30.68	**22.89**	24.75	15.35	**23.28**	29.85	30.09	14.39
2002	33.19	**24.63**	23.97	14.59	**25.88**	29.10	28.83	16.67
2003	30.50	**22.86**	24.45	15.83	**27.31**	30.20	30.70	13.28
2004	32.87	**25.95**	27.55	18.99	**27.83**	30.21	29.10	16.71
2005	32.09	**25.74**	25.81	21.02	**29.55**	31.57	30.91	15.62
2006	34.40	28.62	**28.45**	24.19	36.47	34.48	**32.92**	19.42
2007	42.91	36.02	**30.06**	25.95	32.87	31.83	**31.76**	16.19

Table 6.8: MAPE of predicted mortality for males (left) and females (right) given different mortality models for ages 0-89. The models are arranged in order of increasing number of parameters with the model with the largest number of parameters on the right (LC). The smallest MAPE for a given year for the Heligman-Pollard based models is indicated in **bold** numbers.

the largest number of parameters.

The models for Australian females show no gain in accuracy improvements when the interdependencies are captured. The univariate MR models outperform the BVAR and VAR models. The correlations between the parameters of the female Heligman-Pollard parameters quite small. Therefore, capturing the interdependencies in the Heligman-Pollard parameters does not improve the forecasts.

The approach presented in Sherris and Gaille (2010a) is similar to the HP-VAR approach in this chapter but does not use the advanced Bayesian VAR to address the over-parameterization in the VAR model. A key difference is that Sherris and Gaille (2010a) analyses cause-specific mortality rates and for some causes some of the Heligman-Pollard parameters are not subject to significant variation through time. Therefore, a smaller number of parameters are modelled. An improvement over the univariate method in McNown and Rogers (1992) is realised as the econometric approach captures common stochastic trends across the age structure of mortality just like the HP-VAR and the HP-BVAR models in this thesis.

Brandt and Freeman (2006) compares the eigenvector decomposition during the construction of Bayesian error bands in Sims and Zha (1999) to a dynamic factor analysis and concluded that they are similar because both methods account for the main sources of variation in the responses over time. It is possible that a dynamic factor model could outperform a Bayesian VAR model (Gupta and Kabundi, 2011) when directly applied to modelling mortality trends but the additional dimension reduction by use of Heligman-Pollard model complicates matters. A dynamic factor mortality model has been recently developed in O'Hare and French (2011).

6.6 Recommendation for Further Research

The models described in this chapter capture the interdependencies in the parameters of the Heligman-Pollard model for Australian males and Females. A key assumption in this chapter is that the age distribution is accurately described by and will remain in the shape described by the Heligman-Pollard mortality law. The Heligman-Pollard mortality law has several shortcomings. The accident hump is disappearing and is described as more of a bulge than a hump (Pollard, 1996). For females, other shortcomings include the definition of the accident/maternity hump. Improved medical practices are significantly reducing maternal deaths (Hannerz, 2001b). Rather than having a single "accident" hump it possible that mortality will have several small humps or bulges that assume different shapes and locations (Hannerz, 2001b). A mortality law that considers a population where females have low infant mortality and high life expectancy as in Hannerz (2001a,c) should be considered as a possible alternative to the Heligman-Pollard model. There are several other possibilities for alternative models (e.g. Carrière, 1992; Rogers, 1983). The evolution of the parameters estimated can be modelled using econometric techniques such as those in this thesis to transform it into dynamic mortality laws.

Further investigations into the most suitable set of hyperparameters for making the HP-BVAR model generalizable for use in a variety of populations is another area for potential research.

In addition to obtaining more realistic forecasts from VAR and B-VAR models than from univariate models, conditions can be imposed on the forecasts of VAR models and Bayesian VAR models to give conditional forecasts under different scenarios. There is potential for further research in this area as it will be useful in analysis of longevity risk by considering the potential effects of changes in different aspects of mortality.

6.7 Conclusion

In this chapter, the method developed in McNown and Rogers (1989) is extended to capture the effects of common trends in the Australian population mortality. The HP-BVAR and HP-VAR models considered the correlation between the parameters of the parametric model through time capturing trends from young ages and taking into account the richer age structure of mortality improvement from young ages to middle and then into older ages by utilizing the dependencies between ages.

The main result in this chapter is a parsimonious model that improves the accuracy of mortality forecasts and captures longevity risk (including parameter uncertainty). The developed model performs better for males than females. The Heligman-Pollard model is transformed from a static to a dynamic parametric mortality model. Various econometric models including VAR models, VECM models (and cointegration tests) and Bayesian VAR are used to model and project the uncertainty in future mortality.

The uncertainty from the model parameters is captured in the Bayesian VAR. Parameter uncertainty is a significant risk due to the limited amount of data available for use in model estimation. The uncertainty in longevity risk that is contributed by uncertainty from parameter estimation is ignored in several other models. The uncertainty from the parameters estimated and the models chosen is estimable because probability distributions have been used to explicitly incorporate data and uncertainties in parameter estimation and model choice in a coherent and transparent way. This has led to realistic probabilistic projections.

7

Measures of Longevity Shocks for Longevity Stress Margins in Risk-Based Capital

Introduction

In this chapter a longevity stress margin is calibrated using an internal model based on the Heligman-Pollard and Bayesian Vector Autoregression (HP-BVAR) model developed in 6. A portfolio of life annuities is used to estimate the longevity stress margin as the probability of a one in two hundred year event occurring in a one year horizon and is used to determine the corresponding constant permanent decline in mortality rates. The longevity stress margin for the insurance risk in life annuities and deferred life annuities is calibrated. The longevity risk stress margin is defined in APRA's technical documents APRA (2010b, 2011) and the calculation of longevity risk margin is simplified by assuming a permanent 25% decrease in mortality. This simplification is a trade-off between complexity and detail and is similar to the original simplification assumption in Solvency II which has since been revised to assuming a permanent 20% decrease in mortality. In this chapter the assumption that a permanent decline in mortality of 25% is checked against the 99.5% probability of sufficiency when the underlying mortality model is the HP-BVAR. When Solvency II's simplification for the computation of capital to meet adverse experiences in longevity (longevity stress), there was a lot of debate regarding the assumption. In Australia, however, there has been no discussion on APRA's assumptions in its simplification on the longevity stress to date. This is evidenced by the lack of detailed discussion on the longevity stress margin in APRA (2011). This study is intended to stimulate further discussion on the suitability of the APRA-specified simplification for the calculation of the longevity stress margin.

7.1 Background

The Australian life annuity market is small and the demand for life annuities is low and falling (Sherris and Evans, 2010). There exists a need for provision of more choices for financing retirement income through a variety of products that provide living benefits such as different types of life annuities. To directly quote the Australia's Future Tax System Henry (2009) Chapter A section A.2-3 *"As people live longer, they will require more options to manage their assets over a longer period"*. It is not mandatory for retirees to invest their benefits as a lump sum in any particular product at the moment and retirees do not generally purchase life annuities or other longevity insurance products on a voluntary basis. Two key factors that encourage this behaviour include expensive life annuities and the availability of the Age Pension later in life. In addition, the taxation of income streams from life annuities does not motivate retirees to purchase them.

A key consideration that annuity providers must account for is the cost of holding capital. This cost is important because the price of the product must cover the cost of capital. Life annuities in Australia are already expensive and therefore any regulations that will affect the amount of capital annuity providers will be required to hold (and by extension the cost of the life annuities) must be considered carefully. The purpose of this chapter is to comprehensively consider the regulations for capital that is directly related to longevity risk - the longevity stress margin (LSM). The APRA-specified method for calculating the LSM is analysed with reference to its magnitude and structure.

Since the most efficient solution to the life annuity market challenge in Australia would be the compulsory annuitisation of retirement lump sums (Sherris and Evans, 2010) this study is based on population mortality. Compulsory annuitisation would eliminate the problem of adverse selection[1] which directly affects the cost of capital.

The following significant assumptions are made in this chapter:

1. There is no basis risk. This chapter involves male population mortality rates. This assumption is primarily due to data limitation. However, some studies already exist that compare the mortality of members of industry pension schemes to that of members of the general population (Knox and Nelson, 2006; Stevenson and Wilson, 2008) and their results can be extrapolated to make further conclusions if needed.

[1]Adverse selection in Life Annuities occurs when very healthy retirees who are likely to live longer than their peers buy life annuities while their peers do not buy them.

2. There is no interest rate risk. Börger (2010) shows that interest rates do not have a significant impact on the results of an analysis on longevity stress margins.

This study is an application of the model developed in chapter 6.

7.2 Simulation Based Analysis of Longevity Insurance Products

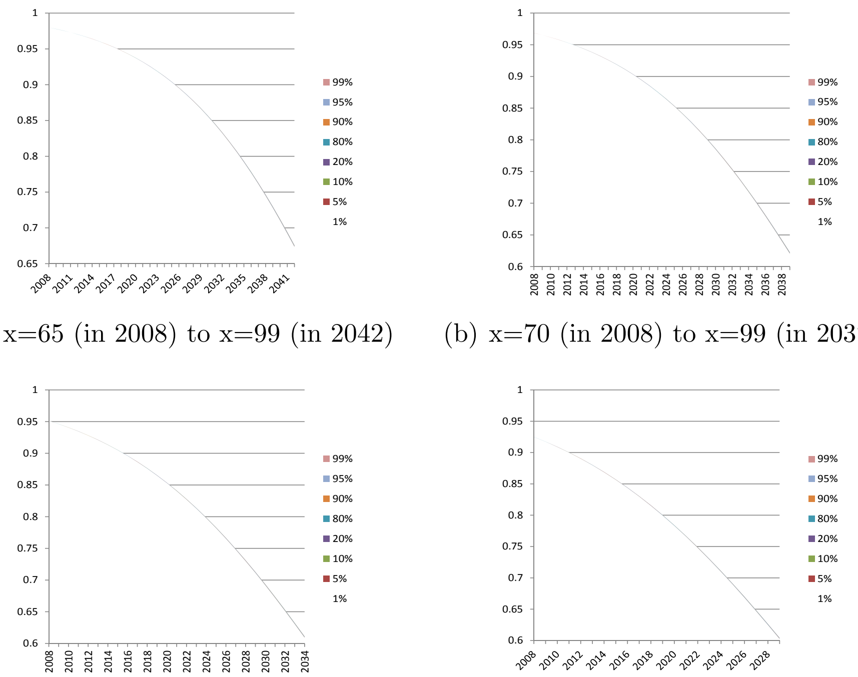
A simulation based analysis is performed in this chapter. This is due to two main reasons. The lack of data and the non-linearity of the variables of interest. The former is due to the practically non-existent Australian life annuity market as described in the previous section. The latter is because analytical or closed form representations of longevity models are not possible when multiple sources of uncertainty need to be combined and when the variables of interest, such as expected annuity values, are non-linear functions of the parameters of the model. Consequently simulation techniques must be used (Renshaw and Haberman, 2008). The age-period Heligman-Pollard Two-Stage mortality models have uncertainty from the estimated parameters of the Heligman-Pollard parameters and additionally there is parameter risk in the HP-BVAR.

