MASTERS DISSERTATION

MORTALITY MODELS: COMPARISON AND APPLICATION IN OLD-AGE POPULATIONS OF SELECTED COUNTRIES

Brian Jin-Wei Hu

Dissertation to be submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science

Johannesburg, 2014

DECLARATION

I declare that this Dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other university.

____

(Signature of Candidate)

_____________ day of ____________________ 2014 in Johannesburg.

ABSTRACT

This research examined which of the five well-known chosen extrapolative mortality models best captured the trends in old-age population mortality for different age groupings in four different countries. Mortality rates from the Human Mortality Database for the United Kingdom, Poland, Japan and Taiwan were used, encompassing males and females in the 65-89 age group. This allowed assessments to be made across developed and emerging economies, and across Europe and Asia. Comparisons were made across models to understand why some work better for some age groupings in some countries. The research considered the goodness-of-fit of these well-known mortality models to historical population mortality rates, assessed the range of projected future mortality rates, and evaluated the financial impact of mortality uncertainty on annuity prices across the subject populations.

Some of the findings which emerged were that the Booth-Maindonald-Smith model tended to work best for most of the selected populations, particularly for female or Asian populations. Perhaps surprisingly, retiring females in the emerging economies can be expected to possibly outlive males in the developed economies selected. In a low yield environment, uncertainty around mortality has a noticeable impact on the range of pricing of annuities. The extent of mortality uncertainty is expected to be less for developed than in emerging economies, and less for females than males.

TABLE OF CONTENTS **PAGE NUMBER**

LIST OF FIGURES

FIGURE **PAGE**

LIST OF TABLES

TABLES **PAGE**

LIST OF SYMBOLS

Calendar year t = from time t to t+1

Log x = Natural logarithm of x to the base e

Logit x = log $\left(\frac{x}{1-x}\right)$

m(t,x) = central crude/unsmoothed death rate for age x in calendar year t

$$= \frac{\text{Number of deaths during calendar year t aged x last birthday}}{\text{Average population during year t, aged x last birthday}}$$

q(t,x) = probability an individual aged x at time t dies between t and t+1

µ(t,x) = probability an individual aged x at time t dies instantaneously

HMD = Human Mortality Database

Models used

LC = Lee-Carter model

RH = Renshaw-Haberman model

CBD = Cairns-Blake-Dowd model (original version before later extensions)

BMS = Booth-Maindonald-Smith model

HU = Hyndman-Ullah model

CHAPTER ONE – INTRODUCTION

1.1 General Introduction

Declining rates of population mortality (namely improving population longevity) is a topic which has become increasingly prominent across the world in recent years. This has featured in various actuarial publications, such as Volume 15 of the British Actuarial Journal (2009). It has also been discussed in various international publications, such as in International Monetary Fund (2012a) and World Economic Forum (2010, 2012).

The forecasting of mortality is an inherently challenging topic, due to the many factors that could be relevant in explaining developments in mortality. Examples are historical trends, socio-economic conditions, medical impairments, and even genetics. Hence, the forecasting of mortality remains an ongoing field of research in actuarial circles.

To date most studies have looked at the improvement in population longevity mainly in developed economies in Europe, North America and Australasia. Examples are Booth, Tickle and Smith (2005) and Booth *et al* (2006), which looked at Australia, Canada and eight other European countries. Cairns *et al* (2007) considered England and Wales, and the United States. Based on the papers examined in the literature review, mortality models could be divided into two regions in terms of their usage and reference: United Kingdom-Europe and Asia Pacific-North America. Studies originating from researchers in a particular country tend to use the prevalent model for that region.

This research considered five mortality models that are well-known in the literature: the Lee-Carter (LC), Renshaw-Haberman (RH), the original Cairns-Blake-Dowd (CBD), Booth-Maindonald-Smith (BMS) and Hyndman-Ullah (HU) models. The RH (Renshaw and Haberman, 2006) and CBD (Cairns, Blake and Dowd, 2006) models were developed by senior researchers based in the United Kingdom. In contrast, the BMS (Booth, Maindonald and Smith, 2002) and HU (Hyndman and Ullah, 2007) models were created by those based in Australia.

Hence, for the purposes of this research, the RH and CBD models were loosely classified in the "UK-Europe" group, and the BMS and HU models in the "Asia Pacific-North America" group of models.

Although the LC model (Lee and Carter, 1992) originated from the United States, due to its early founding and history, it has been adopted internationally and appears to have become the starting model and the basis against which improvements are compared across the world. For example, the projections from the LC model were compared against those from later models in Booth, Tickle and Smith (2005), Booth *et al* (2006), Cairns *et al* (2007), Hyndman and Ullah (2007) and Wang and Liu (2010). Later models were often extensions of or compared against the LC model. For example, the RH model was created as one with a cohort extension to the LC model. The BMS model was another modification of the LC, by selectively fitting only over the years when the assumption of linear improvements in mortality held.

Given that to date, research on the application of mortality models on old-age populations have been in developed economies, this research aims to contribute by applying mortality models in old-age populations of emerging economies, as represented by Poland and Taiwan. The results are compared against applications in developed economies, as represented by the United Kingdom and Japan. In the process, models from the two regions identified are applied across all four countries in the two continents being studied. This allows the applicability and relevance of the models to be compared across countries, geographies and age groupings.

1.2 Purpose and Objectives

The purpose of this research is to find models that describe trends in old-age population mortality rates in selected countries from the developed and emerging economies. Four countries were selected, two of which are from the developed economies and two from emerging. The classifications are consistent with those shown in International Monetary Fund (2012b). Further reasons for selecting them are to understand if there are differences in the patterns of old-age mortality between developed and emerging economies.

The objectives are to

- Examine the trends in old-age population mortality rates in the selected countries for different old-age groupings
- Consider the usefulness of different mortality models in capturing and describing those trends
- Make projections of population mortality rates and illustrate the uncertainty around these projections

The scope of this research is limited to the countries selected. The results and conclusions from this research are therefore unlikely to be representative of all developed and emerging economies in general, because they have different socio-demographic backgrounds and levels of economic development. Wealthier economies may have populations with lower mortality rates and longer life expectancies. For example, the World Health Organisation (2008, p4) found a positive relationship between annual GDP per capita, and life expectancy (albeit at birth) in 169 countries, in 1975 and 2005. A sharp rise in life expectancy occurred when GDP per capita moved towards the $10 000 a year level. Thereafter the rate of increase in life expectancy declined with increasingly higher levels of per capita GDP.

The International Monetary Fund (2012c) calculated the annual GDP per capita in United States (US) dollar ($) terms in 2011 for 185 countries. Of the countries selected in this research, Japan was ranked 18 in the world with a figure of $45 870, the UK ranked closely at 23 with $38 811. Taiwan ranked 40 with $20 083 and Poland ranked 51 with $13 469. Due to their high per capita GDP figures and the proximity of their rankings, Japan and the UK were classified under the developed economies group. Therefore Taiwan and Poland were classified under the emerging economies group. Although not considered by comparison and analysis in this study, as a reference South Africa ranked 72 with $8 078. This research will also examine if there are differences in the patterns of expected population mortality in the four countries selected.

1.3 Hypotheses and Questions

The hypotheses are that different models are required to adequately capture trends in old-age population mortality between developed and emerging economies, because:

1. They are at different stages of economic development on a per capita basis. Old people in developed economies are likely to have accessed (during their working life) and be accessing better-quality healthcare and more advanced medical technologies, resulting in higher old-age life expectancies.
2. As developed economies may already have experienced the phase of fast improving mortality, there may be more room for improvements to continue in emerging economies. Similarly the rate of mortality improvement in developed economies may also have stabilized.
3. The countries selected also have very different ethnic and socio-demographic backgrounds. For example, Japan (Tsuda, 2001) and Taiwan (Ma, 2007) are respectively ethnically homogeneous in their own borders, and similarly for those remaining in Poland (Dustmann, Frattini and Rosso, 2012). In contrast the United Kingdom has a more diverse ethnic composition due to the inflow of immigrants in the twentieth century, as discussed in Bunting (2001), BBC (2005) and Migration Watch UK (2006).

The questions that would be answered are

- How well do the selected models fit the old-age populations in the selected countries?
- What is the extent of uncertainty around mortality in the selected countries?
- What is the financial implication of mortality uncertainty on annuity prices?

1.4 Importance of the topic

Understanding developments in old-age population mortality is important not only for governments with social security and welfare systems, but also for pension funds and insurance companies providing pension benefits. In an insurance context, a better understanding of uncertainty and fluctuations in old-age mortality is likely to be important for actuaries involved in the pricing of annuities and capital allocation and reserving using an internal model under the Solvency Assessment and Management regime in South Africa (Financial Services Board, 2012) and Solvency II regime in Europe (European Insurance and Occupational Pensions Authority, 2011).

This research also aims to raise awareness around trends in population mortality rates outside the developed economies. The scope of this research is however limited to the population mortality rates of countries on the Human Mortality Database. As countries listed there needed to have complete histories of mortality over a fairly long period (usually for at least 40 years), the emerging economies considered tended to be the more affluent ones (as seen in the relatively high GDP per capita numbers in Section 1.2), where improvements in mortality have started taking place. Nevertheless, the framework used in this research can be extended to other countries when longer histories of reliable population mortality data become available.

CHAPTER TWO – LITERATURE REVIEW

2.1 Review of models

2.1.1 Different classes of models: expectation, explanation and extrapolative

According to Booth and Tickle (2008), there are three main classes of models in the field of mortality modelling: expectation, explanation and extrapolative. The paper by Booth and Tickle (2008) forms the basis of this section.

The "expectation" class of models relies on expert opinion and specification of scenarios. According to Booth and Tickle (2008, p9), the advantage of using this class is the potential incorporation of qualitative knowledge around demographics and epidemiology. The disadvantage however, is the inherent subjectivity and inclination to bias by the experts. These traits are generally not regarded as being good for mortality forecasting, particularly as at a population level, experts have historically underestimated mortality improvements. For example, Shaw (2007, p16) considered the life expectancy at birth for both males and females in the United Kingdom. When actual experience was compared against projections made in 1971, 1977, 1981, 1985, 1989, 1991 and 1992, the projections had all underestimated period life expectancy for both males and females. For males, after 20 years since the year of initial projection, the underestimation was around four years. For females, this was around three years.

In the words of Booth and Tickle (2008, p31), the "expectation" class of models is *"generally not a good basis for mortality forecasting, either at the individual or population level".* They elaborated by stating that

> "*Individual expectations are relevant only to the very short-term future and have limited applicability. At the population level, the conservativeness [around decline in mortality] of expert opinion-based targets has been a persistent source of inaccuracy."*

The “explanation” class of models forecasts mortality based on relationships between mortality and medical diseases or risk factors. There is reliance on medical knowledge and information on behavioural and environmental factors. The main advantage of this class is the logical supposition of a link between medical risk factors or conditions, and mortality. The disadvantages are the lack of data containing reliable classifications of cause of death, and an imperfect understanding of links between different risk factors and mortality.

Booth and Tickle (2008) remarked that epidemiological models are used mainly in understanding the “*effect on morbidity and mortality of policy changes affecting the risk factors, rather than in forecasting per se*” (p12). They concluded that the “explanation” approach is *“not useful for overall mortality forecasting”*, mainly because it requires decomposing mortality data by the cause or determinant of death. They continued by stating that

> *“causal models are not widely used because of a lack of sufficient data on the determinants … Furthermore, the difficulty (at current levels of knowledge) in identifying cause of death at the advanced (and increasing) ages at which most deaths occur severely limits the usefulness of the explanatory approach”*. (p31 - 32)

The disadvantages with this approach were cited in other sources too. McNown and Rogers (1992) concluded using US male and female mortality data from 1960 to 1985 that “*disaggregation by cause of death has produced little or no gain in the accuracy of forecasting total mortality*”. The (United Kingdom) Government Actuary’s Department (2001) also recommended against cause-of-death decomposition.

The “extrapolative” class of models is quantitative in nature, which can be expanded to provide stochastic mortality projections. These models take historical trends into account in projecting into the future. They commonly take the form of mortality being explained by factors relating to age, time period and cohort, plus an error term. The Lee-Carter model (Lee and Carter, 1992) is an example of a two-factor extrapolative model (with the factors being age and time period). The Renshaw-Haberman model would be an example of a three-factor model (Renshaw and Haberman, 2006), with the third factor being cohort.

Booth and Tickle (2008) observed that most research has focused on looking at two- and three-factor models, as zero- and one-factor models were generally inadequate and led to parameters which lacked an intuitive interpretation. This was a result of trying to explain mortality trends with too few parameters. An example of a zero-factor model was given in Bell (1997), which applied a simple random walk with drift model to the logged age-specific mortality rates of white US males and females. Booth and Tickle (2008, p32) noted that zero-factor models "*provide no information about changes in the age pattern, and the independent forecasting of age-specific rates may produce irregular and implausible age patterns*". The one-factor models suffer from having to explain changes in mortality with just one factor, which may result in that factor losing an intuitive meaning. In comparison, the two- and three-factor models have the freedom to use intuitive factors like age, time period or cohort to explain mortality development.

Applying these models in the 21st Century means using a computing package to fit parameters, after taking amongst others, mortality trends, age structure, time period and cohorts of the underlying dataset into account. Once tests such as goodness-of-fit have been performed, forecasts of mortality can then be made.

Whilst the 'extrapolative' class of models presents many advantages, by removing qualitative overlays such as expert judgment and ignoring relationships with medical causes, this class of models is dependent on the quality and appropriateness of the input data. The type of model and parameters specified will affect the quality of projections from this class of models.

Whilst each class of models has its advantages and disadvantages, on balance, the 'extrapolative' class of models presents more advantages than disadvantages. Hence, this research only considered the "extrapolative" class of models.

2.1.2 Lee-Carter (LC) model

In this research, five mortality models are considered. The Lee-Carter model is considered first, as it was the earliest published in 1992, calibrated using mortality data of the United States over the period 1933 to 1987. Since then it has become widely used, and formed the foundation for later models published by Booth, Maindonald and Smith (2002) and Renshaw and Haberman (2006).

The model described by Lee and Carter (1992) postulates that the mortality rate at a particular age x in a particular year t can be explained by factors relating to the age of the individual and year of interest. Hence, it can be expressed in the following form:

$$\text{Log } m(t,x) = A_x^{(1)} + A_x^{(2)}P_t^{(2)} + E(t,x) \qquad (2.1)$$

Where

$m(t,x)$ is the death rate for age x in calendar year t

$A_x^{(1)}$ is the value of the “age effect” parameter at age x

$A_x^{(2)}$ is the value of the “age interaction” parameter at age x

$P_t^{(2)}$ is the value of the “time interaction” parameter in year t

$E(t,x)$ is the value of the error term at age x and year t

The natural logarithm of mortality rates is used to ensure that the projected mortality rates are not negative.

To avoid potential problems with identification of the model, Lee and Carter (1992, p661) defined the following “normalising” constraints for their model:

$$\sum_t P_t^{(2)} = 0 \text{ and}$$

$$\sum_x A_x^{(2)} = 1$$

The first constraint on $P_t^{(2)}$ implies that for each age x, the estimated $A_x^{(1)}$ would be the average log m(t,x) over time for that age x. Cairns *et al* (2007, p13) stated that there was no natural choice for the second constraint on $A_x^{(2)}$, and different choices have appeared in the academic literature on the application of the LC model. Importantly though, in their view, the choice of this constraint had no impact on the quality of fit or forecasts of mortality in any case.