Here, HP-VAR and HP-BVAR models are estimated using Australian male mortality data from 1946-2007. 10,000 paths of 40 year predictions are simulated from these two models to obtain sample paths of the parameters of the Heligman-Pollard parameters. Monte Carlo simulation is used in the HP-VAR model while a Gibbs Sampler is used for HP-BVAR model. The distribution of $q_{x,t}$ is obtained for each $\tau = 1, \ldots, 40$ with $t = 2007 + \tau$ by simulating the underlying parameters θ_t. Then, the confidence and prediction intervals for life expectancy are computed. The actuarially fair price of a life annuity (a_x) and a deferred life annuity (${}_{t|}a_x$) with a fixed interest rate are calculated as well as the changing cost of longevity as reflected by the price of an annuity. Longevity risk is measured in terms on the percentiles of the distributions is used to estimate the corresponding capital requirements.

7.3 Uncertainty in Future Mortality

In this section 10,000 paths of θ_t are simulated and substituted in equation (2.16) to obtain 10,000 simulated paths of $q_{x,t}$ over a chosen horizon. The survival rates $S_{x,t}$ are calculated from $q_{x,t}$ using $S_{x,t+1} = (1 - q_{x,t})S_{x,t}$.

Blake et al. (2008) estimated probabilities of survival using the Cairns-Blake-Dowd model (Cairns et al., 2006b) using two formulations (one with parameter risk and one without). The uncertainty in future survivorship was illustrated using fan charts which they defined as “charts of projected probability densities over each year in a specified forecast period and show the likely confidence interval to which a dynamic quantity may belong in a particular future year”. The confidence intervals of survival paths from $q_{x,t}$ based on the simulated sample paths of the parameters of the Heligman-Pollard parameters are shown using fan charts in this section, in a manner similar to Blake et al. (2008).



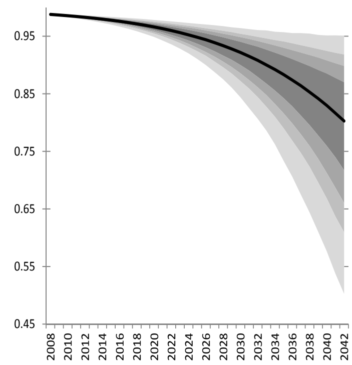
(a) x=65 (in 2008) to x=99 (in 2042) (b) x=70 (in 2008) to x=99 (in 2037)

(c) x=75 (in 2008) to x=99 (in 2032) (d) x=80 (in 2008) to x=99 (in 2027)

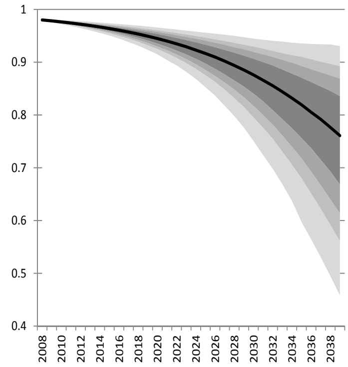
Figure 7.1: Fan charts of p_x without parameter uncertainty.

Allowing for parameter risk leads to wider confidence intervals and therefore a wider width of the fan chart bounds in Cairns et al. (2006b). In chapter 6 the difference in the width of the confidence intervals of the HP-BVAR and HP-VAR forecasts was presented and in a similar manner the confidence bands for the HP-BVAR are wider than those of the HP-VAR model (see figures 7.1 and 7.2).

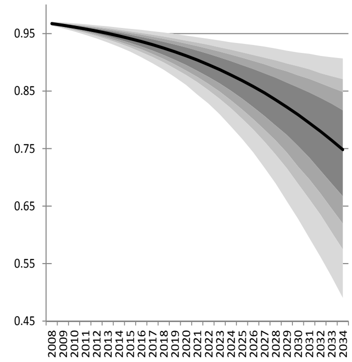
It is noteworthy that as age increases the confidence bands become wider. This reflects the high uncertainty in mortality at higher ages. In addition, the confidence bands become wider after a shorter time as age increases. For instance, a comparison of 7.2(a) and 7.2(b-d) is that as the starting age increases from 65 to 80 the point at which the confidence bands grow wider moves closer to the starting year. This further verifies that uncertainty in mortality increases with age.



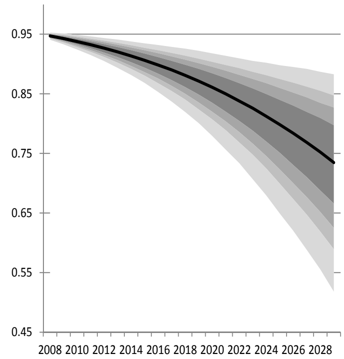
(a) x=65 (in 2008) to x=99 (in 2042)



(b) x=70 (in 2008) to x=99 (in 2037)



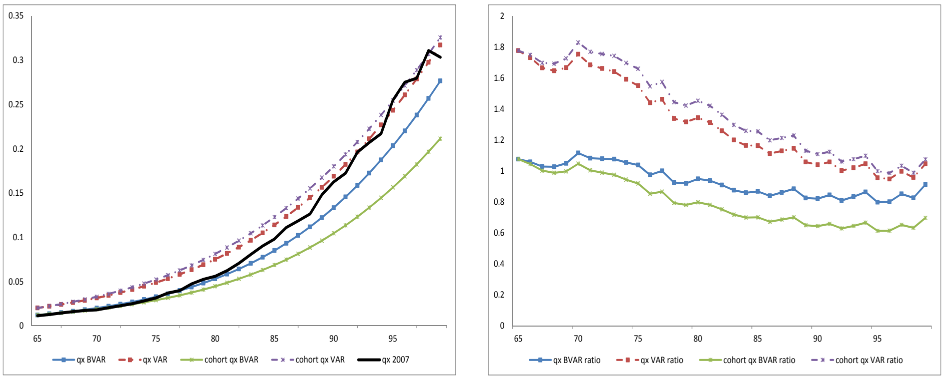
(c) x=75 (in 2008) to x=99 (in 2032)



(d) x=80 (in 2008) to x=99 (in 2027)

Figure 7.2: Fan charts of p_x with parameter uncertainty. As age increases the bands become wider after a shorter time reflecting the high uncertainty in mortality at old ages.

One year forecasts of age-specific mortality rates and the corresponding permanent percentage reduction for the best estimates are shown in figure 7.3 and the 99.5 percentile are shown in figure 7.4. The projections are extracted in two ways for both the HP-VAR and HP-BVAR models. In the first method age-period projections (the columns of the projected matrix) are extracted. The second method extracts cohort projections (the diagonals of the projected matrix). Both methods are used because the HP-VAR and HP-BVAR models do not explicitly account for the cohort effect and a life table is required for valuation of longevity insurance products.

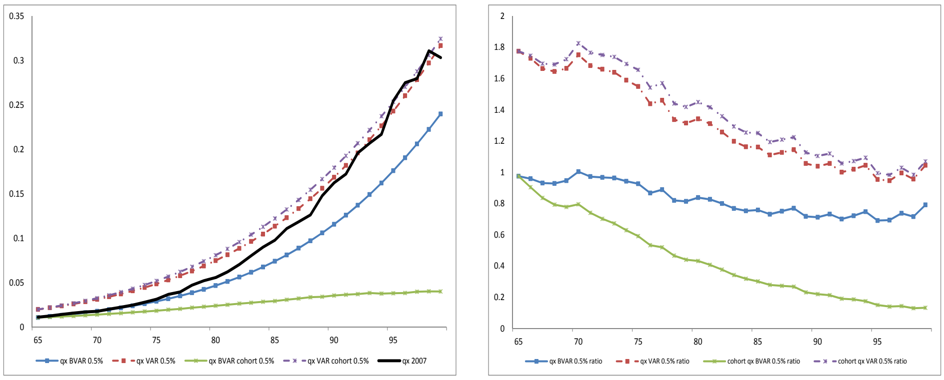


(a) Different Best Estimates $\hat{q}_x^{2008}$ compared to the observed q_x^{2007}

(b) Ratio $\hat{q}_x^{2008}/q_x^{2007}$

Figure 7.3: Probability of Death projected under different models. The solid lines are the HP-BVAR model; the dot-dashed lines are the HP-VAR model. The Square markers are the age-period projections; the cross markers are the cohort projections. The HP-BVAR reflects that the mortality from age 65 to 75 in 2007 and in the 2008 projections are similar under both the age-period and the cohort projections. After age 75 mortality improves to a greater extent in the HP-BVAR cohort projection (green cross marker solid line) than in the HP-BVAR age-period (blue square marker solid line). The HP-VAR projections overestimate mortality up to age 90 and at the oldest ages mortality in 2008 is similar to that in 2007. The HP-VAR is overestimates mortality.

The ratio of the best estimates of the projected probability of death in 2008 to the observed probability of death in 2007 shows that the decrease in mortality does not occur at a constant rate. Different ages have different rates of mortality decline and at some ages the probability of death increases as shown in figure 7.3. On average, from table 7.1, the HP-VAR suggests that mortality will increase and this is very unlikely based on past experience. HP-BVAR model suggests that mortality will decrease, with greater decreases (56%) under the HP-BVAR cohort estimates. The HP-BVAR age-period projections suggest that an (unweighted) average mortality decline of 17 percent over all ages 65 to 99 is a one in 200 year event. Further, as an unweighted average, this figure does not capture the fact that as age increases the population size decreases. Therefore, the weighted average mortality decline is expected to be even lower. In the valuation of liabilities for life annuities the weighting is incorporated as part of the price of the annuity and a more realistic average percentage decline will be estimated. Observing that the unweighted average constant decline is significantly less than 25 percent decline proposed to represent the 1 in 200 year event under APRA's proposed framework suggests that the percentage decline that will be required to make the 99.5% VaR equivalent to the value of liabilities due to a constant shock will definitely be less than 25%.



(a) 0.5% Probability of Death $\hat{q}_{x,0.5\%}^{2008}$ compared to the observed q_x^{2007} under different models

(b) Ratio $\hat{q}_{x0.5\%}^{2008}/q_x^{2007}$

Figure 7.4: 99.5% Quantile of Probability of Death projected under different models. The confidence bands of the HP-VAR model are not wide so the 5% quantiles are not significantly different from the best estimate. The confidence bands of the HP-BVAR model are wider. The amount of uncertainty in the HP-BVAR cohort projections increases by both time and age. This leads to a significant decrease in $\hat{q}_{x0.5\%}^{2008}$ HP-BVAR cohort projections (green cross marker solid line.)

	HP-BVAR	HP-VAR	HP-BVAR Cohort	HP-VAR Cohort
Best Estimate % Δq_x	-0.0670	0.3102	-0.2037	0.3834
0.5% Quantile % Δq_x	-0.1757	0.3089	-0.5635	0.3801

Table 7.1: (Unweighted) Average percentage change in q_x from 2007 to 2008

A life table based on the HP-BVAR age period estimate is a good approximation to the ABS life table estimates. The life expectancy at age 65 in 2007 is 18.62 while the estimated life expectancy at age 65 in 2008 when estimated using the age period estimates from HP-VAR is 16.05 and HP-BVAR is 18.99. However, the life expectancy estimated by following the cohorts when estimated from HP-VAR is 15.60 and from HP-BVAR is 20.17. For comparison, the male life expectancy at age 65 based on the Australia Bureau of Statistics (ABS) life table 2005-2007 is 18.5 years and based on the ABS life table 2007-2009 is 18.7 years. Therefore, HP-BVAR age period estimate is the best option to consider for the best estimate life table because it is only 0.29 years (3 months) more than the ABS 2007-2009 estimate.

The model that is closest to reality is the HP-BVAR age period model. The HP-BVAR model integrates the correlation of the Heligman-Pollard parameters as they

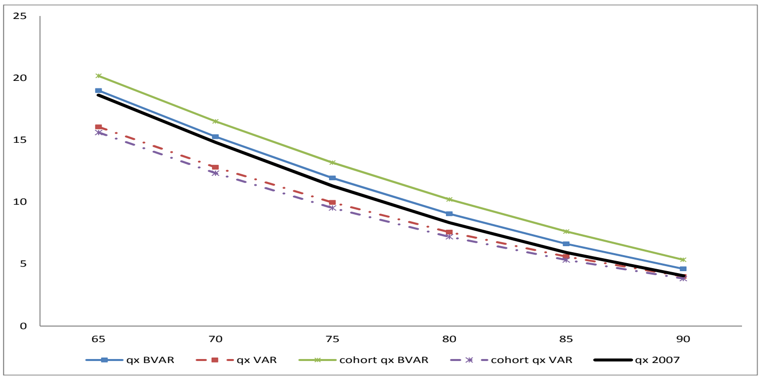


Figure 7.5: Life Expectancy in 2008 compared to 2007.

evolve. Since each parameter is describes mortality in one of three phases of the lifespan the HP-BVAR model captures changes in a cohort's mortality. Consequently, a life table extracted using the age-period projections is sufficient.

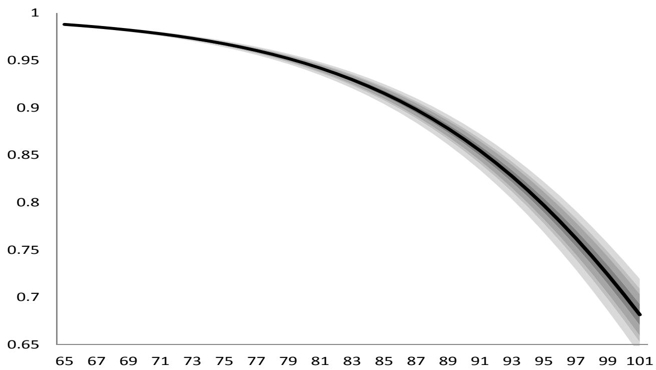


Figure 7.6: One Year Survival Confidence Intervals

As age increases the uncertainty in survival increases. This is shown in in figure 7.6 where one year survival confidence intervals get wider as age increases.

7.4 Calibration of the Longevity Risk Stress Margin

Different magnitudes of the longevity shock, γ, affect annuity values at different ages to different extents. For an immediate life annuity The q-duration defined in equation (4.23) for different ages is shown in figure 7.7. The ages closer to age 65 have a steeper curve than the older ages and as a result it is concluded that the sensitivity of the value of liabilities for a life annuity to a 1% change in mortality decreases with increasing age.

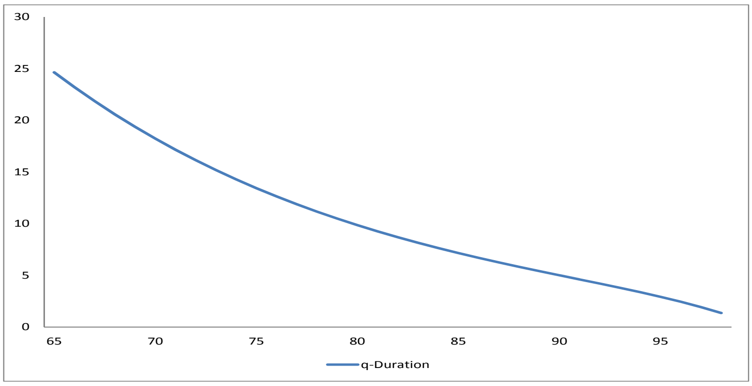
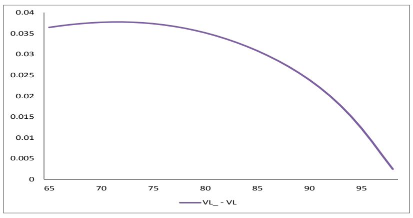
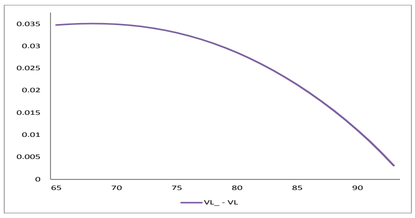


Figure 7.7: q-Duration of a life annuity at different ages. The steepness of the curve decreases as age increases. The value of liabilities for an annuitant aged 65 is more sensitive to a 1% change in mortality than the value of liabilities for an annuitant aged 85.

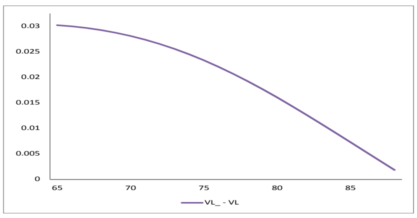
The effect of a 1% decline in mortality rates varies for different annuity products. The length of the deferral period is a factor that affects the change in the value of liabilities when a 1% mortality decline is experienced. The change in the value of liabilities for an immediate life annuities and deferred life annuities due to a 1% decline in mortality at all ages is shown in figure 7.8.



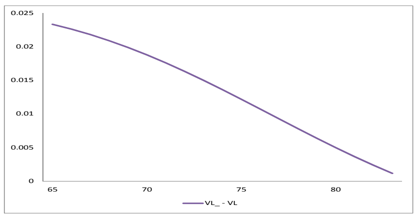
(a) a_x



(b) ${}_{5|}a_x$



(c) ${}_{10|}a_x$



(d) ${}_{15|}a_x$

Figure 7.8: Change in the value of liabilities when mortality declines by 1% for different life annuity products. The value of liabilities of an immediate annuity initially increases as age increases and then decreases. The value of the change is smaller for longer deferment periods.

All the figures represent a product where \$1 paid annually to an annuitant and the change in the value of liabilities due to a 1% decline in mortality is shown. In general, as age increases the additional value of liabilities for a 1% decline decreases. However, for the immediate annuity there is an increase in $VL_{-} - VL$ then a decrease as age increases as shown in figure 7.8(a). This is due to the fact that since the mortality rates become larger as age increases the shocks will also become larger. For ages 65 to 75 a shock with a constant structure will possibly lead to problems in determining the longevity stress margin. After age 75 the value of the liabilities decreases. A further observation is that annuity products with different deferral periods are affected to a different extent by a 1% decline in mortality.

7.5 Financial Implications

The longer annuitants survive, the more the amount of annuity payments made and consequently the higher the present value of future liabilities. The survival patterns of annuitants can be described by internal models. From the results in section 7.4 the HP-BVAR model quantifies longevity risk adequately. Therefore in this section the HP-BVAR is the "internal model" used to measure the financial implications of the longevity stress margin.

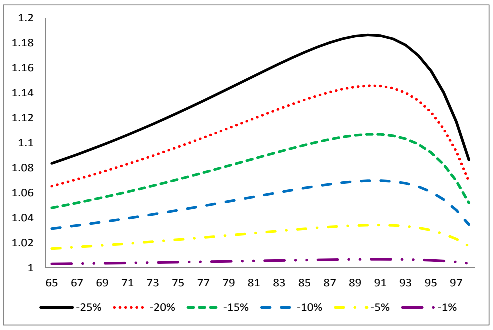
This section is an analysis whether $LSM^{VaR} = LSM^{SHOCK}$ at γ=-25%, where γ is the constant permanent percentage change in best estimate of future mortality.

γ	65	70	75	80	85	90	95
-0.25	13.15	11.38	9.54	7.70	5.93	4.23	2.39
-0.20	12.93	11.14	9.30	7.48	5.74	4.09	2.32
-0.15	12.72	10.92	9.08	7.27	5.55	3.95	2.25
-0.10	12.52	10.71	8.87	7.07	5.38	3.82	2.19
-0.05	12.33	10.51	8.67	6.88	5.22	3.69	2.12
BE	12.14	10.32	8.48	6.70	5.06	3.57	2.06
99.5% VaR	12.57	10.78	8.96	7.17	5.49	3.90	2.23

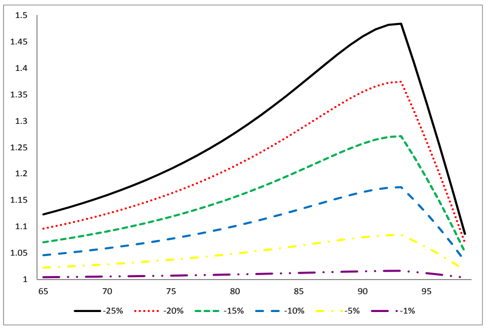
Table 7.2: Value of Liabilities at different ages for a_x that pays \$1 per annum.

Consider a life annuity that pays \$1 at the end of the year that is sold to Australian Males at the beginning of 2008. Assuming a fixed interest rate of 4 percent, the value of liabilities for an immediate annuity subject to a shock of magnitude γ are presented in table 7.2. A comparison of the value of liabilities at the end of the year when the annuity is subjected to 5 levels of constant mortality shocks, the 99.5 percentile of the mortality rates and the best estimate of mortality (mortality shock with $\gamma = 0$) suggests that the constant mortality shock should be a decline that lies between 10% and 15%. This is consistent for all ages from 65 to 95.