The model effectively postulates that the log mortality rate for a person aged x last birthday at time t, can be explained by a term related to age x ($A_x^{(1)}$), and another by the interaction between age x and time t ($A_x^{(2)}P_t^{(2)}$).

$A_x^{(1)}$ can thus be interpreted as the average log mortality over time, $P_t^{(2)}$ the overall level of mortality in year t and $A_x^{(2)}$ the sensitivity at age x to changes in mortality over time.

2.1.3 Renshaw-Haberman (RH) model

Renshaw and Haberman (2006) extended the LC model by adding a cohort effect, applied to data in England and Wales over the period 1961 to 2003. In other words the year c (where c = t-x) in which an individual was born was also thought to be important. The RH model is presented in (2.2):

$$\text{Log } m(t,x) = A_x^{(1)} + A_x^{(2)}P_t^{(2)} + A_x^{(3)}C_c^{(3)} + E(t,x) \qquad (2.2)$$

Where
m(t,x) is the death rate for age x in calendar year t
$A_x^{(1)}$ is the value of the “age effect” parameter at age x
$A_x^{(2)}$ is the value of the “age interaction” parameter with time at age x
$P_t^{(2)}$ is the value of the “time interaction” parameter in year t
$A_x^{(3)}$ is the value of the “age interaction” parameter with cohort at age x
$C_c^{(3)}$ is the value of the “cohort interaction” parameter in cohort year c, where c = t-x
E(t,x) is the value of the error term at age x and year t

As for the LC model, constraints were applied to avoid identification problems for the RH model. Renshaw and Haberman (2006, p562) and more clearly, Cairns *et al* (2007, p14) used the following constraints:

$$\sum_{t} P_t^{(2)} = 0$$

$$\sum_{x} A_x^{(2)} = 1$$

$$\sum_{c} C_{t-x}^{(3)} = 0 \text{ (where c = t-x) and}$$

$$\sum_{x} A_x^{(3)} = 1$$

The first and third constraints on $P_t^{(2)}$ and $C_c^{(3)}$ imply that for each age x, the estimated $A_x^{(1)}$ would also be the average log m(t,x) over time for that age x. Cairns *et al* (2007) stated again that there are no natural choices for the second and fourth constraints on $A_x^{(2)}$ and $A_x^{(3)}$. This was similar to their comments about the lack of a natural constraint on $A_x^{(2)}$ in the LC model. Importantly, the choice of constraint had no impact on the quality of fit in any case. Hence, for stylistic consistency, $A_x^{(2)}$ and $A_x^{(3)}$ were constrained to sum to 1 in the RH model.

This model postulates that the log mortality rate for a person aged x last birthday at time t, can be explained by a term related to age x ($A_x^{(1)}$), by the addition of an interaction term between age x and time t ($A_x^{(2)}P_t^{(2)}$), and by the addition of another interaction term between age x and cohort t-x ($A_x^{(3)}C_{t-x}^{(3)}$).

Booth and Tickle (2008, p21) noted that although in theory, adding a cohort term should improve forecasting, in practice using this model requires many more years of data to allow for complete cohorts. If an entire age range across the population is being studied, data covering a century or so would give only one complete cohort, and a much longer series of annual data would be needed to produce forecasts. The bigger the age range of interest, the more years of data are required for the analysis. Even when the data are available, results may depend heavily on the experience of cohorts born in the nineteenth century, which may not be appropriate (Tabeau *et al*, 2001). These problems around data availability and applicability are reduced when the age range of interest is more restricted (Booth and Tickle, 2008). This is another reason why this research focuses only on the 65 – 89 age group.

Analogous to the LC model, $A_x^{(1)}$ can thus be interpreted as the average log mortality over time, $P_t^{(2)}$ the overall level of mortality in year t, $A_x^{(2)}$ the sensitivity at age x to changes in mortality over time, $C_c^{(3)}$ the overall level of mortality for the cohort born in year (t-x), and $A_x^{(3)}$ the sensitivity at age x to changes in cohort mortality.

2.1.4 Cairns-Blake-Dowd (CBD) model

Different from the LC framework, Cairns, Blake and Dowd (2006) looked at the logarithm of the ratio of the mortality rate to the survival rate. This was different from considering the natural logarithm of the mortality rate, used in the LC and RH models. The CBD model postulates that this ratio can be described by a parameter related to the year of interest t, and the interaction between another year-related parameter, and the deviation of the age of the individual x from the average age in the population. The model was applied to English and Welsh data over the period 1961 to 2004, is presented as:

$$\begin{aligned}\text{Logit } q(t,x) &= P_t^{(1)} + A_x^{(2)} P_t^{(2)} + E(t,x) \\ &= P_t^{(1)} + (x - \bar{x}) P_t^{(2)} + E(t,x) \qquad (2.3)\end{aligned}$$

Where
$P_t^{(1)}$ is the value of the “year effect” parameter in year t
$P_t^{(2)}$ is the value of the “time interaction” parameter with age x in year t
$\bar{x}$ is the mean age in the sample
$A_x^{(2)}$ takes on the value $(x - \bar{x})$
E(t,x) is the value of the error term at age x and year t

According to Cairns *et al* (2007, p15) and also Li and Chan (2011, p5), there are no identifiability problems with this model, and the following assumptions were made by Cairns *et al* (2007, p3):

1. For integers t and x, and for all $s \geq 0$, $u < 1$, $\mu(t+s,x+u) = \mu(t,x)$, where $\mu(t,x)$ is the force of mortality, the instantaneous death rate at exact time t for an individual aged x at time t

In other words, the force of mortality remains the same up to one calendar year and up to one integer age, before changing

2. The size of the population at all ages remains constant over time. In other words, in the context of this research, it is assumed that the number of deaths, immigrants and emigrants in subsequent years cancel each other out, resulting in an old-age population the size of which does not change.

Cairns *et al* (2007, p3) did remark that these two assumptions “*do not normally hold exactly, but the resulting relationship between m(t,x) and q(t,x) is generally felt to provide an accurate approximation*”.

Nevertheless, these two assumptions imply the following relationships:

a) $m(t,x) = \mu(t,x)$
b) $q(t,x) = 1 – \exp[-\mu(t,x)]$
$\quad\quad = 1 – \exp[-m(t,x)]$ (from a))

Relationship a) has also been used in the analysis of death rate data (Brouhns, Denuit and Vermunt, 2002). Relationship b) between q(t,x) and μ(t,x) could also be found in the formulae book published by the Faculty of Actuaries and Institute of Actuaries (2002, p32).

From a) and b), a relationship is established between m(t,x) and q(t,x). This is useful, because whilst the CBD model produces q(t,x) as its output, the other four models produce m(t,x). This relationship allows the q(t,x) rates from the CBD model to be converted into m(t,x), thus allowing comparability with the projections from the other four models.

For the record, Cairns *et al* (2007) did discuss more advanced versions of the CBD model. Compared to the original version presented in Equation 2.3, more advanced versions had an extra cohort term (for example $(x_c – x)C_C^{(3)}$) added to the model, or had a quadratic term related to the age effect (for example $((x – \bar{x})^2 – m_x^2)P_t^{(3)}$, where m_x^2 is the mean of $(x – \bar{x})^2$) added to the model.

Given that the purpose of this research is to gain an understanding of the different families of models as a starting point, it was decided that only the original CBD model would be considered. More advanced CBD models are not considered in this research.

2.1.5 Booth-Maindonald-Smith (BMS) model

With the LC model, terms $A_x^{(1)}$ and $A_x^{(2)}$ are assumed to be invariant over time and term $P_t^{(2)}$ is assumed to be linear over time (implying a constant rate of mortality decline over time). These assumptions were challenged by Booth, Maindonald and Smith (2002). As described in Booth *et al* (2006), the BMS model built on and improved the LC model. This was by using more interaction terms between age x and year t, and by shortening the fitting period to when the assumptions of constant A_x and linear P_t were better met. Whereas the LC model only used the first terms of the singular value decomposition, the BMS model used “n” terms, allowing second and higher order terms to be used as well. In the words of Booth, Maindonald and Smith (2002), “any systematic variation in the residuals from fitting only the first term would be captured by the second and higher order terms”. The BMS model was applied to Australian data over the period 1907 to 1999 and takes the form described in (2.4):

$$\text{Log } m(t,x) = A_x^{(0)} + \sum_{i=1}^{n} A_x^{(i)} P_t^{(i)} + E(t,x) \qquad (2.4)$$

Where
m(t,x) is the death rate for age x in calendar year t
$A_x^{(0)}$ is the value of the “age effect” parameter at age x
$A_x^{(i)}$ is the value of the “age interaction” parameter at age x
$P_t^{(i)}$ is the value of the “time interaction” parameter in year t
E(t,x) is value of the error term at age x and year t
n is the rank of the approximation
By using more interaction terms, the BMS model aims to improve the fit to the data and account for previously unexplained effects under the LC model.

According to Booth, Maindonald and Smith (2002), when applied to Australian data, the BMS model produced higher forecasted life expectancies relative to the

LC model, and had half of the forecast error. In Booth *et al* (2006), which considered the mortality rates of ten developed countries, the BMS model again provided more accurate forecasts than the LC model, which was attributable to the shorter fitting period used by the BMS model. By working with a shorter fitting period, during which the assumptions of constant A_x and linear P_t were better met, the BMS model was able to achieve a better fitting result than the LC model. The LC model used a longer fitting period, during which these assumptions were met not as well.

2.1.6 Hyndman-Ullah (HU) model

Hyndman and Ullah (2007) took a more computational approach in their model that was applied to French mortality data over the period 1899 to 2001, where factors are not directly attributable to age, period or cohort. Their extension of the LC model is applied to data smoothed using non-parametric methods:

$$\text{Log } m(t,x) = f(t,x) + \sigma(t,x)E(t,x) \qquad (2.5a)$$

$$= \mu(x) + \sum_{i=1}^{n} B(t,i)D(x,i) + E_2(t,x) + \sigma(t,x)E(t,x)$$

$$= \mu(x) + \sum_{i=1}^{n} B(t,i)D(x,i) + \sigma(t,x)E(t,x) + E_2(t,x) \qquad (2.5b)$$

The HU model started off as modelling log mortality rates using a function f(t,x) that is observed with error, as per 2.5a. E(t,x) follows a standard normal distribution, and σ(t,x) allows the amount of noise to change with age (Hyndman and Ullah, 2007, p4943). Function f(t,x) is then developed into a polynomial involving coefficients B(t,i) and orthonormal basis functions D(x,i), with another error term $E_2(t,x)$ that follows a normal distribution with mean 0 (p4944). The terms are re-ordered to arrive at (2.5b) representing the model.

From Booth *et al* (2006, p294), the terms in (2.5b) are interpreted as follows:
μ(x) is the average mortality at age x over time, estimated by applying penalised regression splines to each year of data and taking the average. A penalised regression spline method involves fitting a smooth curve to a scatter of data points.

B(t,i) is a time series coefficient, and D(x,i) a basis function, both estimated using principal component decomposition. A principal component decomposition method identifies the relevant factors that explain the behaviour of a series of data points, with the factors being presented in decreasing order of importance.
n is the number of basis functions used.
The term σ(t,x)E(t,x) represents observational error varying with age, which is the difference between the observed mortality rates (log m(t,x)) and those projected from spline curves (f(t,x)).
E_2(t,x) is the modelling error, which is the difference between the spline curves and fitted curves from the model.

From Hyndman and Ullah (2007), when the HU model was applied to French mortality and Australian fertility data, the forecasts obtained were better than from the LC model and its variants. In Booth *et al* (2006), on average the HU model was found to provide more accurate forecasts of log death rates than the LC and BMS models, although not by a large margin. The LC model tended to underestimate mortality rates (with associated p-values being no more than 0.03), particularly of females. In contrast, the BMS and HU models overestimated male and underestimated female mortality rates, but at levels that were not significant (with p-values being greater than 0.09). To put this in context, the average absolute error across the ten countries studied was calculated for males and females, with the unit of calculation being the absolute difference between two log death rates. For males, the error was 0.31 for the LC, 0.15 for the BMS and HU models. For females, the error was 0.45 for the LC, 0.16 for the BMS and 0.15 for the HU model.

2.2 Review of research by country

For this section, only publications available in English of studies on the mortality rates in the four countries (the United Kingdom, Poland, Japan and Taiwan) have been considered. This would cover original publications in English, or translations thereof. Publications available in Polish, Japanese or Mandarin but not in English, were not covered in this literature review.

2.2.1 United Kingdom

In the study done by the (United Kingdom) Office for National Statistics (2012a), mortality rates for the UK in each calendar year from 1961 to 2009 were smoothed to remove fluctuations from age to age and from year to year. A p-spline model, described in Currie, Durban and Eilers (2004), was then applied to produce a fitted and smoothed mortality surface to the historical data for each gender. Currie (2006) provided an example of how the p-spline method works. It is based on a penalised generalised linear model. The spline method fits through a scatter of data points using a smooth curve, and forecasts made are a function of the smoothing process. The period expectation of life at age 65 had increased steadily for males and females between 1911 and 2010. The rate of increase accelerated after 1971 for both genders. After 1991, males began to experience faster rates of improvements in life expectancy than females. However, when sub-divided into geographies, mortality for Scottish males and females at some ages had improved more slowly (or even worsened) compared to the rest of the UK. Period expectation of life at age 65 was then projected from 2010 to 2085. The rate of improvement in years was expected to be broadly linear at the same pace for both males and females. At age 65, females were generally expected to live longer than males by between two to three years.

Canudas-Romo (2010) showed that between the periods 1980-82, and 2008-10, the median age of death for UK females had increased from 80 to 85. Considering other measures, such as the average and modal age of death, these were found to have increased from 77 to 82, and 84 to 88 respectively.

Cairns *et al* (2007) used various models such as the LC, RH, CBD and its extensions in projecting old-age male mortality for England and Wales. Extensions to the CBD model were introduced, which involved adding extra terms to capture the cohort effect, or to introduce a quadratic term into the age effect. Ages 60 to 89 were considered based on data from 1961 to 2004. The paper concluded that extensions to the CBD model provided the best fit to the historical data, and resulted in more stable parameter estimates. The paper remarked that mortality rates had effectively improved linearly over time at all old ages, improvements had been greater at younger older ages, and there was a significant cohort effect to improvements.

2.2.2 Poland

Matysiak and Nowok (2007) looked at making population projections for Poland across all ages using the cohort-component model, taking factors such as fertility and net migration into account. The authors relied on historical data to formulate point forecasts of age-specific mortality rates. Forecast errors derived from historical data were adjusted to the empirical specification of the error structure estimated by Alho and Spencer (2005). Forecasts were then adjusted by expert judgement. The model was not calibrated or designed to look at old-age mortality projections specifically. The results shown were projected total population sizes, median age and old-age dependency ratios, rather than projected patterns of mortality for old-age populations. Hence although the literature found on Poland was not particularly relevant for this research, the results showed that further population ageing in Poland was expected to continue.