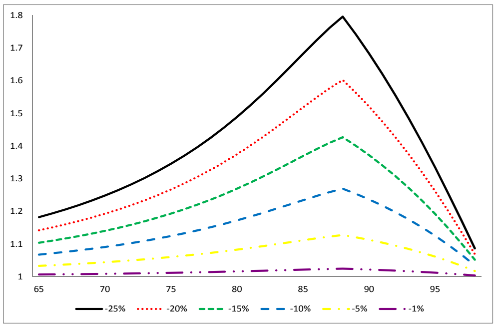
A visualization of the ratio of the value of liabilities at varying stress levels, VL^{γ}, to the best estimate of liabilities, VL^{BE}, is $\frac{VL^{\gamma}}{VL^{BE}}$, $\gamma = \{-25, -20, -15, -10, -5\}$ is given in figure 7.9. From table 7.2 the 99.5% VaR lies between 10% and 15% for the immediate life annuity and figure 7.9 verifies this for deferred life annuities as well.



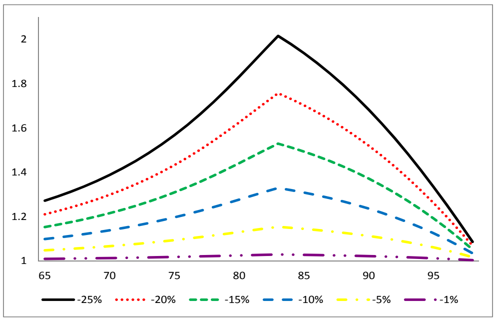
(a) a_x



(b) ${}_{5|}a_x$



(c) ${}_{10|}a_x$

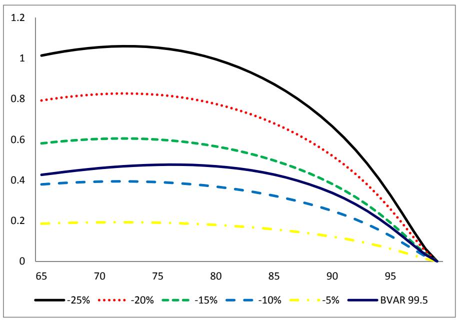


(d) ${}_{15|}a_x$

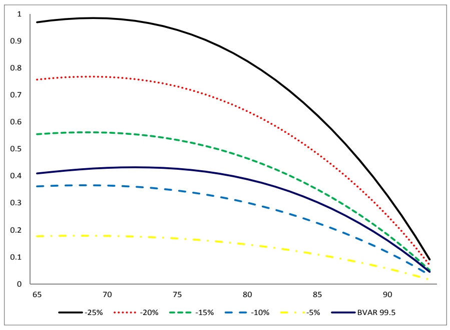
Figure 7.9: $\frac{VL^{\gamma}}{VL^{BE}}$ life annuity at different ages. The curves from bottom to top represent the effect of increasing the magnitude of the shock in 5% increments from 5% to 25%. The top curve (black solid line) is subject to $\gamma = -25\%$ shows a slow increase in the ratio with increases in age up to age 90 and a more rapid decline for age 90+. In comparison, the bottom curve (purple dot dot dash line) at γ = -1% is flat. As the magnitude of the shock increases, $\frac{VL^{\gamma}}{VL^{BE}}$ increases.

Figure 7.10 is a visualization of the longevity stress margins for four different products - an immediate life annuity, 5 year, 10 year and 15 year deferred life annuities. The 99.5% VaR lies between 10 and 15% for all ages. For the very elderly (age 95+) there is no difference between the LSMs with shocks of different magnitudes.

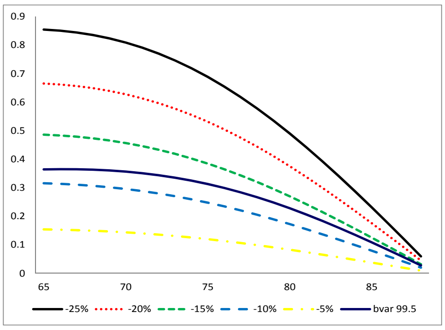
The constant permanent decline in mortality rates at all ages that will result in the same amount of capital as that needed to be sufficient with 99.5% probability is significantly less than 25%. It lies between 10% and 15%. From table 7.3 the assumption that the 1 in 200 year shock is equivalent to a 25% constant mortality decline in not accurate. It leads to a capital requirement of at least double the 1 in 200 year event's requirement. The conclusion is that LSM^{VaR} is not equal to



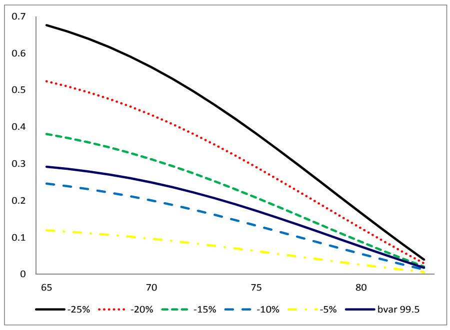
(a) a_x



(b) ${}_{5|}a_x$



(c) ${}_{10|}a_x$



(d) ${}_{15|}a_x$

Figure 7.10: The LSM given shocks of different magnitudes. The 25% shock is the black solid line while the 99.5% percentile is the purple (grey in black and white) solid line. LSM^{VaR} lies between LSM^{-10} and LSM^{-15}.

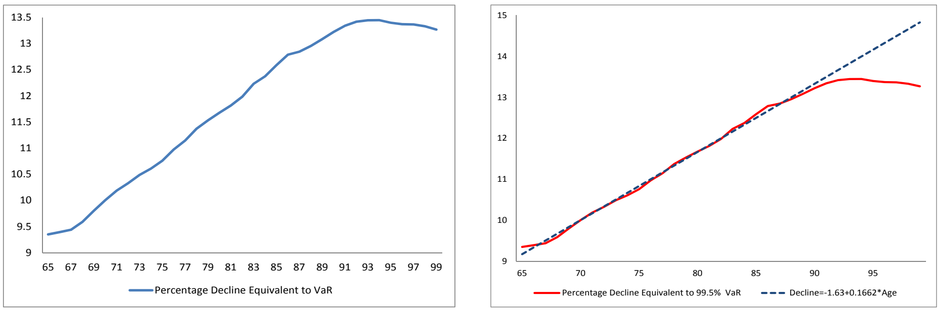
LSM^{SHOCK} at γ=-25%. The results recommend that it is necessary to review the magnitude of the constant value in the assumption of a constant mortality decline.

x	a_x			${}_{5\|}a_x$			${}_{10\|}a_x$			${}_{15\|}a_x$		
	-25	VaR	$\frac{LSM^{-25}}{LSM^{VaR}}$	-25	VaR	$\frac{LSM^{-25}}{LSM^{VaR}}$	-25	VaR	$\frac{LSM^{-25}}{LSM^{VaR}}$	-25	VaR	$\frac{LSM^{-25}}{LSM^{VaR}}$
65	1.01	0.43	2.38	0.97	0.41	2.37	0.85	0.36	2.34	0.68	0.29	2.32
70	1.05	0.46	2.30	0.98	0.43	2.29	0.81	0.36	2.27	0.56	0.25	2.25
75	1.05	0.48	2.21	0.94	0.43	2.21	0.69	0.31	2.20	0.38	0.17	2.21
80	0.99	0.47	2.13	0.83	0.39	2.13	0.49	0.23	2.15	0.17	0.06	2.21
85	0.87	0.43	2.04	0.62	0.30	2.06	0.23	0.11	2.13	-	-	-

Table 7.3: A Comparison of LSM^{VaR} and $LSM^{Shock=-25\%}$ for different products. $LSM^{VaR} \neq LSM^{SHOCK}$ at γ=-25%

7.6 Age Dependent Mortality Decline

The percentage decline in mortality at individual ages that is equivalent to the 99.5% percentile is approximately linear until the very elderly ages (90+) as shown in figure 7.11(a). Instead of using a simplification that assumes a constant percentage decline in mortality at all ages, a simplification that assumes the percentage decline increases with age is more realistic.

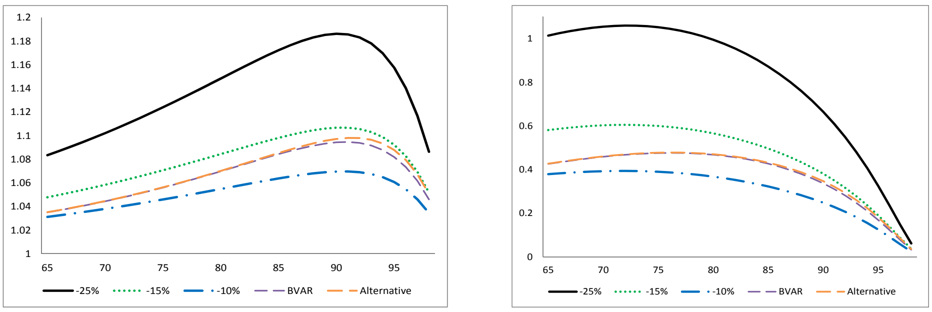


(a) The percentage decline that is equivalent to the 99.5% percentile

(b) A linear function representing the age dependent longevity stress.

Figure 7.11: An age dependent percentage mortality decline. The linear function used is Percentage Decline $= -1.63 + 0.1662 \times$ Age

A linear regression is fitted to the percentage declines between age 65 and 84. The estimated line is shown in figure 7.11(b). The line of regression used is Percentage Decline $= -1.63 + 0.1662 \times$ Age. For a more prudent measure, the line can be shifted upwards. The simplification presented in this section is not complex to implement.



(a) Ratio of the Value of Liabilities given different shock scenarios to the Value of liabilities given the Best Estimate for a_x

(b) $LSM^{-25\%}$, $LSM^{-15\%}$, $LSM^{-10\%}$, LSM^{VaR} and LSM^{Alt}

Figure 7.12: A comparison of the ratios, $\frac{VL^{\gamma}}{VL^{BE}}$ for a_x , and LSM under different constant percentage declines and under the alternative proposed in this thesis representing the percentage decline in mortality at individual ages that is equivalent to the 99.5% percentile.

The ratio of the value of liabilities given the alternative age-dependent stress is almost identical to the ratio of the value of liabilities given the 99.5% percentile of mortality rates (see figure 7.12(a)).

The age-dependent longevity stress margin, denoted by LSM^{Alt}, initially increases slightly and then decreases as age increases.

7.7 Discussion

A one year risk horizon is not appropriate for measuring longevity risk. The revamping of the Australian Prudential Regulatory Standards aims to harmonise capital requirements for both life and non-life insurance therefore it is a necessary shortcoming. The calibration of the longevity stress margin needs to be as risk sensitive as possible to mitigate some of the disadvantages of using such a short risk horizon. In its current form, the APRA-specified longevity stress margin is lacking in detail and is not an adequate measure of a longevity stress event.

One way of increasing the risk sensitivity of the longevity stress margin is to use an age-dependent stress. The APRA-specified longevity stress is a common mortality stress that is independent of age. Uncertainty in longevity has been shown to increase as age increases (see for example Blake et al., 2008). Ages with very wide bounds of prediction intervals represent the 'toxic tail'[1] effect. The 'toxic tail' effect for annuity providers is the high uncertainty surrounding number of annuitants that will survive to old ages (over 85) leading to uncertainty regarding the value of liabilities that will be paid out.

For an age-dependent stress there are two sets of arguments surrounding the uncertainty of mortality as age increases. Both agree that reductions in mortality rates are different over age but the degree of uncertainty in the reductions is where they diverge. In the first camp, Tabeau et al. (2002) and Börger (2010) argue that a reduction in uncertainty in mortality among the elderly diminishes by age. This is attributed to the fact that the elderly are exposed to more causes of death. This is interpreted to imply that if medical advancements eradicate a single cause of death, the gains in life-expectancy will be moderate due to the myriad of other causes of death that the elderly are still exposed to. Therefore, as age increases uncertainty in longevity does not increase. In the second camp, publications including Blake et al. (2008) and Sherris and Njenga (2011) show that longevity risk increases with age. In such studies the increased uncertainty is because mortality models are estimated using a limited amount of data and is therefore subject to parameter risk. This risk must be accounted for when projecting mortality trends. Keeping in mind the purpose of longevity stress margin as a reflection of the uncertainty in mortality due to changes in the level, trend and volatility of mortality rates the latter argument is preferable.[2]

[1]The phrase 'toxic tail' is used in Blake et al. (2008) and Tom Boardman of the (UK) Prudential is credited with coining the phrase in his address to "Longevity Two: The Second International Conference on Longevity Risk and Capital Market Solutions Conference" (Chicago) in 2006.

[2]Blake et al. (2008) find that for the CBD model (Cairns et al., 2006b) "As a rough rule of

In light of this, the issues regarding the magnitude and structure of the APRA-specified longevity stress scenario can be discussed with a possible alternative age-dependent stress scenario in mind. The magnitude of the longevity stress scenario is currently set as a constant 25% decline in mortality rates at all ages. Since mortality rates generally increase with age in absolute terms the size of the longevity stress scenario increases with age (intuitively, if the rates were 0.03 and 0.3 at ages 65 and 90 respectively, the longevity stress scenarios would be 0.0225 and 0.225). The result is an increase in the amount of capital required or the longevity stress margin given the shock.

The magnitude of the shock should be revised downwards from 25% to between 10% and 15%. At a stress of 25% the longevity stress margin is more than double the longevity stress margin with the stochastic stress for the immediate and deferred life annuities. This implies that in its current form the APRA-specified longevity stress is highly likely lead to overcapitalization leading to more expensive life annuities as the cost of capital will be passed on to retirees who purchase the annuities.

The longevity stress margin increases and then decreases with age (see figure 7.10). For the immediate life annuity, the ages with the maximum LSM for the constant shock and for the VaR are different. The constant shock's LSM attains its maximum value between ages 70 and 75 while for the VaR the LSM has its maximum between ages 80 and 85. Therefore, the VaR recognizes that there is significant uncertainty in longevity for Australian males aged 80-85. This is consistent with the observation in Blake et al. (2008) that the uncertainty in longevity increases with age and therefore additional capital required to account for the insurer's exposure to the 'toxic tail' of the elderly.

An age dependent stress which increases with age is proposed as an alternative to the constant mortality decline stress. This is contrary to Börger (2010) where an age dependent stress with smaller relative reductions for old ages is proposed as an alternative. The beauty of this alternative is that in addition to its simplicity, it is flexible and can be shifted upwards when a more prudent longevity stress margin is desired.

The linear mortality stress that increases with age which is proposed in this chapter results in a LSM which is approximately equal to the LSM of the 99.5% VaR.

thumb, allowing for uncertain parameters nearly doubles the dispersion of each of our fan charts over the age ranges where there is serious uncertainty about future survival rates. Failing to allow for parameter uncertainty, therefore, leads to fan chart forecasts that are far too narrow and grossly underestimate longevity risk". In the HP-VAR and HP-BVAR models the width of the HP-BVAR is much wider than the HP-VAR model. In the simulation scenario the HP-VAR confidence intervals for p_x are so narrow they appear as a single line.

It addresses the issues of magnitude and structure that reduce the risk sensitivity of the APRA-specified simplification where the longevity stress event is a constant decline in mortality at all ages. The alternative longevity stress presented in this chapter is flexible and simple to implement.

7.8 Recommendation for Further Research

This analysis is based on population data for males. Further research using data from a portfolio of annuities should be done to check how different annuitant mortality is from population mortality for Australia. Also, research that considers the movements of interest rates and their effects on the longevity stress margin will also be beneficial to insurers, reinsurers and regulators.

More studies comparing pensioner mortality in different industries to population mortality can also be done. Then, a population based longevity stress scenario such as the one presented in this thesis can be used as a basis for calculating the longevity stress margins for an annuitant who previously worked in a given industry.

A similar study should also be done for female mortality. Granularization of the longevity stress event by gender should be considered. Male and female mortality behaves differently therefore it is possible that a one in two hundred year mortality improvements will be different by gender.

7.9 Conclusion

The insurance capital regulator, APRA, has specified a simple formula for determining a one in two hundred year improvement in mortality. The assumptions regarding the magnitude and structure of the longevity stress have been shown to be inadequate. An alternative simple age-dependent stress based on a linear equation is found to be a better reflection of the one in two hundred year improvement in mortality. It is attractive as it gives greater uncertainty to mortality of at older ages which is realistic because of the high uncertainty surrounding number of annuitants that will survive to old ages. The linear alternative is therefore more risk sensitive and appropriate for setting capital standards.

8

Conclusion

This thesis has covered an analysis of longevity risk modelling and, as an application, looked at the quantification of longevity risk to insurers' risk based capital margins. It begun with a brief introduction where the motivation for the study and the research questions were presented. The challenge of quantifying longevity risk was presented. Next, existing relevant literature was reviewed and some models were introduced. After that the data used in the study was described in Chapter 3. The methodology followed in Chapter 4. Chapters 5, 6 and 7 presented and discussed the results of the three studies that formed this thesis. The contributions of this thesis to demography and actuarial science are summarised in the following paragraphs where the key findings and conclusions are reiterated.

8.1 Summary of Key Findings

The change in mortality rates varies by country and gender. There is dependence across age within a country and this presents a problem because it leads to systematic risk. The first study presented in Chapter 5 was an analysis of the historical features of mortality levels, trends and volatilities. The aim of the first study was to quantify and give an in-depth understanding of mortality trends and volatility. A variety of dimension reduction techniques were used to bring out information about different features of the mortality rates and their changes. The behaviour of mortality rates in different countries with comparable standards of living was analysed. The observable forces (principal components) and the unobservable forces (factors) that underlie mortality were clearly distinguished from each other using principal component analysis and factor analysis respectively.

It was found in relation to each other, the number of factors and principal components needed to explain mortality trends varied by country and gender but in general 7 or 8 factors are sufficient and females tend to require one less factor than males. Period or time trends in mortality rates are explained by fewer factors (between 2 and 5) while the cohort trends are explained by more factors (between 2 and 10). This is important because the time trends only capture information due to changes in time while the cohort trends include changes in time and by age and need additional factors to explain the added information. It is interesting to note that for the two countries from the same geographical region, UK and Norway, although the number of factors required was different the behaviour of the factors was similar. Some countries in the study also exhibited a stronger cohort presence than others. For example, the USA requires 9 (males) and 10 (females) factors to explain cohort trends compared to Norway's 3 (males) and 4 (females). In comparison, one principal component is sufficient to explain over 90% of the cumulative variation in mortality trends for all the countries studied except Norway. The number of principal components required to explain time trends varies by gender and country with Japanese males requiring the least, followed by Japanese females. In general, the number of principal components that explain 90% of the cumulative variation in time trends was found to be less than the number of principal components that explain 90% of the cumulative variation in cohort trends. This is surprising because of change in age provides an additional source of variation in cohort trends.

Econometric models were also used to examine the historical features of the levels of mortality trends with emphasis on cross-country mortality trends. Unit root and stationarity tests were performed on the standardised mortality rate time series and their first differences for the period from 1963-2007 and confirmed that the standardised mortality rates for all the countries in the study were first order integrated, I(1). A first order Vector Autoregressive model (VAR(1)) was found to be sufficient but surprisingly there were no common stochastic country trends and therefore the first differences of the standardised mortality rates should be modelled.

The second study applied some of the findings from the first study to quantify the uncertainty in mortality projections. The aim of the second study was to develop a dynamic parametric mortality model that is parsimonious and quantifies longevity risk including the portion of uncertainty in longevity risk due to parameter risk. In Chapter 6 an existing static parametric mortality model was extended to capture the effects of common trends in a given population using an innovative combination of demography models and econometric techniques. First, a parametric mortality law was used to reduce the dimension of the data using an 8 factor non-linear model - the Heligman-Pollard model. Then, the interdependencies in the parameters of the Heligman-Pollard models were captured using econometric techniques thereby transforming a static parametric mortality model into a dynamic parametric mortality model. Unrestricted and Bayesian Vector Autoregression models were used to estimate the HP-VAR and HP-BVAR models respectively. The HP-BVAR model is based on a parsimonious Bayesian Vector Autoregression model which captures parameter risk. The resulting model performed better than previous attempts at making a static model dynamic such as McNown and Rogers (1989). In particular, the model presented in this thesis is based on a trade-off between under-parameterization due to using univariate methods and over-parameterization due to using the unrestricted VAR. The portion of uncertainty in longevity risk that is contributed by uncertainty from parameter estimation was incorporated. The model performed well for Australian males whose Heligman-Pollard parameters were highly correlated and produced realistic probabilistic projections.

Chapter 7 was an application of the model estimated in the second study. The aim of the third study was to test the adequacy of the magnitude and structure of the APRA specified simplification of longevity stress margin. Insurer longevity risk based capital stress margins were examined to check the adequacy of the insurance capital regulator's assumption regarding the longevity stress. The magnitude and structure of the APRA specified simplification of longevity stress margin were considered. It was concluded that its current form the APRA-specified simplification will lead to over-capitalization. An age dependent stress captured the one in two hundred year event better than the current specification.

8.2 Implications of Findings

In most of the countries studied in Chapter 5 had a similar number of factors driving changes in the mortality rates. The econometric analysis found that standardized mortality rates across the countries studied had stochastic trends based on the historical data but the stochastic trends were not common to all the countries in this analysis. The findings of the research in Chapter 5 are based on a range of models. The econometric models presented allowed for volatilities and correlations between the mortality rates and provide a relatively parsimonious structure. These models allow the quantification of the benefits of diversification in portfolios as well as a consistent framework for modelling multivariate risk factors where some of these risk factors are non-stationary and others are stationary. This is relevant for reinsurers who are exposed to mortality risk on an international horizon.