2.2.3 Japan

Ozeki (2005) looked at fitting mortality models to historical male Japanese life tables. The models considered were the Heligman-Pollard, Mixed Weibull, Lee-Carter and Ozeki’s own simulation model which was based on the “Vitality” concept proposed by Furukawa (1996). This paper remarked that the Mixed Weibull model achieved better fits than the Heligman-Pollard and simulation models. It was not clear if the Mixed Weibull was better than the Lee-Carter model, as the paper did not apply statistical measures to quantify the goodness-of-fit of each model. Goodness-of-fit was evaluated visually (based on how close forecasted mortality rates were to simple projections of historical mortality rates) for each model separately. Quantitative comparisons across models were not made, and no stochastic projection of mortality with confidence intervals was made. Hence, similar to Poland, the literature found for Japan was not particularly relevant for this research.

2.2.4 Taiwan

Wang and Liu (2010) considered fitting four stochastic mortality models to Taiwanese male and female mortality data from 1970 to 2000. The models were the LC, Lee-Miller, BMS and HU models. Goodness-of-fit was assessed by considering which model gave the lowest measures under five criteria: mean error, mean absolute error, mean squared error, mean percentage error and mean absolute percentage error. The authors used the period from 1970 to 2007, and divided that between the fitting and forecast periods. When the period over which the model was fitted was long (of 31 years from 1970 to 2000), the BMS model achieved better fits for males and the LC model for females, albeit over a short forecast period of only 7 years. However, when the period of fitting was shorter (from 1970 to 1990, or from 1970 to 1980), the HU model achieved the best fit for both genders, over a longer forecast period of 17 or 27 years. Whilst the findings were interesting, the study considered the male and female population across all ages (with ages above 95 grouped). This differs from the focus of this research, which is in the 65 to 89 age group.

2.2.5 Comparison across countries

With the exception of the United Kingdom, literature on mortality projections for old-age populations has been limited. What was found for Poland and Japan were not directed to this research. The results found for Taiwan were of use, but it considered the entire population across the full age spectrum, not just at the old ages. The study by Cairns *et al* (2007) was the most relevant. Importantly, it noted the existence of a cohort effect, and observed that mortality improvements had been linear. The UK section of this research expands on Cairns *et al* (2007) by considering the entire UK population beyond just the male population in England and Wales, and bringing the BMS and HU models into the comparison.

CHAPTER THREE – METHODOLOGY

3.1 Sources of data

This research made use of data from the Human Mortality Database (HMD), which can be found at www.mortality.org. This is a comprehensive database containing population mortality data across many countries, and it continues to expand its coverage. As at August 2012, there were 37 countries on the database. The underlying data were collected and compiled predominantly by the relevant national statistical agency of that country, which lend credibility to the quality of the data. This database has been used widely.

As examples, Booth, Tickle and Smith (2005), Booth *et al* (2006) and Shang, Booth and Hyndman (2011) used data from over ten countries downloaded from the HMD. Scandinavian data from the HMD were used in Li and Chan (2005), Koissi, Shapiro and Högnäs (2006) and Hyndman, Booth and Yasmeen (2011). Data for England and Wales were used in Li (2010) and Sweeting (2011). Data for the United States and Canada were used in Li, Hardy and Tan (2009, 2010), Li, Chan and Cheung (2011) and Wang *et al* (2010). Wang and Liu (2010) used data for Taiwan.

For this research, data from the United Kingdom, Poland, Japan and Taiwan were downloaded from the HMD in March 2012. The number of years used in this study (up to the year 2009) for the UK and Japan was 63, for Poland it was 52, and for Taiwan it was 40. The HMD provides period mortality rates. For illustration, this means that the mortality rate for age 65 in 2008 relates to those born in 1943, and that in 2009 relates to those born in 1944. The mortality rates analysed were for males and females at a total country level, in three age sub-groups of 65-69, 75-79 and 85-89, and in the 65-89 age group.

The format of the period mortality data is shown in Appendix 1.

These countries were chosen so as to bring emerging economies (Poland and Taiwan) into the analysis along with developed economies (United Kingdom and

Japan), as well as introduce geographical diversity (two countries from each of Europe and Asia).

Unfortunately mortality rates of African countries (including South Africa) were not available on the HMD. Efforts were made to see if data at the required level of detail were available for South Africa from other sources. However, no continuous history of data over a long enough period (of at least 10 years until 2005) was available at a national level for the entire population of South Africa. The main problem has been the lack of reliable data for black South Africans. In contrast, for the four countries selected, data were available at a national level for at least 40 years to 2009. A length of 40 years was deemed necessary to conduct more meaningful out-of-sample testing of the mortality model fitted in this research. For this reason South Africa was not included in this research.

The reason for looking at three different old-age sub-groups is to examine if there are differences in the patterns of mortality across different old-age groups. For example, mortality at extremely old ages (85-89) may behave differently from mortality at “younger” old ages (65-69). Different mortality models may therefore work better in different age sub-groups.

Age 65 was chosen as the starting point as that tends to be a common reference point for retirement. For example, the population mortality study by the UK’s Office for National Statistics (2012a) used age 65 as the reference point for projecting life expectancy in old age, although the default retirement age in the UK has been abolished, according to the BBC (2011). The BBC (2012) reported that the current retirement age for men in Poland is age 65, but will be raised to 67. Warnock (2012) reported on legislation in Japan effectively raising the mandatory retirement age from 60 to 65. The China Post (2008) reported that the revised Labor Standards Law in Taiwan raised the retirement age for insured workers from 60 to 65.

Age 90 was chosen as the cutoff point of analysis. This is because beyond this point, the size of the population becomes noticeably smaller, lending less credibility to the analysis. Also, problems with inaccurate age reporting and age heaping may also start to take place. For example, A’Hearn, Baten and Crayen

(2006) mentioned that "*age data frequently display excess frequencies at round or attractive ages, such as even numbers and multiples of five*". Hence, for this research, five-year (instead of single-year) age bands were used to mitigate these potential problems, so that any significant trend in mortality seen is more likely to be prominent.

3.2 Computing language

Application of the five mortality models discussed in 2.1 was done in the R computing language, and the numerical and graphical summaries of results were done in Microsoft Excel. The following R programme modules were used to run the models:

- Lifemetrics toolkit, with code and user guide written by Cairns (2007).
- "Demography" package for R, with code and user guide written by Hyndman (2011a)
- "Forecast" package for R, with code and user guide written by Hyndman (2011b)

Additional code in Visual Basic was written by myself specifically to convert the raw data from the HMD into the format required by these R-based programme modules.

The code written by Cairns (2007) in the Lifemetrics toolkit was used to fit the LC, RH and CBD models. Professor Cairns is also one of the developers of the Cairns-Blake-Dowd mortality models, which adds a degree of reliability to the quality of the code. To check that his code was being used correctly for this research, the code was run on the same datasets used in Cairns *et al* (2007), for England and Wales and the United States, which were included with the code. Results generated around parameter estimates and Bayes' Information Criteria for the models matched those published in Cairns *et al* (2007). This replication suggested that the code was being used correctly.

The "Demography" package written by Hyndman (2011a) was used to fit the BMS and HU models. Professor Hyndman is the co-author of the paper by Booth *et al* (2006), and is one of the developers of the HU model, which again adds a degree of reliability to the quality of the packages. Instructions in Hyndman (2011a) were followed to run the package.

Once the parameters in the five models had been fitted, the “Forecast” package (also written by Professor Hyndman) was used to make projections for all five models for consistency. Instructions in Hyndman (2011b) were followed to run the package.

3.3 Assessing goodness-of-fit

For each of the three age sub-groups (65-69, 75-79, 85-89) and the 65-89 age group in each gender in each of the four countries, the following scenario analyses were done:

Scenario A: Analysis of model-fitted mortality rates over the full period of data available for that country, compared to actual period mortality rates (in-sample testing)

Scenario B: Analysis of model-fitted mortality rates over a subset of the full period, leaving 20 years (from 1990 to 2009) available to compare fitted against actual period mortality (out-of-sample testing)

Scenario C: Analysis of model-fitted mortality rates for a cohort aged 67 in 1990, compared to actual cohort mortality rates to 2009 (another out-of-sample testing)

In scenario A, the purpose was to assess how well each of the five models could capture the pattern of historical period mortality rates, over the full period of data availability.

Having gone past scenario A, scenario B intended to assess how closely projected mortality rates compared with actual period mortality rates in the ensuing 20 years (from 1990 to 2009) after the fitting period. According to Booth, Maindonald and Smith (2002, p335), the maximum length of the forecast period should be no longer than the fitting period. As the country with the shortest history of data had 40 years, 20 years was chosen as the forecast period, which was no shorter than the length of the fitting period. Choosing a common 20 years for forecasting across the four countries also introduced consistency into the study.

Scenario C expands on Scenario B by considering the projection of cohort mortality rates at age 67 in 1990. Age 67 was chosen, as that represented the mid-point of the 65-69 age sub-group. The 65-69 age sub-group was regarded as the most important, as it is the largest five-year sub-group by the number of people, and is most representative of recently-retired people considering the purchase of annuity products.

To date, the mortality rates considered were period-based. Cohort mortality differs from period mortality, as cohort mortality tracks the survivorship of a group of people born in the same year over some period. In contrast period mortality considers the survivorship over one year, and then moves on to a younger cohort born in the following year. For example, suppose as a start, the cohort and period mortality rate of age 67 (those born in 1923) in 1990 are the same. The cohort mortality of age 67 in 1991 is the mortality rate at age 68 of those born in 1923. In contrast, the period mortality of age 67 in 1991 considers those born in 1924, not 1923. In other words, cohort mortality is people-dependent, and period mortality is time-dependent, with more details available in Appendix 2. Cohort mortality is more useful when tracking the mortality development of a group of people buying an annuity. For this reason a good model should also perform well under Scenario C.

In the full period of analysis (Scenario A), the goodness-of-fit criteria used to assess model appropriateness were which model

- maximised the Bayes' Information Criterion (where the maximum likelihood is reduced by the number of parameters used),
- generated the smallest residuals (defined as the differences between actual period and fitted mortality rates) on average, and if possible also
- generated the smallest absolute residuals (defined as the absolute differences between actual period and fitted mortality rates) on average

In the sub-period of analysis (Scenario B), only the residuals and absolute residuals were considered.

In Scenario C the cohort mortality projections from the models were compared graphically against actual mortality rates.

The three goodness-of-fit criteria are discussed in more detail:

1. Bayes' Information Criterion (BIC)
2. Residuals
3. Absolute residuals

3.3.1 Bayes' Information Criterion (BIC)

Cairns *et al* (2007) defined the BIC to be of the following form:

$$\text{BIC} = \text{ll} - 0.5 \times p \times \log N \qquad (3.1)$$

where
ll is the log-likelihood function for the mortality model
p is the effective number of parameters being estimated
log is the natural logarithm with base e
N is the number of observations

The best-fitting model would have the highest (or least negative) BIC measure.

In estimating the ll, firstly the following notation was used:
D(t,x) = number of deaths in year t recorded as aged x last birthday
E(t,x) = size of population in year t aged x last birthday
m(t,x) = mortality rate for person in year t aged x last birthday

It was assumed in Cairns *et al* (2007) that D(t,x) followed a Poisson distribution, with mean E(t,x) x m(t,x).

Next the derivation of the ll is shown.

The general probability density function for a Poisson distribution takes the form of:

$$p(x) = \frac{e^{-\mu}\mu^{x}}{x!} \quad , x = 0, 1, 2, \ldots \qquad (3.2)$$

$$= 0 \qquad \text{elsewhere}$$

where x is the variable of interest, and μ is the mean and parameter of x which is greater than 0.

The likelihood function would then be

$$\prod_{x,t} \frac{e^{-\mu}\mu^{x}}{x!} \qquad (3.3)$$

Now let x represent D(t,x), and μ becomes E(t,x)m(t,x) based on the assumption in Cairns *et al* (2007).

The log-likelihood function is then

$$\ln \prod_{x,t} \frac{e^{-\mu}\mu^{x}}{x!}$$

$$= \sum_{x,t} \ln\left(\frac{e^{-E(t,x)m(t,x)} E(t,x)m(t,x)^{D(t,x)}}{D(t,x)!} \right)$$

$$= \sum_{x,t} \left[D(t,x)\ln[E(t,x)m(t,x)] - E(t,x)m(t,x) - \ln D(t,x)! \right] \qquad (3.4)$$

Which corresponds to the equation shown on p11 of Cairns *et al* (2007).

The BIC measure is appropriate across models that use the same number of years in the fitting process. As discussed in 2.1.5, the BMS model fits selectively only over years that meet the linearity of mortality decline assumption. Therefore, it would be expected to have the least negative log-likelihood function and best BIC measure.

Hence, for this research, the BIC would only be used as an indicator of the goodness of fit, rather than as a conclusive selection criterion across models. To compare across all five models, alternative assessments of the goodness-of-fit were used, namely the residuals and absolute residuals.

3.3.2 Residuals

The residual is defined as the average difference between fitted and actual mortality, over the years and age groupings considered, which is

$$\overline{d} = \frac{1}{mn}\sum_{x}\sum_{t}[\hat{m}(t,x) - m(t,x)] \qquad (3.5)$$

where

x represents age, and there are m ages in total

t represents year, and there are n years in total

The residual is not necessarily expected to converge to zero, because the models considered were not designed to over-fit the mortality data in order to deliver a result of zero residual. To start off, the model is fitted either on the natural logarithm or the logit function of the mortality rate, not on the mortality rate itself. Also, the fitting process is not the same as an ordinary regression, because there are no known regression variables to regress against. The parameters in the models are not a series of ages or years that can be regressed against. Lastly, out-of-sample tests are done (as described in Scenario B in 3.3), which presents a mismatch between the earlier fitting period and the later forecast period, which makes it highly unlikely that the residual in the forecast period will converge to zero.

Assuming that the residual follows a Normal distribution, the following metric was used to conduct tests on the statistical significance of fit:

$$t = \frac{\overline{d}}{S_d / \sqrt{mn}} \qquad (3.6)$$

where

s_d = standard deviation of $\overline{d}$, the difference between fitted and actual mortality across ages and years
m and n as defined in 3.5.

This is basically a paired t-test which follows a t distribution with (mn – 1) degrees of freedom (Hogg and Tanis, 2001, p442). Having a statistical distribution underpin allows p-values to be calculated. Here, the null hypothesis is that the residuals generated from the model are not significantly different from zero. The alternative hypothesis is that they are significantly different from zero. With a good model, the null hypothesis would fail to be rejected, with the result being statistically “insignificant”. The 5% level was chosen as the level of insignificance. A good model should thus have a low value for the residuals, a statistically “insignificant” result, and a high p-value (of at least above 5%).

3.3.3 Absolute residuals

If only residuals were considered, then a model whose forecasts under- and over-shoot actual rates may still pass as a good model, provided that the positive and negative residuals cancel each other out. For this reason, absolute residuals are introduced. This criterion aims to reward models whose forecasts are not far from actual rates in either direction, above or below. Absolute residuals were also used as a criterion (in addition to residuals) in Booth *et al* (2006).

The absolute residual is defined as

$$\overline{d_A} = \frac{1}{mn}\sum_x \sum_t |\hat{m}(t,x) - m(t,x)| \qquad (3.7)$$

where
x represents age, and there are m ages in total
t represents year, and there are n years in total

It is also assumed that the absolute residual follows a Normal distribution. A similar t-statistic to that defined in 3.6 is used to calculate the p-value, which is

$$t = \frac{\overline{d_A}}{S_{dA} / \sqrt{mn}} \qquad (3.8)$$

where

s_{dA} = standard deviation of $\overline{d}_A$, the absolute difference between fitted and actual mortality across ages and years

As in section 3.3.2, a good model should thus have a low value for the absolute residuals, a statistically “insignificant” result and a high p-value.