The combination of advanced econometric techniques and static parametric mortality models to form a dynamic parametric mortality model (the HP-BVAR model) in Chapter 6 has shown that parametric mortality models are important in stochastic mortality modelling. Incorporating interdependencies between a parametric model's parameters and the parameter risk due to the limited amount of data available can lead to improved forecasts that are realistic and probabilistic. The parameters of the Heligman-Pollard model for males have often been observed to be highly correlated leading to problems in modelling the evolution of the parameters (Hartmann, 1987; Booth and Tickle, 2008). Using the Bayesian Vector Autoregressive model with suitable hyper parameters improved the forecasted mortality rates for Australian males studied in this thesis. These findings imply that the time evolution of parameters of a wide variety of mortality models can be modelled using econometric models and used to predict the changing shape of mortality profiles.

The results in Chapter 7 suggest that the magnitude and structure of the simplification of a longevity stress based on the assumption of a constant decline in mortality at all ages is not a realistic reflection of a one in two hundred year longevity stress event. It is very likely that the current specification of the longevity stress will lead to over-capitalization. It is certain that the magnitude and structure of the simplification needs re-examination. A simple alternative structure that considers the different amount of uncertainty in longevity at different ages is the linear age-dependent mortality stress determined in Chapter.

8.3 Limitations of the research and Recommendations for future work

The methods and findings in this thesis form a small part of the vast research area in longevity risk. There are several limitations to the research in this thesis. Looking at the limitations of this research is useful for putting forward recommendations for future work therefore this section combines the limitations of the three studies in this thesis as well as potential future research based on this thesis.

With regards to the first study in Chapter 5, the main limitation is that a small number of countries is analysed. This introduces bias into the study as it solely concentrates on five countries - Australia, Japan, Norway, the UK and the USA. Further research using a broader base of countries will add to the knowledge bank on the forces that drive trends in mortality levels as well as the forces that drive their volatilities as measured by mortality time trends and mortality cohort trends.

As for the second study in Chapter 6 the static parametric mortality model used is from Heligman and Pollard (1980). This parametric model is adequate but not perfect. It is the parametric mortality model for which previous attempts to make it dynamic have been published (Forfar and Smith, 1987; McNown and Rogers, 1989, 1992). Therefore, this model was suitable for comparability of the findings of this thesis to the results of previous attempts to convert a static Heligman-Pollard mortality model into a dynamic Heligman-Pollard mortality model. Future work can be done on other parametric mortality models and on different data sets such as cause-specific mortality data. The multivariate econometric techniques used for analysis on longitudinal data in this thesis certainly merit further investigation for use with parametric models for fertility and migration.

The assumptions in Chapter 7 are two key sources of the limitations of the results of the the third study. The data and literature on Australian annuitant mortality is very limited and is not readily available for academic research. Therefore, the first significant limitation is the third study does not account for basis risk because it is based on population mortality rates. More work can be done using annuitant data from annuity providers and pensioner data from superannuation schemes and used to model mortality and longevity uncertainty. In particular, longevity stress margins for joint lives (on a joint life and last survivor annuity) would be an interesting research area due to the differences in male and female longevity. Research into a different specification of the simplified longevity stress margin also has a lot of potential. APRA's specification will be used to simplify the calculation of longevity stress margins in one country compared to CEIOPS's specification which needs to cover more than 30 countries. Therefore, a detailed simplification with a better structure than the existing constant shock can be developed. Further, this can be done at a granular level catering for mortality experience for different groups (it is very probable that improvements in longevity for Australians varies by socio-economic groups) and taking into account different longevity insurance products (with different q-durations). The second major limitation of the third study is that interest rate risk is not considered. Future research that incorporates the impact of economic uncertainty as well as uncertainty surrounding future longevity (with basis risk) will provide additional important information that can be used to calibrate insurer longevity risk based capital stress margin in a more realistic way.

8.4 Summary

The three main contributions of this thesis can each be summarised in one sentence each as:

- A better understanding of mortality trends and their volatilities.
- A dynamic parametric model that produces realistic probabilistic projections.
- An age-dependent longevity stress for computation of insurer's risk based capital.

References

1. Abel, G. J., Bijak, J., Forster, J. J., Raymer, J., and Smith, P. W. (2010). What Do Bayesian Methods Offer Population Forecasters? ESRC Centre for Population Change, UK, Working Paper, 6.

2. Alai, D. H. and Sherris, M. (2011). Rethinking Age-Period-Cohort Mortality Trend Models. *UNSW Australian School of Business Research Paper No. 2011 ACTL2009*. Available at http://ssrn.com/paper=1838805.

3. Alexander, C. (1999). Optimal hedging using cointegration. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 357(1758):2039–2058.

4. Alho, J. (2000). Discussion of Lee (2000). *North American Actuarial Journal*, 4:91–93.

5. Andreev, K. F. and Vaupel, J. W. (2005). Patterns of Mortality Improvement over Age and Time: Estimation, Presentation and Implications. Presented to the Population Association of America 2005 Annual Meeting, 31 March - 2 April 2005.

6. Andreev, K. F. and Vaupel, J. W. (2006). Forecasts of cohort mortality after age 50. *MPIDR Working Paper*. Max Planck Institute for Demographic Research, WP-2006-012.

7. APRA (2010a). Discussion Paper: Review of capital standards for general insurers and life insurers. Discussion paper, Australian Prudential Regulation Authority. Available at http://www.apra.gov.au/Policy/upload/GLI_DP_RCSGILI_032010_v7.pdf.

8. APRA (2010b). Review of capital standards for general insurers and life insurers: Capital base and insurance risk capital charge for life insurers. Technical report, Australian Prudential Regulation Authority. Available at http://www.apra.gov.au/Policy/upload/GLI_TP_CBIRCC_072010_v8.pdf.

9. APRA (2011). Technical Specifications for QIS2: Review of capital standards for general insurers and life insurers - Life insurance (including friendly societies). Technical report, Australian Prudential Regulation Authority. Available at http://www.apra.gov.au/Policy/upload/LI-QIS2-Technical-Specifications-13-May-2011.pdf.

10. Baltagi, B. (2002). *Econometrics*. Springer-Verlag, New York, 3 edition.

11. Banerjee, A. (1993). *Co-integration, error correction, and the econometric analysis of non-stationary data*. Recent advances in econometrics series. Oxford University Press.

12. Bartlett, M. S. (1954). A Note on the Multiplying Factors for Various χ^2 Approximations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 16(2):296–298.

13. Bauwens, L., Lubrano, M., and Richard, J. (1999). *Bayesian inference in dynamic econometric models*. Advanced texts in econometrics. Oxford University Press.

14. Bell, W. R. (1997). Comparing and Assessing Time Series Methods for Forecasting Age-Specific Fertility and Mortality Rates. *Journal of Official Statistics*, 13(3):279–303.

15. Bell, W. R. and Monsell, B. C. (1991). Using principal components in time series modeling and forecasting of age-specific mortality rates. Paper presented at the American Statistical Association 1991 Proceedings of the Social Statistics Section.

16. Biffis, E. (2005). Affine processes for dynamic mortality and actuarial valuations. *Insurance: Mathematics and Economics*, 37(3):443 – 468.

17. Blake, D., Dowd, K., and Cairns, A. J. G. (2008). Longevity risk and the Grim Reaper's toxic tail: The survivor fan charts. *Insurance: Mathematics and Economics*, 42(3):1062–1066.

18. Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 22(3):547 – 581.

19. Booth, H., Hyndman, R., Tickle, L., and de Jong, P. (2006). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research*, 15(9):289–310.

20. Booth, H., Maindonald, J., and Smith, L. (2002a). Age-time interactions in mortality projection: applying Lee-Carter to Australia. Working paper, Australian National University, Australia. Available at http://adsri.anu.edu.au/pubs/demog-pubs/WorkingPapers/85.pdf.

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

22. Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, 3:3–43(41).

23. Börger, M. (2010). Deterministic shock vs. stochastic value-at-risk an analysis of the Solvency II standard model approach to longevity risk. *Blätter der DGVFM*, 31(2):225–259.

24. Box, G. and Jenkins, G. (1976). *Time Series Analysis: Forecasting and Control (Wiley Series in Probability and Statistics)*. Holden-Day San-Francisco.

25. Bozik, J. E. and Bell, W. R. (1987). Forecasting age specific fertility using principal components. *Paper presented at the Proceedings of the Social Statistics Section, American Statistical Association, San Francisco, California.*

26. Brandt, P. (2011). *MSBVAR: Markov-Switching, Bayesian, Vector Autoregression Models*. R package version 0.6-0 Available at http://CRAN.R-project.org/package=MSBVAR.

27. Brandt, P. T. and Freeman, J. R. (2006). Advances in Bayesian Time Series Modeling and the Study of Politics: Theory Testing, Forecasting, and Policy Analysis. *Political Analysis*, 14:1–36.

28. Brooks, C. (2008). *Introductory econometrics for finance*. Cambridge University Press.

29. Cairns, A. J. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics*, 27(3):313 – 330.

30. Cairns, A. J., Blake, D., and Dowd, K. (2006a). Pricing Death Frameworks for the Valuation and Securitization of Mortality Risk. *ASTIN Bulletin*, 36(1):79 –120.

31. Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., and Khalaf-Allah, M. (2011). Bayesian Stochastic Mortality Modelling for Two Populations. *To appear in ASTIN Bulletin.*

32. Cairns, A. J. G., Blake, D., and Dowd, K. (2006b). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance*, 73(4):687–718.

33. Carrière, J. F. (1992). Parametric models for life tables. *Transactions of the Society of Actuaries*, 44:77–99.

34. Carter, L. R. and Prskawetz, A. (2001). Examining structural shifts in mortality using the Lee-Carter method. *Demographic Research*, pages 1–16. Max Planck Institute for Demographic Research Working Papers.

35. Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, 1(2):245–276.

36. Cattell, R. B. (1983). This Week's Citation Classic. *Current Contents - Social & Behavioral Sciences*, 1(5):16.

37. CEIOPS (2007). Quantitative impact study 3. Technical report, Committee of European Insurance and Occupational Pensions Supervisors.

38. CEIOPS (2008). CEIOPS' Report on its fourth Quantitative Impact Study (QIS4) for Solvency II. Technical report, Committee of European Insurance and Occupational Pensions Supervisors.

39. CEIOPS (2009). CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: Standard formula SCR - Article 109 c Life underwriting risk. Technical report, Committee of European Insurance and Occupational Pensions Supervisors.

40. CEIOPS (2010). QIS5 Technical Specifications: Annex to Call for Advice from CEIOPS on QIS5. Technical report, Committee of European Insurance and Occupational Pensions Supervisors.

41. Chan, C.-L. and Ting, H.-W. (2011). Constructing a novel mortality prediction model with Bayes theorem and genetic algorithm. *Expert Systems with Applications*, 38:7924–7928.