3.4 Selecting a model

It is logical that a model based on historical data will fit to the historical dataset better than forecast beyond the dataset. Thus, in selecting a model, more weight is placed on a model’s forecasting than fitting performance. The final model is selected based on the following model-selection criteria (in descending order of preference):

1. Performance of the model in the out-of-sample forecasting (under Scenario B of 3.3). Over the period 1990 to 2009, were the model’s residuals and absolute residuals associated with p-values greater than 5% over different age groupings?
2. Performance of the model in forecasting mortality rates for a cohort aged 67 in 1990 (Scenario C of 3.3), based on fitting over the preceding data sub-period. How close were the projections to actual cohort mortality rates from 1990 to 2009?
3. Performance of the model in the in-sample testing over the full period of data availability (under Scenario A of 3.3). Within this period more reliance was placed on residuals and absolute residuals, with the BIC used only as an indicator, for reasons discussed in 3.3.1.

If more than one model did well under criterion 1, then criterion 2 would be considered. If no clear choice emerged by then, then criterion 3 would be brought into consideration.

Although a model’s in-sample performance was regarded as less important than out-of-sample, for the purposes of flow and logical development, Chapters 4 to 7 still began by considering model performance over the full data period (Scenario A of 3.3) before moving onto sub-periods (Scenario B) and cohort mortality projections (Scenario C).

3.5 Application of selected model

Once a model has been selected for a gender in a country, and after fitting the model over the full period of data available for that country, mortality projections for a cohort aged 67 in 2010 (after 2009, the last year when historical data from the HMD were available) were made for the remaining 20 years to 2029.. This was to gauge the development of mortality beyond the dataset provided by the HMD. In assessing the likely development of mortality for a group of people buying an annuity, cohort mortality is more realistic than period (with more details in Appendix 2), and more useful to consider for pricing and risk reserving purposes. O’Connell (2011) commented that period mortality should be used to “*summarise the level of mortality in a population in one period*”, whereas cohort mortality should be used to estimate the “*lifespan a person of a defined cohort might expect*”. Historically, life expectancy calculated using period mortality could have underestimated the more accurate life expectancy calculated using cohort mortality.

In the cohort mortality projections, the best estimate mortality rates were shown, along with the 70% and 95% confidence levels. These projections are based on the best-fitting ARIMA model chosen by the code described in Hyndman (2011b), which also calculated the mortality rates associated with the confidence intervals. The levels of confidence 70% and 95% were chosen as they are fairly closely associated with a one- and two-standard-deviation distribution of outcomes under the Normal distribution respectively, as well as for convenience and ease of reference. Introducing confidence intervals would provide a useful idea of the range of outcomes and uncertainty on future mortality, using the model deemed most appropriate for that population.

To quantify this uncertainty financially, the fair price of a theoretical 20-year level annuity paying an income of one each year, ending upon the death of the annuitant if earlier, was calculated. This was sensitivity-tested for different scenarios of mortality, whilst keeping the interest rate component constant. This was to understand the impact of uncertain mortality on the correct price of an annuity. To allow comparisons across countries on the extent of mortality uncertainty, the same flat annual interest rate of 1.5% was used initially in the calculation of annuity prices for all countries. This rate was based on the 10-year yield of a UK government-issued nominal bond in August 2012. As the interest rate was the same, differences in annuity prices would be due to differences in life expectancies and mortality uncertainty from country to country.

For the purposes of interpretation, choosing this interest rate was similar to analysing the price of annuities denominated in British Pounds, on the different populations in different countries. To do this one would need to hedge out future cross movements between the local currency and the British Pound using derivatives. The theory of interest rate parity was used, which was also mentioned in Bekaert, Wei and Xing (2002). If exchange rate movements between the local currency and the British Pound were hedged out using forward contracts, then the resulting interest rate available for pricing should be the UK one. Otherwise there would be an arbitrage opportunity for others to profit in the forward currency market. This could be done by buying currencies undervalued relative to the level predicted by the interest rate parity theory, and selling those overvalued. Over time the forward currency rates would adjust to fair levels according to interest rate parity.

To illustrate, suppose relative to the UK, country A has a higher annual interest rate of 4%, and country B has a lower rate of 0%. If an insurer in A wants to write annuities in British Pounds, it could in theory write annuities in the currency of A at an annual interest rate of 4%. However, in order to convert the annuity payment into British Pounds using derivatives, the insurer would buy into a forward that implies a currency depreciation of A to the Pound at a rate of 2.5% (=4% - 1.5%), from the interest rate parity theory. This means that the insurer has effectively written annuities in British Pounds at an annual rate of 1.5% (=4% - 2.5%).

Likewise, with country B, in local currency terms the annuity is written at 0%. However, because of the implied currency appreciation of B to the British Pound at 1.5% (= 1.5% - 0%), the insurer in B can write the annuity denominated in Pounds at an annual interest rate of 1.5%.

Using a common interest rate also presented other advantages. The absolute annuity prices calculated would be proportional to the life expectancies in the different populations. In other words, populations with longer life expectancies would attract higher annuity prices, once the interest rate was standardized.

In addition, the percentage difference in price between the best estimate, 70% and 95% lower and upper bound mortality cases (calculated over a 20-year horizon) could be taken as the extent of mortality uncertainty, which could be compared across different populations.

The disadvantage with using a common interest rate is that it may not be relevant for a particular country. If the rate is too low, then the financial impact on annuity pricing from mortality uncertainty is exaggerated, and vice versa. Hence, in local currency terms, the financial effect of mortality uncertainty that emerges may not be relevant.

Given that this research also aims to compare mortality projections across countries, the advantages with using a common interest rate outweigh the disadvantages. For this reason, in section 8.3, results using other annual interest rates such as 0.5% and 2.5% were also considered.

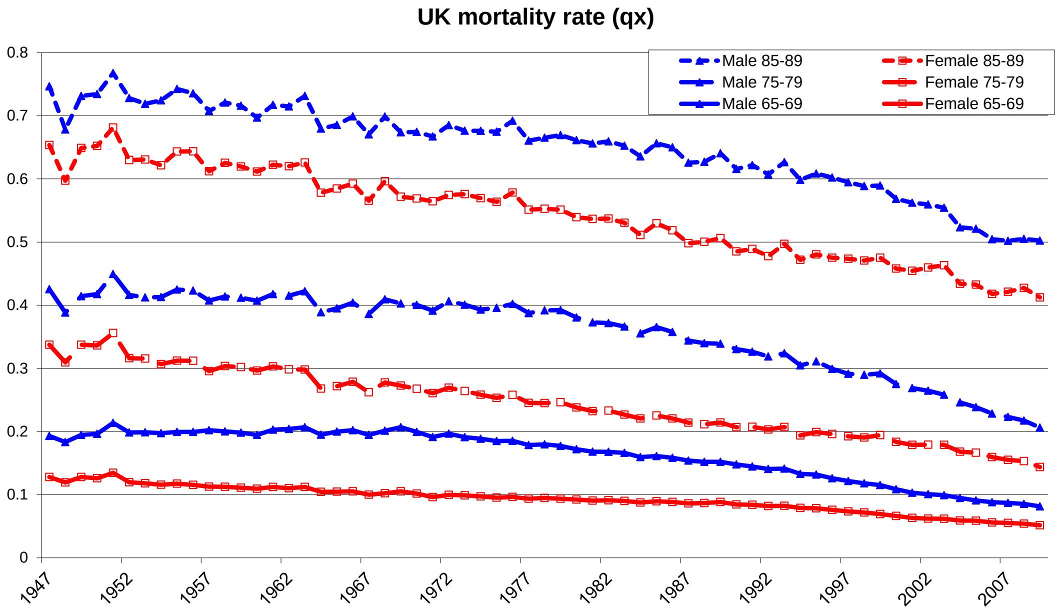
CHAPTER FOUR – UNITED KINGDOM

4.1 Introduction

The underlying data sources for mortality rates in the United Kingdom were the Office for National Statistics, Northern Ireland Statistics and Research Agency and General Register Office for Scotland

Data were available from 1922 to 2009. To remove the effect of World War II on male mortality and to maintain consistency of the starting point with Japan, only the data from 1947 were used. Figure 4.1 illustrates the period mortality rates for UK males and females over the period 1947 to 2009.

Figure 4.1. Mortality rates over time for both genders and different age sub-groups in the UK



For both genders in all three age sub-groups considered, the trend of decreasing mortality was very clear over time. The gap between males and females was more pronounced in the older age sub-groups of 75-79, and 85-89. In all age sub-groups, the gap between male and female mortality was narrowing over time.

4.2 Results and Discussion

Having examined historical trends, mortality models were introduced into the analysis using the methodology described in 3.3. Results for males are presented, followed by females.

4.2.1 Males

4.2.1.1 Analysis over the full period, 1947 – 2009

The parameters of the five models were calibrated based on the full 63 years of mortality history from 1947 to 2009 across ages 65 to 89. However, unlike the other four models, the BMS model only used 21 years (from 1989 to 2009) for fitting. This was because it deemed only this sub-period to meet the LC model's assumption of linearity of mortality improvement, as discussed in 2.1.5. Periods which did not meet this assumption were not considered for fitting purposes by this model.

The BIC was calculated for each model using equation (3.1):

Model	LC	RH	CBD	BMS	HU
BIC	-14266	**-10374**	-12973	-4939	-12978

Table 4.1. BIC statistics, UK males

From the discussion in 3.3.1, as it used much less data, the BMS model had a much less negative log-likelihood measure, less negative BIC and an artificially better BIC measure. Therefore it should not be favoured over the other four models based on the BIC indicator alone.

That aside, the RH model showed the least negative (highest) BIC score, which is shown in bold in Table 4.1.

Having considered the BIC measure, the size of residuals was considered across four age groupings: 65-69, 75-79, 85-89 and 65-89 in summary. In

Table 4.2, residuals are shown in scientific notation (to be multiplied by 10^{-4}). A two-sided t-test for statistical significance was done. Where the associated p-value for the model in that age grouping is greater than 5% (or 0.05), it is given in brackets after the residual and shown in bold.

UK Males	**1947-2009**			
Model	65-89	65-69	75-79	85-89
LC	-3	**0 (0.89)**	**0 (0.8)**	-15
RH	13	2	**0 (0.95)**	65
CBD	**1 (0.2)**	**0 (0.64)**	**-2 (0.11)**	**5 (0.13)**
BMS	**1 (0.3)**	**0 (0.29)**	**1 (0.34)**	**1 (0.39)**
HU	-2	-1	-2	**-3 (0.18)**

Table 4.2. Residuals and p-values (in brackets), 1947-2009, UK males

The BMS and CBD models produced residuals associated with p-values greater than 5% in all four age groupings. Other than the HU model, the remaining four models achieved good fits in the 75-79 age sub-group.

Absolute residuals were then considered, based on 3.3.3. Table 4.3 is set out similarly to Table 4.2, with the absolute residuals shown in scientific notation (to be multiplied by 10^{-4}). The lowest absolute residuals in each age grouping are highlighted in bold. Where the associated p-value for the model in that age grouping is greater than 5% (or 0.05), it is given in brackets after the residual.

UK Males	**1947-2009**			
Model	65-89	65-69	75-79	85-89
LC	31	8	17	85
RH	26	7	**8**	97
CBD	21	7	14	48
BMS	16	**5**	12	39
HU	**15**	6	13	**31**

Table 4.3. Absolute residuals and p-values (in brackets), 1947-2009, UK males

As all associated p-values were below 5%, they were not shown. No model was convincingly the best if evaluated using the absolute residual criterion at a statistically significant level.

At this point, the most credible criterion was the residuals. This was because the BIC usually favours the BMS model, and all p-values under the absolute residuals criterion were below 5%. The BMS and CBD models performed well under the residuals criterion, with p-values greater than 5% in all four age groupings.

Out-of-sample mortality forecasts were then done to assess how projected compared against actual historical mortality, over the 20 years 1990 to 2009. This involved re-calibrating the models based on a shorter fitting period from 1947 to 1989.

4.2.1.2 Forecast over the period 1990 – 2009, calibrated on 1947 - 1989

In this section, the parameters of the models are calibrated to historical mortality rates over the 43 years from 1947 to 1989. The objective is to examine the accuracy of mortality forecasts over the following 20 years (from 1990 to 2009) relative to actual historical mortality. Residuals and absolute residuals are used to assess the quality of fit.

Residuals under the five models are shown in Table 4.4, with the smallest residuals in absolute terms given in bold. Similar to Table 4.2, scientific notation is used (to be multiplied by 10^{-4}), with associated p-values given in brackets if these are greater than 5%.

UK Males	**Forecast:1990-2009**			
Model	65-89	65-69	75-79	85-89
LC	131	67	146	155
RH	**-4 (0.26)**	47	**16**	**-109**
CBD	146	71	139	239
BMS	100	**32**	93	169
HU	154	72	151	242

Table 4.4. Residuals and p-values (in brackets), 1990 – 2009, UK males

The RH model produced the lowest residuals in three of the four age groupings. It was also the only model with a p-value greater than 5% in this out-of-sample testing, importantly for the 65-89 age group.

The absolute residuals (x10^{-4}) are also calculated and shown in Table 4.5, with the smallest highlighted in bold.

UK Males	**Forecast:1990-2009**			
Model	65-89	65-69	75-79	85-89
LC	134	67	146	167
RH	**54**	49	**32**	**116**
CBD	146	71	139	239
BMS	101	**33**	93	173
HU	154	72	151	243

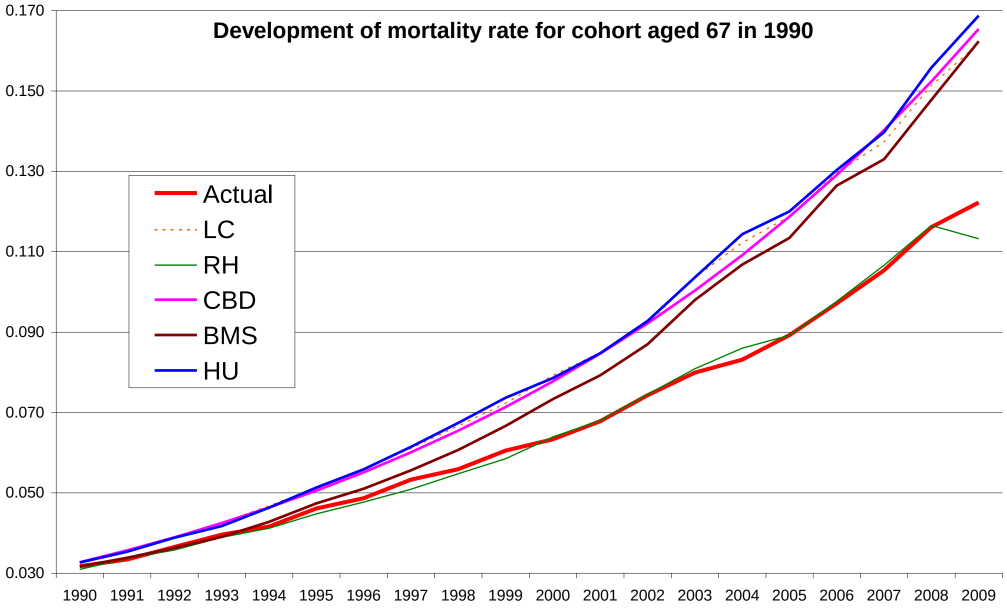
Table 4.5. Absolute residuals and p-values (in brackets), 1990 – 2009, UK males

Similar to the results on residuals in Table 4.4, the RH model produced the smallest absolute residuals in three of the four age groupings. The associated p-values are not shown as they were less than 5%.

Based on the results of the residuals in Table 4.4 and absolute residuals in Table 4.5, the RH model is regarded as having produced mortality forecasts most consistent with subsequent actual experience.

As a check, the mortality rates for a cohort aged 67 in 1990 are projected using the five models and compared against actual historical experience from 1990 to 2009. These are shown in Figure 4.2.

Figure 4.2. Development of mortality rates for cohort aged 67 in 1990, UK males



Projections using the RH model (in green) tracked actual experience (in red) very closely up to 2008, and underestimated mortality in 2009. As Figure 4.2 considered cohort mortality, the rate in 2009 was the mortality rate of age 86 in 2009. The unusual dip in mortality from 2008 to 2009 in the RH model (fitted over the period 1947 to 1989) was due to the forecasted mortality rate of age 86 in 2009 being lower than of age 85 in 2008. This surprising improvement was due to the interaction of fitted and forecasted age, period and cohort parameters. With reference to (2.2), the term $A_{86}^{(2)}P_{2009}^{(2)}$ was markedly more negative (smaller) than $A_{85}^{(2)}P_{2008}^{(2)}$. This was because $P_{2009}^{(2)}$ was smaller than $P_{2008}^{(2)}$ suggesting lighter mortality from the year factor alone, and $A_{86}^{(2)}$ was bigger than $A_{85}^{(2)}$ suggesting greater sensitivity to the year factor. This explained the lower mortality for age 86 in 2009, and more details are given in Appendix 3.

Despite the unusual projection for 2009, relative to the other four models which all overestimated mortality rates, the RH model tracked actual experience much more closely.

In addition to the results in Tables 4.4 and 4.5, Figure 4.2 bolstered the case for the RH model, which also had a good BIC measure (in Table 4.1). Hence, it is chosen as the final model for UK males. Choosing the RH model is also consistent with the finding in Cairns *et al* (2007) of the existence of a cohort effect in male English and Welsh data.

4.2.1.3 Forecast over the period 2010-2029, calibrated on 1947 - 2009

Having chosen the RH model for UK males, projections of mortality rates were made for the 20 years after 2009, in order to gauge the possible extent of developments in mortality.

From section 2.1.3, the RH model is framed as follows:

$$\text{Log } m(t,x) = A_x^{(1)} + A_x^{(2)}P_t^{(2)} + A_x^{(3)}C_c^{(3)} + E(t,x)$$

The full period 1947 to 2009 was used to calibrate the $A_x^{(1)}$, $A_x^{(2)}$, $A_x^{(3)}$, $P_t^{(2)}$ and $C_c^{(3)}$ parameters. For each age x (from age 65 to 89), a value for $A_x^{(1)}$, $A_x^{(2)}$ and $A_x^{(3)}$ was calculated. $P_t^{(2)}$ values were calculated for years 1947 to 2009, and $C_c^{(3)}$ values for cohorts born in 1858 to 1944.

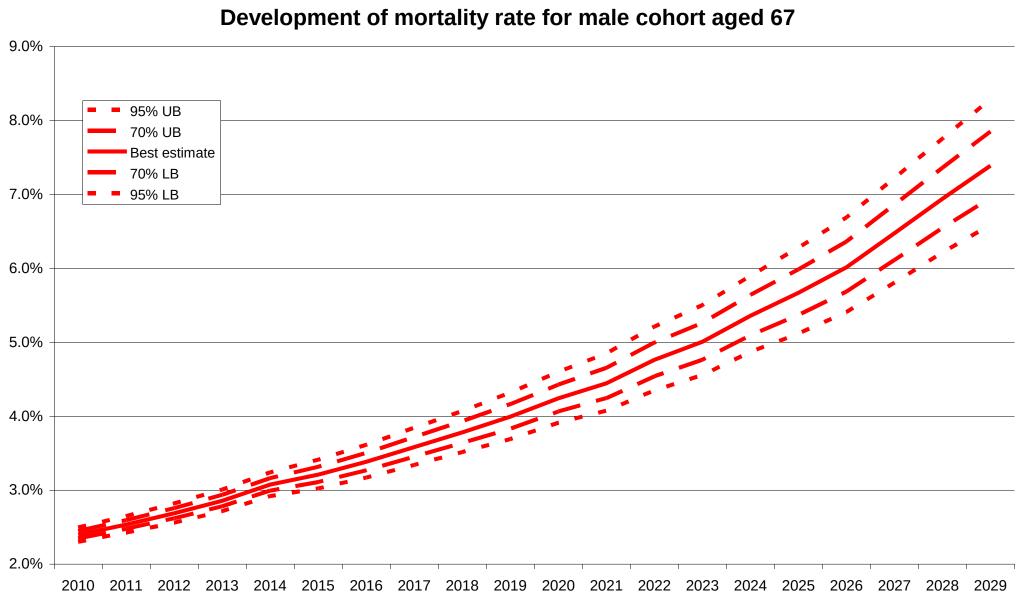
As parameters $P_t^{(2)}$ and $C_c^{(3)}$ were time- and cohort-sensitive respectively, forecasts were obtained by applying ARIMA forecast models to them in R. Using the code described in Hyndman (2011b), the ARIMA(0,1,1) model was selected for $P_t^{(2)}$, and ARIMA(0,1,0) for $C_c^{(3)}$.

Once the parameters had been calibrated, forecasts of mortality rates from 2010 to 2029 were made. Projected mortality rates for males aged 67 in 2010 over the next 20 years were shown, including the best estimate case, and 70% and 95% confidence intervals. This is shown in Figure 4.3.

To put the width of the confidence intervals into perspective, ten years into the projections in 2019, the best estimate mortality rate was 4% for the cohort aged 76. The rates at the 70% and 95% confidence levels were different by 0.2% and 0.3% respectively, producing intervals of 3.8% to 4.2%, and 3.7% to 4.3%

respectively. Twenty years into the projections in 2029, the width of the confidence intervals had increased. The best estimate mortality rate was 7.4% for the cohort aged 86. The 70% interval was 7% to 7.8%, and 95% interval 6.6% to 8.3%. The 95% interval was not symmetric, with a slightly larger room for higher mortality. The mathematical explanation relates to the fact that the log mortality rate is being modelled. When the mortality rate is worked out by raising to the exponent, all else equal a more positive number will be raised to a proportionally larger positive number by the exponential function. This explains the asymmetry mentioned, and is visible at the first decimal point when the 95% confidence interval is considered. The asymmetry at the 70% level is not visible at the first decimal point. This is explained in more detail in Appendix 4. The modelling of the log mortality rate and use of the exponential function also make intuitive sense, because at advanced ages the chance of higher mortality should be greater than lower mortality.

Figure 4.3. Forecasted mortality rates for UK males, 2010-2029



LB = Lower Bound; UB = Upper Bound

4.2.1.4 Quantification of impact of mortality

From section 3.4, to quantify the financial impact of uncertainty around longevity, the theoretical price to purchase a 20-year level annuity paying one pound each year, which stops payment if the annuitant dies before 20 years, was calculated. The objective was to understand how this price would vary, given differences in the forecasts of mortality, under the RH model chosen for UK males. The mortality projections for the best estimate, the 70% and the 95% confidence interval cases were used in this analysis.

For simplicity, an interest rate of 1.5% pa was used, which was close to the 10-year yield on UK government-issued nominal bonds in August 2012, which was at a historically low level. Besides market consistency, the other benefit of using a low level of interest rate was to make the impact of longevity on pricing more pronounced (relative to using a higher interest rate), giving a more representative picture of ‘worst-case’ scenarios if longevity continued to improve for insurance companies offering these annuities.

The prices of the theoretical annuity for different mortality projection scenarios were given, along with the percentage differences in price from the base estimate scenario:

	95% LB	70% LB	BE	70% UB	95% UB
Annuity Price	12.6	12.5	12.3	12.2	12.1
% Difference	2.1%	1.1%	0.0%	-1.1%	-2.1%

LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

Table 4.6. Effect of mortality forecasts on annuity price, UK males

Table 4.6 illustrates the limited effect of different mortality outcomes, even at the 95% confidence level at a low level of interest rate, on the price required to purchase this theoretical annuity. A difference in price of just 2.1% even at extreme mortality levels (with just a 5% chance of occurring) suggesting relative stability in the financial effect of the mortality forecasts, and a relatively small degree of uncertainty around the financial effect of mortality outcomes for UK males.

4.2.2 Females

4.2.2.1 Analysis over the full period, 1947 – 2009

Analysis of UK females was conducted in a similar manner to that for UK males. The parameters of the five models were calibrated based on the full 63 years of mortality history across ages 65 to 89. However, due to its construction, the BMS model only used 22 years of history (from 1988 to 2009). This was because it deemed only this sub-period to meet the LC model's assumption of linearity of mortality improvement, as discussed in 2.1.5.

Using 3.1, the BIC was calculated for each model using equation (3.1).

Model	LC	RH	CBD	BMS	HU
BIC	-14142	**-10604**	-13801	-5137	-12739

Table 4.7. BIC statistics, UK females

Similar to 4.2.1.1, as the BMS model used much less data, by definition it had a much less negative log-likelihood measure and a better BIC measure. That aside, the RH model showed the least negative (highest) BIC score, which was consistent with the result for males in 4.2.1.1.

Having considered the BIC measure, the residuals (x10^{-4}) were calculated across the four age groupings: 65-69, 75-79, 85-89 and 65-89 in summary.

UK Females	**1947-2009**			
Model	65-89	65-69	75-79	85-89
LC	**0 (0.86)**	**1 (0.18)**	**0 (0.68)**	**-1 (0.73)**
RH	5	2	**0 (0.94)**	22
CBD	2	-2	**0 (0.61)**	13
BMS	**0 (0.31)**	**0 (0.34)**	**0 (0.35)**	**1 (0.38)**
HU	-1	**0 (0.07)**	-2	**-2 (0.35)**

Table 4.8. Residuals and p-values (in brackets), 1947-2009, UK females

The BMS and LC models produced residuals associated with p-values greater than 5% in all four age groupings. Other than the HU model, the other four models achieved good fits in the 75-79 age sub-group, as occurred for UK males in Table 4.2.

Absolute residuals (x10^{-4}) were then calculated and shown in Table 4.9.

UK Females	**1947-2009**			
Model	65-89	65-69	75-79	85-89
LC	14	6	9	31
RH	13	4	**6**	40
CBD	16	5	12	40
BMS	**10**	4	7	23
HU	**10**	**3**	9	**21**

Table 4.9. Absolute residuals and p-values (in brackets), 1947-2009, UK females

The associated p-values were not shown as they were all less than 5%. No model was convincingly the best if evaluated using the absolute residual criterion.

At this stage, from Table 4.8, the LC and BMS models were deemed best able to capture the historical mortality trends.

Out-of-sample mortality forecasts were then considered against actual historical mortality.

4.2.2.2 Forecast over the period 1990 – 2009, calibrated on 1947 - 1989

Similar to the analysis for UK males in 4.2.1.2, models were fitted for UK females over the 43 years from 1947 to 1989. Residuals (x10^{-4}) were calculated, with the smallest ones shown in bold.

UK Females	Forecast:1990-2009			
Model	65-89	65-69	75-79	85-89
LC	26	16	16	42
RH	**18**	33	33	-32
CBD	20	**7**	31	**14**
BMS	20	23	**9**	30
HU	37	22	21	72

Table 4.10. Residuals and p-values (in brackets), 1990 – 2009, UK females

All associated p-values were less than 5%.

Absolute residuals were then considered, but all associated p-values were also less than 5%.

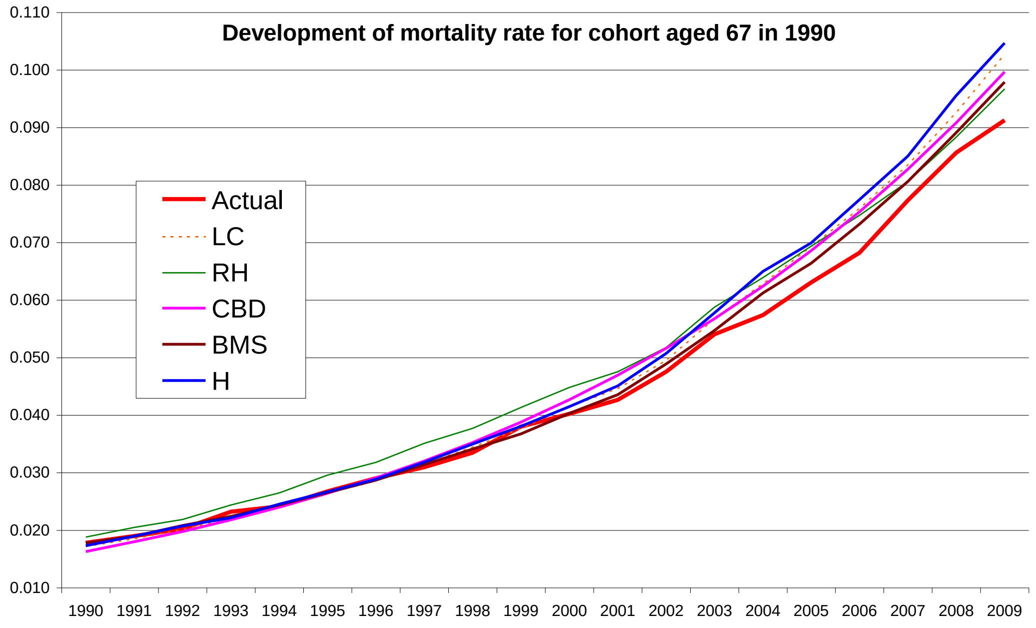
UK Females	Forecast:1990-2009			
Model	65-89	65-69	75-79	85-89
LC	30	18	21	49
RH	38	33	38	45
CBD	28	**13**	32	**37**
BMS	**26**	23	**17**	43
HU	39	22	25	73

Table 4.11. Absolute residuals and p-values (in brackets), 1990 – 2009, UK females

From Tables 4.10 and 4.11, no model performed conclusively better than the other in the out-of-sample testing.

As an extension, mortality rates for a cohort aged 67 in 1990 were projected using the five models and compared against actual experience. The results are shown in Figure 4.4.

Figure 4.4. Development of mortality rates for cohort aged 67 in 1990, UK females



Whilst all models overestimated mortality rates relative to actual experience (in red) over this period, the BMS model (in brown) generally kept most closely to actual experience. This was not surprising from Cairns *et al* (2007), which found historical mortality improvements to be linear in general. When considered with the results in Table 4.8, the BMS model was chosen as the final model for UK females.