42. Chan, N. (2002). *Time series: applications to finance*. Wiley-Interscience.

43. Chunn, J. L., Raftery, A. E., and Gerland, P. (2010). Bayesian Probabilistic Projections of Mortality. Technical report, Center for Statistics and the Social Sciences. Presented to the Population Association of America (PAA) 2010 Annual Meeting.

44. Congdon, P. (1993). Statistical Graduation in Local Demographic Analysis and Projection. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 156(2):237–270.

45. Coughlan, G., Epstein, D., Ong, A., Sinha, A., Hevia-Portocarrero, J., Gingrich, E., Khalaf-Allah, M., and Joseph, P. (2007). Lifemetrics: A toolkit for measuring and managing longevity and mortality risks. Technical report, JP Morgan: Pension Advisory Group.

46. Cox, S. H., Lin, Y., and Pedersen, H. (2010). Mortality risk modeling: Applications to insurance securitization. *Insurance: Mathematics and Economics*, 46(1):242 – 253.

47. Cureton, E. and D'Agostino, R. (1993). *Factor Analysis: An Applied Approach*. Lawrence Erlbaum Associates.

48. Darkiewicz, G. and Hoedemakers, T. (2004). How the co-integration analysis can help in mortality forecasting. Open Access publications from Katholieke Universiteit Leuven urn:hdl:123456789/85470, Katholieke Universiteit Leuven.

49. Deaton, A. and Paxson, C. (2001). Mortality, Income, and Income Inequality Over Time in Britain and the United States. Working Paper 8534, National Bureau of Economic Research.

50. Dellaportas, P., Smith, A. F. M., and Stavropoulos, P. (2001). Bayesian analysis of mortality data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164(2):275–291.

51. Dellinger, J. K. (2006). *The Handbook of Variable Income Annuities*. Finance. Wiley.

52. Denton, F. T., Feaver, C. H., and Spencer, B. G. (2005). Time Series Analysis and Stochastic Forecasting: An Econometric Study of Mortality and Life Expectancy. *Journal of Population Economics*, 18(2):pp. 203–227.

53. Denuit, M. and Frostig, E. (2009). Life insurance mathematics with random life tables. *North American Actuarial Journal*, 13(3):339–355.

54. Dickey, D. A. and Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74(366):pp. 427–431.

55. Doan, T., Litterman, R., and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3(1):1 – 100.

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

57. Drèze, J. H. and Richard, J. F. (1983). Bayesian analysis of simultaneous equation systems. volume 1, chapter 9, pages 517–598.

58. Dunteman, G. (1989). *Principal Components Analysis*. Quantitative Applications in the Social Sciences Series. Sage.

59. Forfar, D. O. and Smith, D. M. (1987). The changing shape of English Life Tables. *Transactions of the Faculty of Actuaries*, (40):98–134.

60. Fries, J. F. (1980). Aging, Natural Death, and the Compression of Morbidity. *New England Journal of Medicine*, 303(3):130–135.

61. Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85(410):398–409.

62. Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741.

63. Girosi, F. and King, G. (2007). Understanding the Lee-Carter Mortality Forecasting Method. Available at http://gking.harvard.edu/sites/scholar.iq.harvard.edu/files/gking/files/lc.pdf.

64. Girosi, F. and King, G. (2008). *Demographic forecasting*. Princeton University Press.

65. Gupta, R. and Kabundi, A. (2011). A large factor model for forecasting macroeconomic variables in South Africa. *International Journal of Forecasting*, In Press, Corrected Proof:–.

66. Haberman, S. and Renshaw, A. (2008). Mortality, longevity and experiments with the Lee-Carter model. *Lifetime Data Analysis*, 14:286–315.

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

68. Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press.

69. Hanewald, K. (2010). Explaining Mortality Dynamics: The Role of Macroeconomic Fluctuations and Cause of Death Trends. *North American Actuarial Journal, Forthcoming*. Available at http://ssrn.com/paper=1336888.

70. Hannerz, H. (2001a). An extension of relational methods in mortality estimations. *Demographic Research*, 4(10):337–368.

71. Hannerz, H. (2001b). Manhood Trials and the Law of Mortality. *Demographic Research*, 4(7):185–202.

72. Hannerz, H. (2001c). Presentation and derivation of a five-parameter survival function intended to model mortality in modern female populations. *Scandinavian Actuarial Journal*, 2001(2):176–187.

73. Härdle, W. and Simar, L. (2007). *Applied Multivariate Statistical Analysis*. Springer.

74. Hari, N., Waegenaere, A. D., Melenberg, B., and Nijman, T. E. (2008). Estimating the term structure of mortality. *Insurance: Mathematics and Economics*, 42(2):492 – 504.

75. Hartmann, M. (1987). Past and recent attempts to model mortality at all ages. *Journal of Offical Statistics*, 3(1):19–36.

76. Harvey, A. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.

77. Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

78. Hatzopoulos, P. and Haberman, S. (2011). A dynamic parameterization modeling for the age-period-cohort mortality. *Insurance: Mathematics and Economics*, 49(2):155 – 174.

79. Heligman, L. and Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107:49–80.

80. Henry, K. (2009). Australia's future tax system - The retirement income system: Report on strategic issues.

81. Hyndman, R. J. and Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, 24(3):323–342.

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

83. Jarner, S. F. and Kryger, E. M. (2009). Modelling adult mortality in small populations: The SAINT model. Technical report. Pensions Institute Discussion Paper PI-0902. Available at http://www.pensions-institute.org/workingpapers/wp0902.pdf.

84. Joiner, A. (2001). Monetary Policy Effect in an Australian Bayesian VAR Model. Working papers, Monash University. Australia.

85. Kadiyala, K. R. and Karlsson, S. (1997). Numerical Methods for Estimation and Inference in Bayesian VAR-Models. *Journal of Applied Econometrics*, 12(2):99–132.

86. Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, 20(1):141–151.

87. Klugman, J., Rodriguez, F., Kovacevic, M., Jespersen, E., Orme, W., and Mend, S. (2010). Human Development Report 2010 - 20th Anniversary Edition The Real Wealth of Nations: Pathways to Human Development. Technical report, United Nations Development Programme - Human Development Report Office.

88. Knox, D. and Nelson, M. (2006). The mortality experience of Australian superannuation pensioners: Some surprising results. Presented to the Institute of Actuaries of Australia, Third Financial Services Forum, 11 - 12 May 2006.

89. Kogure, A. and Kurachi, Y. (2010). A bayesian approach to pricing longevity risk based on risk-neutral predictive distributions. *Insurance: Mathematics and Economics*, 46(1):162 – 172.

90. Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1-3):159–178.

91. Lansangan, J. and Barrios, E. (2009). Principal components analysis of nonstationary time series data. *Statistics and Computing*, 19:173–187.

92. Lazar, D. (2004). On forecasting mortality using Lee-Carter method. Presented at the 3rd Conference in Actuarial Science and Finance in Samos, September 2-5, 2004.

93. Ledermann, S. and Breas, J. (1959). Les dimensions de la mortalitè. *Population*, 14(4):637–682.

94. Lee, R. and Miller, T. (2001). Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality. *Demography*, 38(4):537–549.

95. Lee, R. D. (1992). Stochastic demographic forecasting. *International Journal of Forecasting*, 8(3):315–327.

96. Lee, R. D. and Carter, L. R. (1992). Modeling and Forecasting U. S. Mortality. *Journal of the American Statistical Association*, 87(419):659–671.

97. Li, J. S.-H., Chan, W.-S., and Cheung, S.-H. (2011). Structural Changes in the Lee-Carter Mortality Indeces: Detection and Implications. *North American Actuarial Journal*, 15(1):13–31.

98. Litterman, R. B. (1986). Forecasting with Bayesian Vector Autoregressions: Five Years of Experience. *Journal of Business and Economic Statistics*, 4(1):25–38.

99. Lord Penrose (2004). Report of The Equitable Life Inquiry. Available at http://news.bbc.co.uk/nol/shared/bsp/hi/pdfs/penrose_part1_opt.pdf.

100. Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. New York, 1st edition.

101. Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. New York, 2nd edition.

102. McNeil, D. R., Trussell, T. J., and Turner, J. C. (1977). Spline Interpolation of Demographic Data. *Demography*, 14(2):245–252.

103. McNown, R. and Rogers, A. (1989). Forecasting Mortality: A Parameterized Time Series Approach. *Demography*, 26(4):645–660.

104. McNown, R. and Rogers, A. (1992). Forecasting cause-specific mortality using time series methods. *International Journal of Forecasting*, 8(3):413 – 432.

105. Ni, S. and Sun, D. (2003). Noninformative priors and frequentist risks of bayesian estimators of vector-autoregressive models. *Journal of Econometrics*, 115:159–197.

106. O'Hare, C. and French, D. (2011). Beyond Static Factor Mortality Modeling. *SSRN eLibrary*. Available at http://ssrn.com/paper=1843123.

107. Olivieri, A. and Pitacco, E. (2008). Solvency Requirements for Life Annuities: Some Comparisons. *SSRN eLibrary*. Available at SSRN: http://ssrn.com/abstract=1266094.

108. Pedroza, C. (2006). A Bayesian forecasting model: predicting U.S. male mortality. *Biostatistics*, 7(4):530–550.

109. Peng, R. (2008). A Method for Visualizing Multivariate Time Series Data. *Journal of Statistical Software, Code Snippets*, 25(1):1–17.

110. Pfaff, B. (2008). VAR, SVAR and SVEC Models: Implementation Within R Package vars. *Journal of Statistical Software*, 27(4):1–32.

111. Phillips, P. C. B. and Perron, P. (1988). Testing for a Unit Root in Time Series Regression. *Biometrika*, 75(2):335–346.

112. Plat, R. (2010). One-Year Value-at-Risk for Longevity and Mortality. *Pensions Institute Discussion Paper PI-1015*. Available at http://www.pensions-institute.org/workingpapers/wp1015.pdf.

113. Pollard, J. (1996). On the changing shape of the Australian mortality curve. *Health Transition Review, Supplement 6*, pages 283–300.

114. R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0 http://www.R-project.org/.

115. Rachev, S., Mittnik, S., Fabozzi, F. J., Focardi, S. M., and Jašić, T. (2007). *Financial econometrics: from basics to advanced modeling techniques*. Frank J. Fabozzi series. Wiley.

116. Rao, C. R. (1964). The Use and Interpretation of Principal Component Analysis in Applied Research. *Sankhya: The Indian Journal of Statistics, Series A*, 26(4):329–358.

117. Reichmuth, W. and Sarferaz, S. (2008). Modeling and Forecasting Age-Specific Mortality: A Bayesian Approach. Humboldt-Universität zu Berlin, Germany. SFB 649 Discussion Paper No. 2008-052a.

118. Renshaw, A. E. and Haberman, S. (2008). On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee-Carter modelling. *Insurance: Mathematics and Economics*, 42(2):797–816.

119. Risk Management Solutions, Inc. (2010). Longevity Risk. Available at http://www.rms.com/Publications/Longevity_Risk_brochure.pdf.