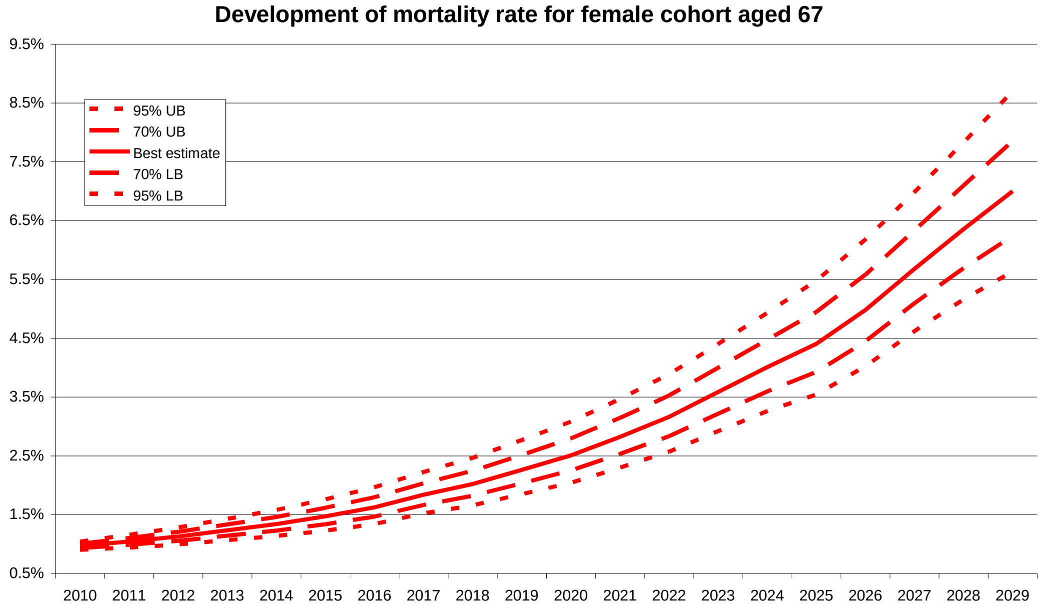
4.2.2.3 Forecast over the period 2010-2029, calibrated on 1947 - 2009

Having chosen the BMS model for UK females, projections of mortality rates for the 20 years after 2009 were made, in order to gauge the possible extent of developments in mortality.

Using R, projected mortality rates for the female cohort aged 67 in 2010 were done on the best estimate basis with 70% and 95% confidence intervals included. These were summarised graphically in Figure 4.5.

Ten years into the projections in 2019, the best estimate mortality rate was 2.3% at age 76. The 70% and 95% confidence levels were 2% to 2.5%, and 1.8% to 2.8% respectively. Twenty years into the projections in 2029, the width of the confidence intervals increased. The best estimate mortality rate was 7% at age 86. The 70% interval was 6.2% to 7.8%, and 95% interval 5.6% to 8.7%. The 95% interval was not symmetric, with a slightly larger room for higher mortality. As the BMS model also models log mortality, the explanation given in Appendix 4 also holds in explaining the asymmetry at the 95% confidence level.

Figure 4.5. Forecasted mortality rates for UK females, 2010-2029



LB = Lower Bound; UB = Upper Bound

Relative to Figure 4.3 for UK males, the range of uncertainty around the best estimate case for females was wider.

4.2.2.4 Quantification of impact of mortality

The analysis for males in 4.2.1.4 was repeated for females. The theoretical prices and percentage differences in price from the best estimate mortality case are shown.

	95% LB	70% LB	BE	70% UB	95% UB
Annuity Price	14.7	14.5	14.3	14.0	13.8
% Difference	2.9%	1.6%	0.0%	-1.7%	-3.3%

LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

Table 4.12. Effect of mortality forecasts on annuity price, UK females

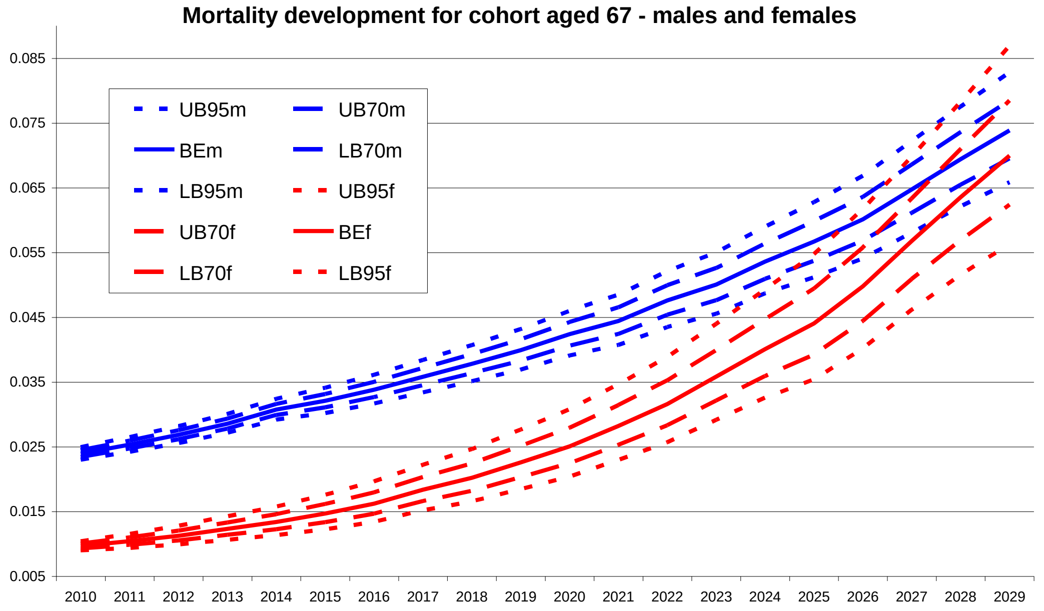
Compared to UK males, the higher uncertainty around mortality observed for UK females in Figure 4.5 was highlighted by Table 4.12, which showed a slightly higher price impact of different mortality outcomes. At the 95% confidence level, there was a 2.1% price impact for males (from Table 4.6), which increased to around 3% for females. Even then, 3% is not too large a value, and this still indicated a reasonable degree of stability in the financial effect of projections around mortality for UK females.

Compared to the results for males in Table 4.6, in Table 4.12 for females there is asymmetry in the percentage price differences between the lower and upper bounds. For females there is a bigger difference at the upper bounds, pointing to greater uncertainty around mortality rates being higher than expected rather than lower. This is possibly because the BMS model deemed higher mortality rates more likely, given the greater mortality improvements females had enjoyed historically already compared to males, as shown in Figure 4.1. Given that there are limits to human life expectancy, as females have already experienced so much in mortality improvements, going forward they are more likely to experience less rather than more of these improvements.

4.3 Conclusion

To summarise the results in 4.2.1.3 and 4.2.2.3, a comparison was made between the best estimate development in cohort mortality for UK males (in blue) and females (in red), along with their 70% and 95% confidence intervals.

Figure 4.6. Comparison of developments in cohort mortality for males and females, for a 67-year old in 2010



LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

Based on the projections, over the ensuing 20 years, the gap between the best estimate mortality for males and females is expected to converge. Intuitively this result makes sense, because there is a limit to human life expectancy. At very advanced ages the mortality differential between males and females is expected to diminish. Hence, it is not surprising to see males beginning to enjoy higher rates of mortality improvement than females. As a result the best estimate mortality rate for males is expected to be only marginally higher than for females by 2029. The 95% upper bound for female mortality would cross the 95% lower bound for male mortality in 13 years, and in 16 years for the 70% bounds.

Figure 4.6 also showed graphically from the width of the bands that uncertainty around female mortality was more than for males. This was possibly because of the models selected to achieve best fit. For males the RH model was chosen, and for females the BMS model. When stochastic projections are made, the model with more parameters leaves more room for uncertainty of outcomes, therefore leading to wider confidence intervals. As the BMS model uses more parameters than RH, this may have contributed towards UK females having wider confidence bands than males.

Interestingly, the convergence of UK male and female mortality over time was also mentioned in Mayhew (2012). He found that the gap between male and female life expectancy at age 30 in England and Wales was closing, and men could catch up with females by 2030. This was thought to be attributable to declining rates of smoking by men and slightly increasing rates for women, as well as other lifestyle factors. In addition, more males were thought to be taking up more sedentary and less hazardous occupations, in a departure from the past. On the other hand, the BBC (2009) reported that in UK school children aged 10 and 11 years, and people aged over 70, males were found to be physically more active than females. In the over-70 age group, men were likely to spend more time in trips outside the house, while women were likely to be doing tasks around the house. Less physical activity was thought to lead to adverse health consequences such as heart disease, diabetes and mental health problems. There were also worries that by exercising less in their youth, girls may be more vulnerable to osteoporosis in later life.

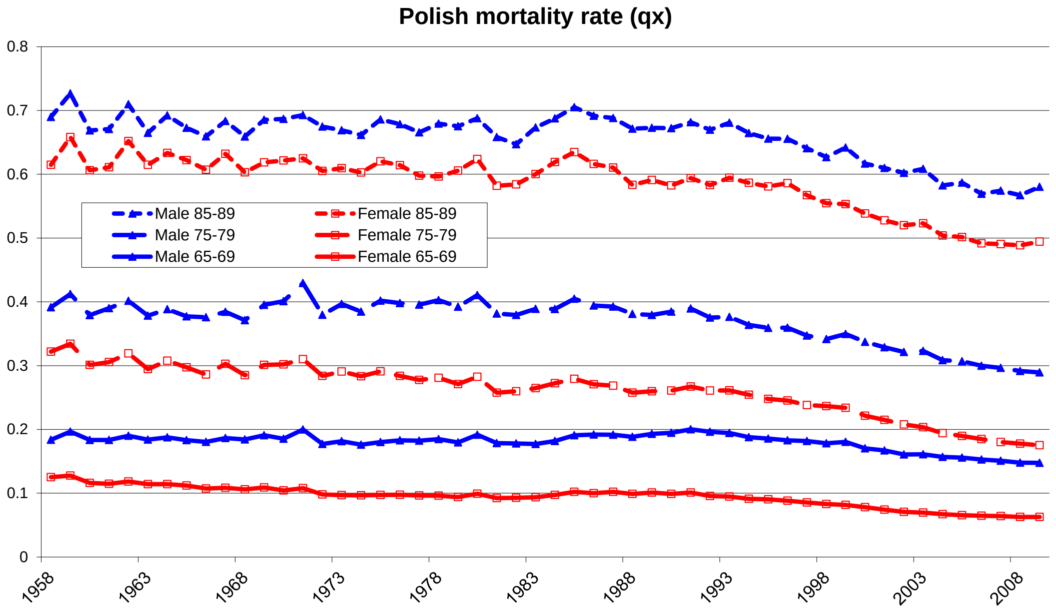
Although different models were chosen for males and females, namely the RH and BMS models respectively, the extent of uncertainty around mortality was not large for old-age UK populations. The financial impact on annuity prices of different mortality outcomes was also manageable, at around 2% and 3% variation respectively even at extreme levels with interest rates at historic lows.

CHAPTER FIVE – POLAND

5.1 Introduction

The underlying data source for Poland was the Central Statistical Office (of Poland). Data were available from 1958 to 2009. Figure 5.1 shows a history of period mortality rates for males and females over this period.

Figure 5.1. Mortality rates over time for both genders and different age sub-groups in Poland



For all three age sub-groups considered, the trend of decreasing mortality was present over time, although not as strongly as compared to the United Kingdom. For this reason, different models may be required to explain adequately the mortality rates in the UK and Poland respectively. The gap between male and female mortality was most noticeable in the 75-79 age sub-group. After 1993, the mortality rates for the 75-79 and 85-89 sub-groups declined linearly.

5.2 Results and Discussion

Having examined historical trends, mortality models were introduced into the analysis using the methodology described in 3.3. Results for males were presented, followed by females.

5.2.1 Males

5.2.1.1 Analysis over the full period, 1958 – 2009

The parameters of the five models were calibrated based on the full 52 years of mortality history across ages 65 to 89. However, due to its construction, the BMS model only used 25 years of history (from 1985 to 2009). This is because it deemed this sub-period to meet the LC model's assumption of linearity of mortality improvement, as discussed in 2.1.5.

Using equation (3.1), the BIC was calculated for each model.

Model	LC	RH	CBD	BMS	HU
BIC	-9111	**-7859**	-9207	-4474	-9000

Table 5.1. BIC statistics, Polish males

As it used much less data, the BMS model by definition had a much less negative log-likelihood measure and hence less negative BIC. For this reason it should not be compared on this measure alone against the other four models.

That aside, the RH model showed the least negative (highest) BIC score, highlighted in bold.

The residuals ($x10^{-4}$) were then calculated for the four age groupings: 65-69, 75-79, 85-89 and 65-89. Table 5.2 follows the same format and convention as Table 4.2.

Polish Males	**1958-2009**			
Model	65-89	65-69	75-79	85-89
LC	**0 (0.94)**	**2 (0.08)**	**0 (0.97)**	**-2 (0.64)**
RH	5	-5	**0 (0.95)**	30
CBD	**-1 (0.69)**	-4	6	**-7 (0.25)**
BMS	**2 (0.1)**	**1 (0.3)**	**2 (0.24)**	**3 (0.2)**
HU	**-1 (0.47)**	**0 (0.75)**	**-1 (0.67)**	**-2 (0.71)**

Table 5.2. Residuals and p-values (in brackets), 1958-2009, Polish males

Based on Table 5.2, the LC, BMS and HU models all performed well, with p-values associated with the residuals all exceeding 5%. In the 65-89 and 85-89 age groupings, all models other than the RH model fitted well. In the 75-79 age sub-group all except the CBD model fitted well.

The absolute residuals (x10^{-4}) are shown in Table 5.3.

Polish Males	**1958-2009**			
Model	65-89	65-69	75-79	85-89
LC	30	13	24	64
RH	28	10	**14**	84
CBD	32	11	26	77
BMS	**23**	9	18	**51**
HU	**23**	**8**	19	**51**

Table 5.3. Absolute residuals and p-values (in brackets), 1958-2009, Polish males

All p-values were less than 5%, so they were not shown. Although the results were not statistically significant, the HU and BMS models were generally associated with low absolute residuals in each of the age groupings considered.

Out-of-sample mortality forecasts against actual historical mortality, over the 20 years 1990 to 2009, were then considered.

5.2.1.2 Forecast over the period 1990 – 2009, calibrated on 1958 - 1989

The parameters of the models were calibrated to historical mortality rates over the 32 years from 1958 to 1989. This was done consistently for all models. The focus was to examine the accuracy of mortality forecasts over the following 20 years, compared to actual mortality over that same period from 1990 to 2009. For this reason, only residuals and absolute residuals were considered.

Table 5.4 shows the residuals (x10^{-4}). The associated p-values were all less than 5%, so were not shown.

Polish Males	**Forecast:1990-2009**			
Model	65-89	65-69	75-79	85-89
LC	166	28	169	298
RH	162	29	166	285
CBD	172	29	164	337
BMS	**147**	**25**	**153**	**262**
HU	166	31	165	305

Table 5.4. Residuals and p-values (in brackets), 1990 – 2009, Polish males

The BMS model produced the lowest residuals for all four age groupings, although none was associated with a p-value greater than 5%.

Table 5.5 shows the absolute residuals (x10^{-4}). The associated p-values were all less than 5%, so were not shown.