120. Risk Management Solutions, Inc. (2010). RMS Unveils New Approach To Quantifying Longevity Risk. Available at http://www.rms.com/news/NewsPress/PR_071210_LongevityRisk.asp.

121. Robertson, J. C. and Tallman, E. W. (1999a). Prior parameter uncertainty: Some implications for forecasting and policy analysis with VAR models. Working Paper 99-13, Federal Reserve Bank of Atlanta.

122. Robertson, J. C. and Tallman, E. W. (1999b). Vector autoregressions: forecasting and reality. *Economic Review*, (Q1):4–18.

123. Rogers, A. (1983). Model: A general program for estimating parameterized model schedules of fertility, mortality, migration and marital and labor force status transitions. Technical Report WP-83-102, Laxenburg.

124. Rogers, A. (1986). Parameterized Multistate Population Dynamics and Projections. *Journal of the American Statistical Association*, 81(393):48–61.

125. Rogers, A. and Gard, K. (1991). Applications of the Heligman/Pollard Model Mortality Schedule. *Population Bulletin of the United Nations*, 30.

126. Sevinç, V. and Ergun, G. (2009). Usage of Different Prior Distributions in Bayesian Vector Autoregressive Models. *Hacettepe Journal of Mathematics and Statistics*, 38(1):85–93.

127. Shang, H. L., Hyndman, R. J., and Booth, H. (2010). A comparison of ten principal component methods for forecasting mortality rates. Monash Econometrics and Business Statistics Working Papers 8/10, Monash University, Department of Econometrics and Business Statistics.

128. Sharrow, D. J., Clark, S. J., Collinson, M. A., Kahn, K., and Tollman, S. M. (2010). The Age-Pattern of Increases in Mortality Affected by HIV: Bayesian Fit of the Heligman-Pollard Model to Data from the Agincourt HDSS Field Site in Rural Northeast South Africa.

129. Sherris, M. and Evans, J. R. (2010). Longevity Risk Management and the Development of a Life Annuity Market in Australia. *SSRN eLibrary*. Available at SSRN: http://ssrn.com/abstract=1585563.

130. Sherris, M. and Gaille, S. (2010a). Improving Longevity and Mortality Risk Models with Common Stochastic Long-Run Trends. *SSRN eLibrary*. Available at http://ssrn.com/paper=1702029.

131. Sherris, M. and Gaille, S. (2010b). Modeling Long-Run Cause of Death Mortality Trends. *SSRN eLibrary*. Available at http://ssrn.com/paper=1705696.

132. Sherris, M. and Njenga, C. N. (2009). Longevity Risk and the Econometric Analysis of Mortality Trends and Volatility. *SSRN eLibrary*. Available at SSRN: http://ssrn.com/abstract=1458084.

133. Sherris, M. and Njenga, C. N. (2011). Modeling Mortality with a Bayesian Vector Autoregression. *SSRN eLibrary*. Available at SSRN: http://ssrn.com/paper=1776532.

134. Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1):1–48.

135. Sims, C. A., Stock, J. H., and Watson, M. W. (1990). Inference in Linear Time Series Models with some Unit Roots. *Econometrica*, 58(1):pp. 113–144.

136. Sims, C. A. and Zha, T. (1998). Bayesian Methods for Dynamic Multivariate Models. *International Economic Review*, 39(4):949–968.

137. Sims, C. A. and Zha, T. (1999). Error Bands for Impulse Responses. *Econometrica*, 67(5):pp. 1113–1155.

138. Sinay, M. S. (2008). *Bayesian inference for linear and generalized linear models with a flexible prior structure on the covariance matrix*. PhD thesis, University of California, Davis.

139. Stevenson, M. and Wilson, A. (2008). Mortality of Public Sector Scheme Pensioners 2005-2007 update. Presented to the Institute of Actuaries of Australia, Fourth Financial Services Forum, 19-20 May 2008.

140. Stock, J. H. (2001). Vector autoregressions. *Journal of Economic Perspectives*, 15(4):101–115.

141. Summers, P. M. (2001). Forecasting Australia's economic performance during the Asian crisis. *International Journal of Forecasting*, 17(3):499–515.

142. Tabeau, E. (2002). A review of demographic forecasting models for mortality. In *Forecasting Mortality in Developed Countries*, volume 9 of *European Studies of Population*, pages 1–32. Springer Netherlands.

143. Tabeau, E., Van Den Berg Jeths, A., and Heathcote, C. (2002). Towards an Integration of the Statistical, Demographic and Epidemiological Perspectives in Forecasting Mortality . In *Forecasting Mortality in Developed Countries*, volume 9 of *European Studies of Population*, pages 281–299. Springer Netherlands.

144. Thatcher, A. R. (1999). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(1):5–43.

145. Thompson, P. A., Bell, W. R., Long, J. F., and Miller, R. B. (1989). Multivariate Time Series Projections of Parameterized Age-Specific Fertility Rates. *Journal of the American Statistical Association*, 84(407):689–699.

146. Thurstone, L. L. (1934). The vectors of mind. *Psychological Review*, 41(1):1–32.

147. Triacca, U. (2002). Cointegration in VAR(1) process: characterization and testing. *Statistical Papers*, 43(3):435.

148. Tsai, J. T., Wang, J. L., and Tzeng, L. Y. (2010). On the optimal product mix in life insurance companies using conditional value at risk. *Insurance: Mathematics and Economics*, 46(1):235 – 241.

149. Tuljapurkar, S. and Edwards, R. D. (2011). Variance in death and its implications for modeling and forecasting mortality. *Demographic Research*, 24(21):497–526.

150. Tuljapurkar, S., Li, N., and Boe, C. (2000). A universal pattern of mortality decline in the G7 countries. *Nature*, 405(6788):789–792.

151. Waggoner, D. F. and Zha, T. (1999). Conditional Forecasts in Dynamic Multivariate Models. *The Review of Economics and Statistics*, 81(4):639–651.

152. Wang, J. L., Huang, H., Yang, S. S., and Tsai, J. T. (2010). An Optimal Product Mix for Hedging Longevity Risk in Life Insurance Companies: The Immunization Theory Approach. *Journal of Risk and Insurance*, 77(2):473–497.

153. Willets, R. C. (2004). Longevity in the 21st Century. *British Actuarial Journal*, 10:685–832(148).

154. Wilmoth, J., Andreev, K., Jdanov, D., and Glei, D. (2007). Methods Protocol for the Human Mortality Database. Technical report, Human Mortality Database.

155. Wolf, D. A. (2004). Another Variation on the Lee-Carter Model. Presented to the Population Association of America 2004 Annual Meeting, 1 April 2004.

156. Wolff, H., Chong, H., and Auffhammer, M. (2011). Classification, Detection and Consequences of Data Error: Evidence from the Human Development Index. *The Economic Journal*, 121(553):843–870.

157. Wong-Fupuy, C. and Haberman, S. (2004). Projecting Mortality Trends: Recent Developments in the United Kingdom and the United States. *North American Actuarial Journal*, 8(2):56–83.

158. Yang, S. S., Yue, J. C., and Huang, H.-C. (2010). Modeling longevity risks using a principal component approach: A comparison with existing stochastic mortality models. *Insurance: Mathematics and Economics*, 46(1):254 – 270.

159. Zivot, E. and Wang, J. (2006). *Modelling Financial Time Series with S-plus.* Second edition.

Appendix A

PCA

A.1 Cattell's Scree Test

A visual way to determine the number of common factors is the scree test developed by Cattell (1966). Cattell (p.16 1983) has the following description by Cattell of his method:

> To my delight, a very simple finding presented itself, namely, that if I plotted the principal components in their sizes, as a diminishing series, and then joined up the points all through the number of variables concerned, a relatively sharp break appeared where the true number of factors ended and the detritus, presumably due to error factors, appeared. From the analogy of the steep descent of a mountain till one comes to the scree of rubble at the foot of it, I decided to call this the scree test.

The estimated number of factors by Cattell's Scree test is the number found just before the scree of ruble at the foot of a mountain.

Appendix B

Unit Root Tests

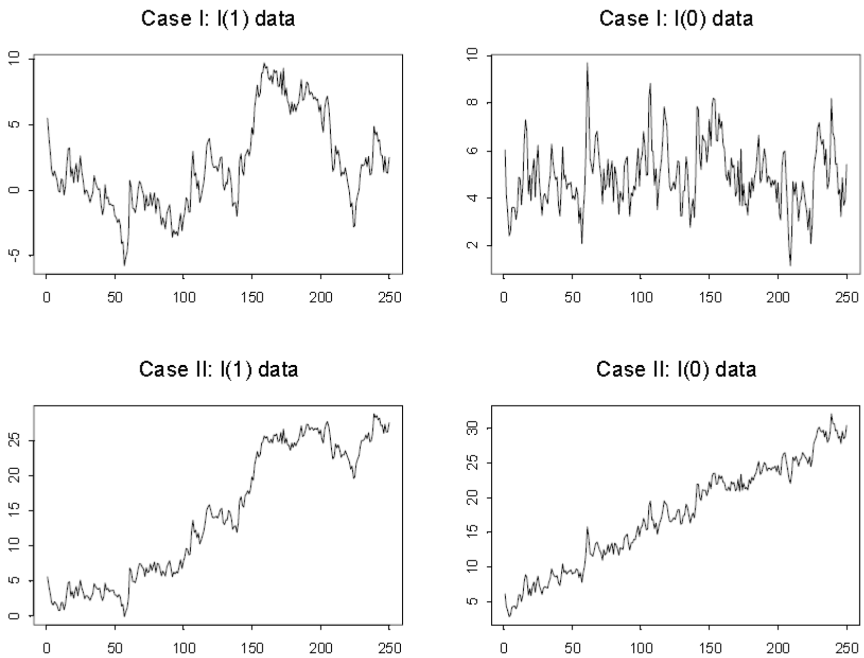


Figure B.1: Different Types of Trend under stationarity $I(0)$ and non-stationarity $I(1)$. This diagram can be used as a visual aid for comparison with the plots of time series in this thesis.

Case I includes a constant to capture the nonzero means and is appropriate for non-trending time series.

Case II includes a constant and deterministic time trend to capture the deterministic trend.

Appendix C

World Health Organisation Standard Values

Age Group	World Average
0-4	8.86
5-9	8.69
10-14	8.6
15-19	8.47
20-24	8.22
25-29	7.93
30-34	7.61
35-39	7.15
40-44	6.59
45-49	6.04
50-54	5.37
55-59	4.55
60-64	3.72
65-69	2.96
70-74	2.21
75-79	1.52
80-84	0.91
85-89	0.44
90-94	0.15
95-99	0.04
100+	0.005

Table C.1: WHO World Standard Values based on world average population between 2000-2025 (values in % - age). These values are used to standardised the mortality rates in Chapter 3