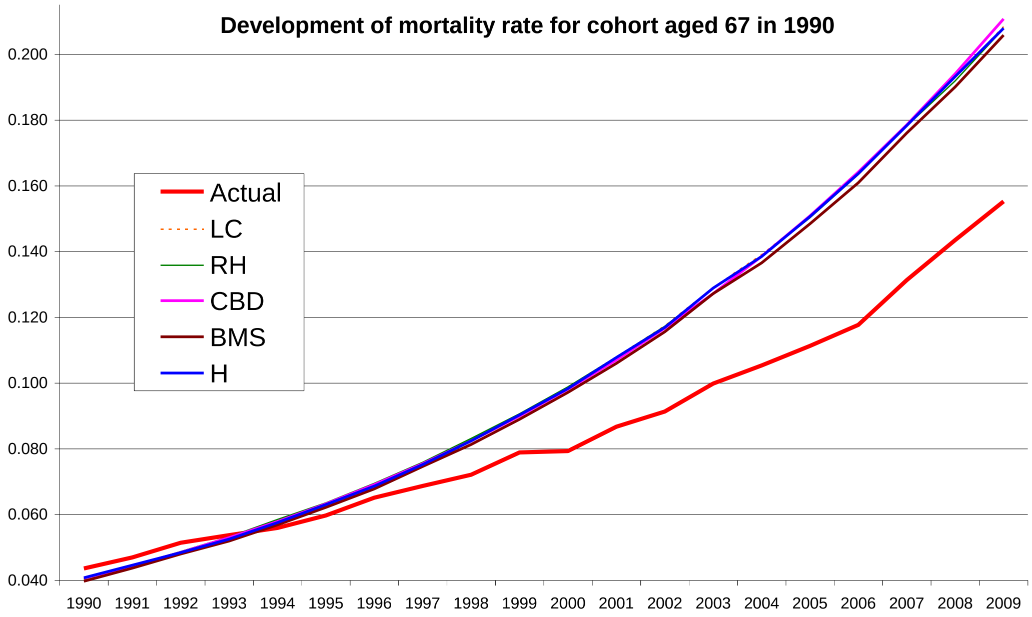
Polish Males	**Forecast:1990-2009**			
Model	65-89	65-69	75-79	85-89
LC	171	**42**	170	305
RH	168	43	171	293
CBD	176	43	165	339
BMS	**157**	**42**	**156**	**279**
HU	171	**42**	167	309

Table 5.5. Absolute residuals and p-values (in brackets), 1990 – 2009, Polish males

As in Table 5.4, in Table 5.5 the BMS model produced forecasts that deviated the least from actual experience. Although none of the associated p-values exceeded 5%, amongst the five models, the BMS model emerged as the best in the group.

To test this further, the mortality rates for a cohort population aged 67 in 1990 were projected using the five models, and compared against actual experience. This is shown in figure 5.2.

Figure 5.2. Development of mortality rates for cohort aged 67 in 1990, Polish males

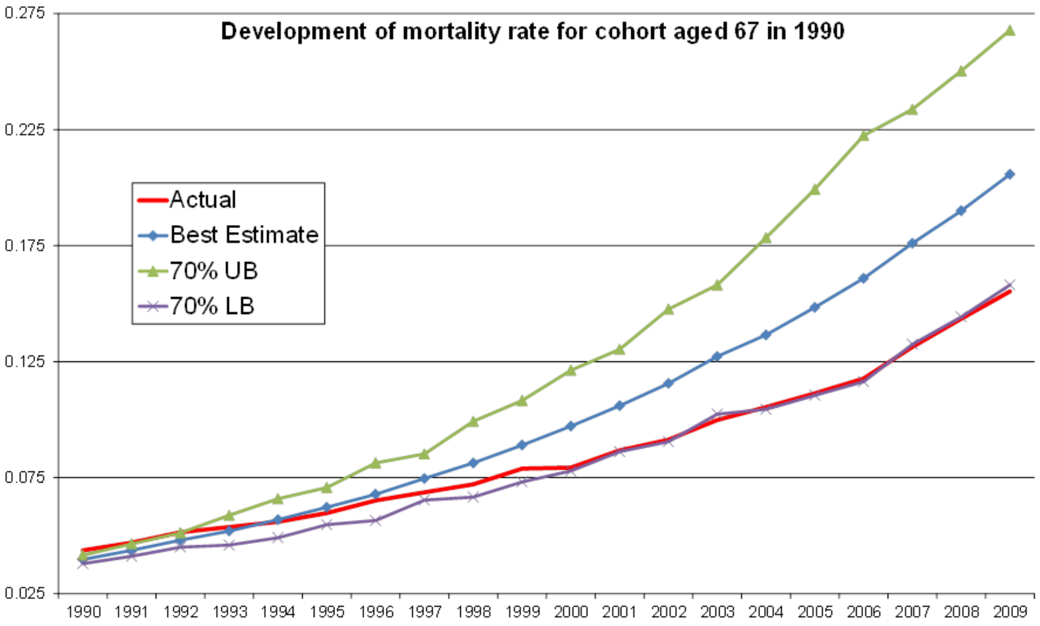


Whilst the best estimate forecasts from all five models overestimated mortality rates significantly relative to actual experience (in red), the BMS model (in brown) was relatively less far away from actual experience.

The most plausible explanation of the poor fit by all models, and the closeness of their projections in the wrong direction, was the choice of fitting period prescribed by the methodology in section 3.4 of this study: the period ending in 1989 (which for Poland is from 1958 to 1989). From Figure 5.1, over the period to 1989, Polish male mortality had fluctuated, at times increasing from one year to the next. After 1989, with democratisation and transition towards a market-based economy in Poland, the male population began to experience more consistent mortality improvements. As these improvements emerged outside the fitting period, all the five models were close in their projections, yet all overestimated the mortality rates that would emerge after 1990. To better calibrate a model for Polish males would require using more data after 1990. However, this is outside the scope of this particular study, but can be part of a study focusing on Polish mortality.

Although the best estimate forecast from the BMS model was far from actual experience, the lower bound of projections at the 70% confidence level was remarkably close to actual experience over the period 1990 to 2009. This is shown in Figure 5.3.

Figure 5.3. Comparison of forecasted mortality rates for Polish males using the BMS model, with actual experience from 2010 to 2029



Even though the mortality improvements were not evident in the data fitting period, the improvements that emerged were still captured by the BMS model at the more extreme ends of its projections. For this reason, the BMS model was picked as the most appropriate of the models considered to conduct further analysis on Polish males.

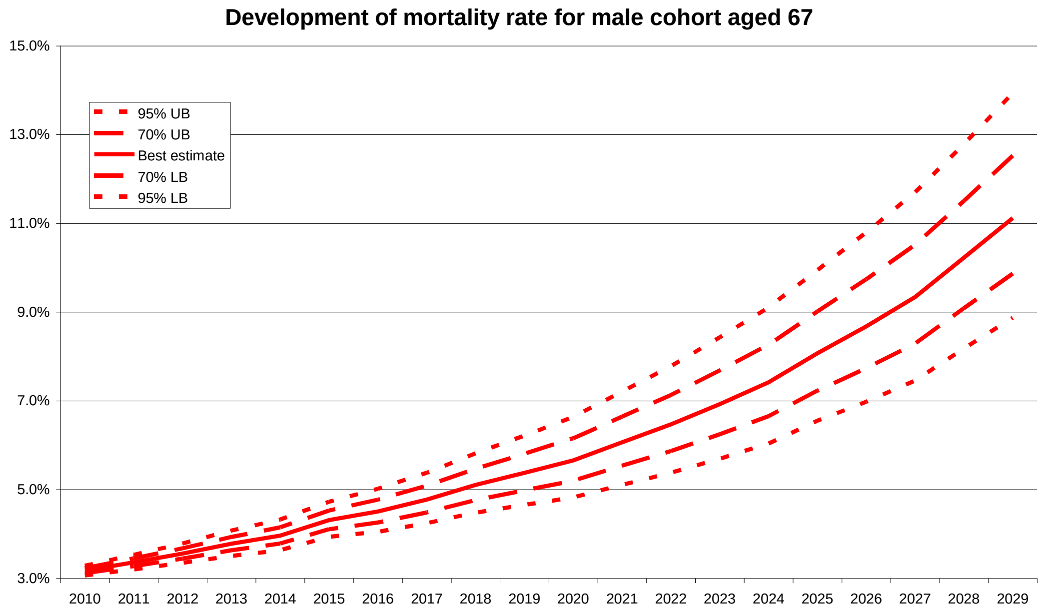
5.2.1.3 Forecast over the period 2010-2029, calibrated on 1958 - 2009

Using the BMS model for Polish males, projections of mortality rates for the 20 years after 2009 were then made, in order to gauge the possible extent of developments in mortality.

Figure 5.4 shows the projected mortality rates for males aged 67 in 2010 over the period to 2029, with the best estimate, and 70% and 95% confidence intervals.

Ten years into the projections in 2019, the best estimate mortality rate for a life aged 76 was 5.4%. The 70% and 95% confidence levels were 5% to 5.8%, and 4.7% to 6.2% respectively. Twenty years into the projections in 2029, the width of the confidence intervals increased. The best estimate mortality rate was 11.1% for a life aged 86. The 70% interval was 9.9% to 12.5%, and 95% interval 8.9% to 13.9%. The 70% and 95% intervals were not symmetric, with a slightly larger room for higher mortality. As the BMS model also models log mortality, the explanation given in Appendix 4 also holds in explaining the asymmetry.

Figure 5.4. Forecasted mortality rates for Polish males, 2010-2029



LB = Lower Bound; UB = Upper Bound

Compared to the forecasts for UK males in Figure 4.3, the extent of uncertainty around mortality was wider for Polish males. This was because compared to the case for UK males, the model did not accurately forecast actual mortality which emerged after 1990, as seen in Figure 5.2. The notable decline in Polish mortality in the 1990s after economic reforms began was not anticipated by the models. The mismatch between fitted and actual mortality was likely a contributor to the wide confidence intervals.

5.2.1.4 Quantification of impact of mortality

Similar to the exercise done for UK pensioners in 4.2.1.4, the financial impact of uncertainty around longevity was quantified by considering the theoretical price to purchase the same 20-year level annuity introduced in 1.5.4 and 3.2.1.4.

Table 5.6 shows the price of the theoretical annuity for different mortality projection scenarios, and the percentage differences from the best estimate scenario:

	95% LB	70% LB	BE	70% UB	95% UB
Annuity Price	11.6	11.4	11.1	10.9	10.6
% Difference	4.4%	2.4%	0.0%	-2.4%	-4.6%

LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

Table 5.6. Effect of mortality forecasts on annuity price, Polish males

Compared to the results for UK males in 4.2.1.4, the absolute prices for the annuities were lower, reflecting the shorter life expectancy for Polish males. However, the uncertainty around Polish male mortality was significantly greater. This was because the relative price impact of different mortality scenarios was around twice as much as for UK males. For example, for Polish males, at the 95% confidence interval, the annuity price could differ from the best estimate case by over 4%. For UK males, this was only around 2%.

5.2.2 Females

5.2.2.1 Analysis over the full period, 1958 – 2009

The parameters of the five models were calibrated based on the full 52 years of mortality history across ages 65 to 89. However, due to its construction, the BMS model only used 21 years of history (from 1989 to 2009). This is because it deemed this sub-period to meet the LC model's assumption of linearity of mortality improvement, as discussed in 2.1.5.

Table 5.7 shows the BIC calculated for each model.

Model	LC	RH	CBD	BMS	HU
BIC	-9266	**-8083**	-9318	-3887	-9259

Table 5.7. BIC statistics, Polish females

As it used much less data, the BMS model by definition had a much less negative log-likelihood measure and hence less negative BIC. For this reason it should not be compared on this measure alone, against the other four models.

That aside, consistent with the result for Polish males, the RH model showed the least negative (highest) BIC score.

Table 5.8 shows the size of residuals (x10^{-4}), across four age groupings: 65-69, 75-79, 85-89 and 65-89. Associated p-values were given in brackets where they exceeded 5%.

Polish Females	**1958-2009**			
Model	65-89	65-69	75-79	85-89
LC	3	**0 (0.53)**	**0 (0.83)**	15
RH	7	-1	**0 (0.84)**	35
CBD	6	-1	**-1 (0.32)**	39
BMS	**1 (0.3)**	**0 (0.36)**	**1 (0.37)**	**1 (0.36)**
HU	**-1 (0.36)**	**0 (0.39)**	**-1 (0.49)**	**-1 (0.84)**

Table 5.8. Residuals and p-values (in brackets), 1958-2009, Polish females

From Table 5.8, the BMS and HU models produced residuals associated with p-values greater than 5% in all age groupings. All models achieved good fits in the 75-79 age sub-group.

Table 5.9 considers similar results for absolute residuals (x10^{-4}). The p-values were not shown as they were below 5%.

Polish Females	**1958-2009**			
Model	65-89	65-69	75-79	85-89
LC	21	6	16	54
RH	20	5	**9**	67
CBD	22	6	17	59
BMS	**14**	**4**	12	34

HU	15	**4**	13	**31**

Table 5.9. Absolute residuals and p-values (in brackets), 1958-2009, Polish females

Although no model was convincingly the best based on Table 5.9, the BMS and HU models still appeared to be better than the others.

Out-of-sample mortality forecasts were then compared against actual historical mortality, over the 20 years 1990 to 2009.

5.2.2.2 Forecast over the period 1990 – 2009, calibrated on 1958 - 1989

The parameters of the models were calibrated to historical mortality rates over the 32 years from 1958 to 1989. The focus was to examine the accuracy of mortality forecasts over the following 20 years, compared to actual mortality over that same period from 1990 to 2009. For this reason, only residuals and absolute residuals were considered.

Table 5.10 shows the residuals (x10^{-4}) and the associated p-values that are greater than 5%.

Polish Females	**Forecast:1990-2009**			
Model	65-89	65-69	75-79	85-89
LC	69	**10**	55	154
RH	**23**	18	**39**	**-16 (0.07)**
CBD	53	18	58	82
BMS	85	18	70	180
HU	84	16	68	184

Table 5.10. Residuals and p-values (in brackets), 1990 – 2009, Polish females

The only p-value greater than 5% was that for the RH model, for the fit to the 85-89 age sub-group, which was 7%.

Table 5.11 shows similar results for absolute residuals (x10^{-4}).

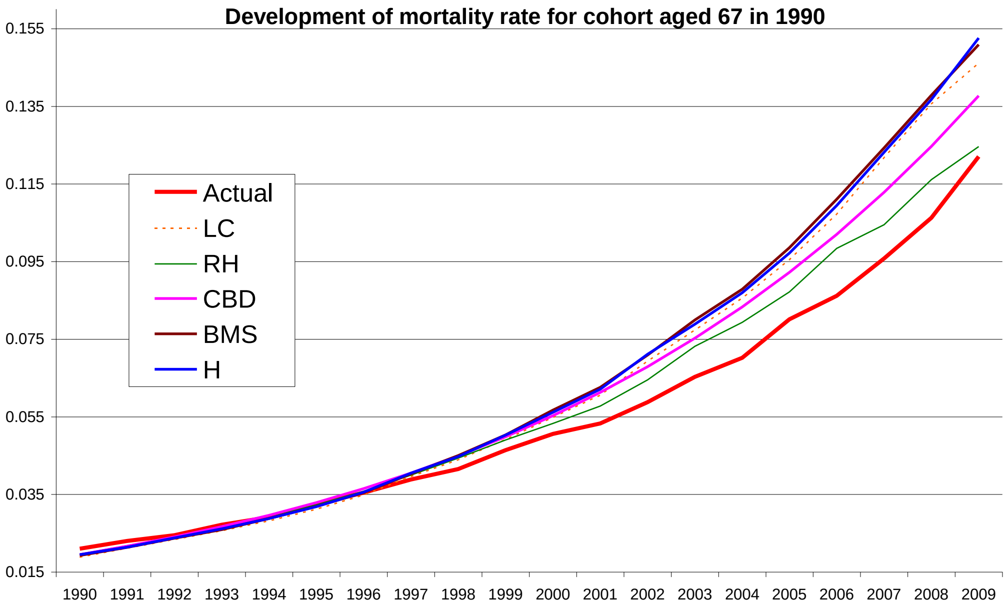
Polish Females	**Forecast:1990-2009**			
Model	65-89	65-69	75-79	85-89
LC	78	**19**	64	164
RH	**49**	22	**52**	**73**
CBD	65	23	67	111
BMS	90	23	76	186
HU	88	21	74	189

Table 5.11. Absolute residuals and p-values (in brackets), 1990 – 2009, Polish females

Both Tables 5.10 and 5.11 suggest the RH model to be more sensible when out-of-sample forecasting was considered.

As a check, mortality rates for a cohort population aged 67 in 1990 were projected using the five models, and compared against actual experience. This is shown in Figure 5.5.

Figure 5.5. Development of mortality rates for cohort aged 67 in 1990, Polish females



Whilst all models overestimated mortality rates over this period, the RH model was not far off from actual experience and was the closest of the models.

Compared to Figure 5.2 for Polish males, the forecasts for females were also closer to actual experience. This was possibly because Polish female mortality rates during the data fitting period were not as volatile as for males.

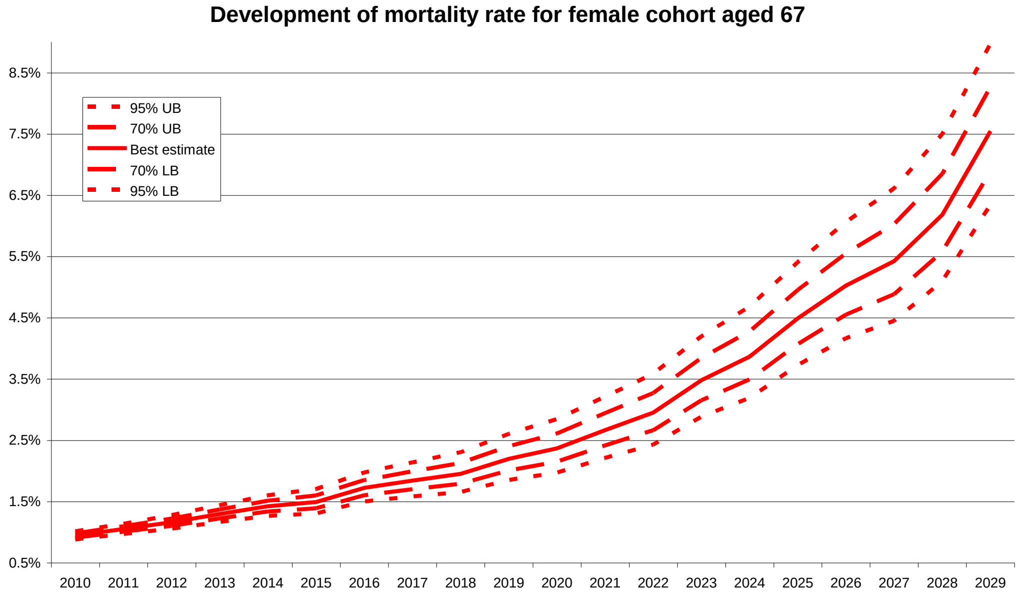
The results in Tables 5.10 and 5.11 and Figure 5.4 supported using the RH model, which also scored well under the BIC criterion in Table 5.7. Hence, the RH model was chosen as the final model for Polish females.

5.2.2.3 Forecast over the period 2010-2029, calibrated on 1958 - 2009

Having chosen the RH model for Polish females, Figure 5.6 shows the projected mortality rates for females aged 67 in 2010 over the next 20 years, showing the best estimate, and 70% and 95% confidence intervals.

Ten years into the projections in 2019, the best estimate mortality rate was 2.2% for a life aged 76. The 70% and 95% confidence levels were 2% to 2.4%, and 1.9% to 2.6% respectively. Twenty years into the projections in 2029, the width of the confidence intervals increased. The best estimate mortality rate was 7.5% for a life aged 86. The 70% interval was 6.9% to 8.3%, and 95% interval 6.4% to 9%. The 70% and 95% intervals were not symmetric, with a larger room for higher mortality. The explanation given in Appendix 4 holds in explaining the asymmetry observed in this instance.

Figure 5.6. Forecasted mortality rates for Polish females, 2010-2029



LB = Lower Bound; UB = Upper Bound

Compared to Figure 5.3 for Polish males, the range of uncertainty around best estimate mortality was noticeably narrower. This was likely because the forecasts from the models were closer to actual experience for females than males, as seen by comparing Figure 5.4 against Figure 5.2.

5.2.2.4 Quantification of impact of mortality

Table 5.12 shows the price of the theoretical annuity for different mortality projection scenarios, and the percentage differences from the best estimate mortality scenario:

	95% LB	70% LB	BE	70% UB	95% UB
Annuity Price	14.6	14.5	14.3	14.1	13.9
% Difference	2.5%	1.3%	0.0%	-1.4%	-2.8%

LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

Table 5.12. Effect of mortality forecasts on annuity price, Polish females

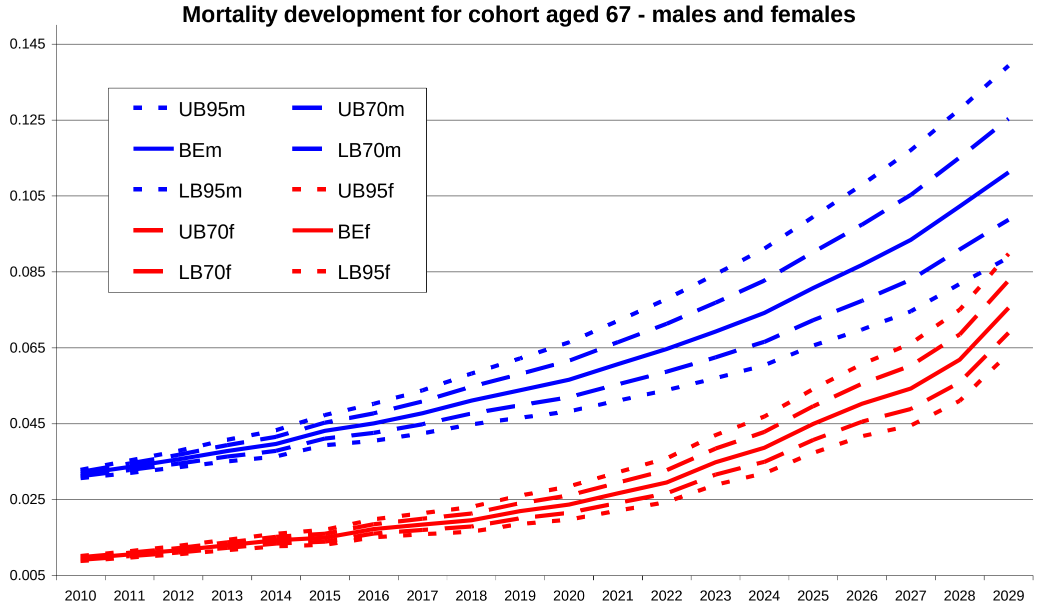
Compared to results for Polish males in Table 5.6, the impact of mortality uncertainty was noticeably less. For example, at the 95% confidence interval, variation to price was expected to be below 3% for females, but higher at above 4% for males. In contrast to Polish males, there is considerably more stability around developments in Polish female mortality.

In fact, the results for Polish females were very similar to those for UK females in Table 4.12. The absolute annuity prices were similar, suggesting similar lengths of life expectancy. In addition, the percentage price differences under different mortality scenarios were also similar between UK and Polish females. This suggested a similar degree of stability around projections in the mortality rates of these two female populations.

5.3 Conclusion

To summarise the results in 5.2.1.3 and 5.2.2.3, Figure 5.7 compared the best estimate development in cohort mortality for Polish males (in blue) and females (in red), along with their 70% and 95% confidence intervals.

Figure 5.7. Comparison of developments in cohort mortality for males and females, for a 67-year old in 2010



LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

Unlike the results for UK pensioners as shown in Figure 4.6, over the next 20 years, the gap between the best estimate mortality for Polish males and females is not expected to narrow noticeably. The 95% upper bound for female mortality only crosses the 95% lower bound for male mortality in 19 years, and no convergence is expected at the 70% level of confidence.

As could be inferred from Tables 5.6 and 5.12, Figure 5.7 also showed graphically from the width of the bands that uncertainty around male mortality was wider than for females.

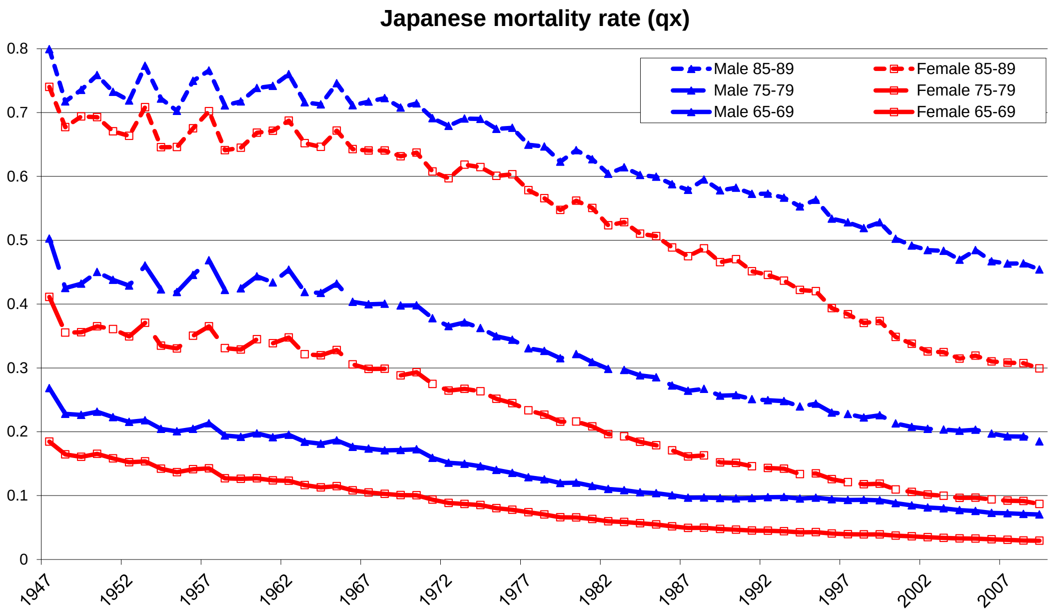
Interestingly, Polish females are expected to experience similar levels of life expectancy in old age to UK females, and uncertainties around mortality are also similar for these two populations. The existence of an adequate model for Polish males is less clear than for females. To the extent that there is one, the output of the model suggests that Polish males are not expected to enjoy as long a life as UK males, and Polish male mortality development is also expected to be more uncertain, based on these projections. Comparing Figure 4.6 with Figure 5.7, relative to UK populations, there is less smoothness in the projected mortality rates for Polish populations. This is possibly due to the following reasons. Firstly, looking at historical mortality rates, since 1950 the UK populations have experienced steady declines in mortality. Looking at Polish mortality rates since 1958, mortality rates had in fact increased in some years. Only after 1990 did mortality rates continue to decline. This naturally introduces more variability in the historical mortality pattern in Poland compared to the UK, which fed through into the projected mortality rates. Secondly, the UK had more years of data (63 years), compared to the 52 for Poland, which gave the underlying models more years of UK data to fit, contributing to a smoother result.

CHAPTER SIX – JAPAN

6.1 Introduction

The underlying data sources were the Statistics Bureau, Ministry of Internal Affairs and Communications, and Ministry of Health and Welfare (of Japan). Data were available from 1947 to 2009. Figure 6.1 shows the mortality rates for males and females over this period.

Figure 6.1. Mortality rates over time for both genders and different age sub-groups in Japan



For both genders in all three age sub-groups considered, the trend of decreasing rates of mortality was very clear over time. The extent of improvements was reflective of the economic progress made by Japan since the end of World War II. This improvement was staggering in the 85-89 age sub-group, especially for females.

6.2 Results and Discussion

The same approach set out in 4.2 and 5.2 were used to analyse mortality rates for Japan.

6.2.1 Males

6.2.1.1 Analysis over the full period, 1947 – 2009

The parameters of the five models were calibrated based on the full 63 years of mortality history across ages 65 to 89. However, due to its design in being selective on which years of data to use, the BMS model only used 22 years of history (from 1988 to 2009). This was because it deemed only this sub-period to meet the LC model's assumption of linearity of mortality improvement, as discussed in 2.1.5.

Table 6.1 shows the BIC calculated for each model.

Model	LC	RH	CBD	BMS	HU
BIC	-18599	**-10700**	-13534	-5672	-13029

Table 6.1. BIC statistics, Japanese males

As it uses much less data, the BMS model by definition has a much less negative log-likelihood measure and hence less negative BIC. For this reason it should not be compared against the other four models on this indicator alone. Excluding the BMS model, the RH model did well on the BIC measure.

Table 6.2 shows the size of residuals (x10^{-4}) and the associated p-values that are greater than 5% in brackets.

Japanese Males	**1947-2009**			
Model	65-89	65-69	75-79	85-89
LC	12	**-1 (0.22)**	4	41
RH	39	**-1 (0.08)**	**0 (0.98)**	194
CBD	5	**0 (0.34)**	**-1 (0.55)**	30
BMS	**1 (0.14)**	**0 (0.43)**	**1 (0.25)**	**2 (0.11)**
HU	-4	-2	-3	-7

Table 6.2. Residuals and p-values (in brackets), 1947-2009, Japanese males

From Table 6.2, the BMS model produced residuals associated with p-values greater than 5% in all age groupings. All models other than the HU model achieved good fits in the 65-69 age sub-group.

Table 6.3 shows similar results for absolute residuals (x10^{-4}).

Japanese Males	**1947-2009**			
Model	65-89	65-69	75-79	85-89
LC	37	12	23	91
RH	54	**4**	8	235
CBD	25	6	16	71
BMS	**10**	6	**7**	**20**
HU	18	6	13	42

Table 6.3. Absolute residuals and p-values (in brackets), 1947-2009, Japanese males

Although no p-value exceeded 5%, from Table 6.3 the BMS model generally produced smaller absolute residuals than the other models.

Out-of-sample mortality forecasts were then assessed against actual historical mortality, over the 20 years 1990 to 2009.

6.2.1.2 Forecast over the period 1990 – 2009, calibrated on 1947 - 1989

The parameters of the models were calibrated to historical mortality rates over the 43 years from 1947 to 1989. The focus was to examine the accuracy of mortality forecasts over the following 20 years, compared to actual mortality over that same period from 1990 to 2009. For this reason, only residuals and absolute residuals were considered.

Table 6.4 considers the residuals (x10^{-4}) and associated p-values.

Japanese Males	**Forecast:1990-2009**			
Model	65-89	65-69	75-79	85-89
LC	**4 (0.07)**	-36	**3 (0.11)**	45
RH	-21	43	**-5 (0.24)**	-119
CBD	18	-32	-5	120
BMS	-34	-39	-32	**-39**
HU	11	**-27**	8	52

Table 6.4. Residuals and p-values (in brackets), 1990 – 2009, Japanese males

Table 6.5 considers similar results for absolute residuals (x10^{-4}).

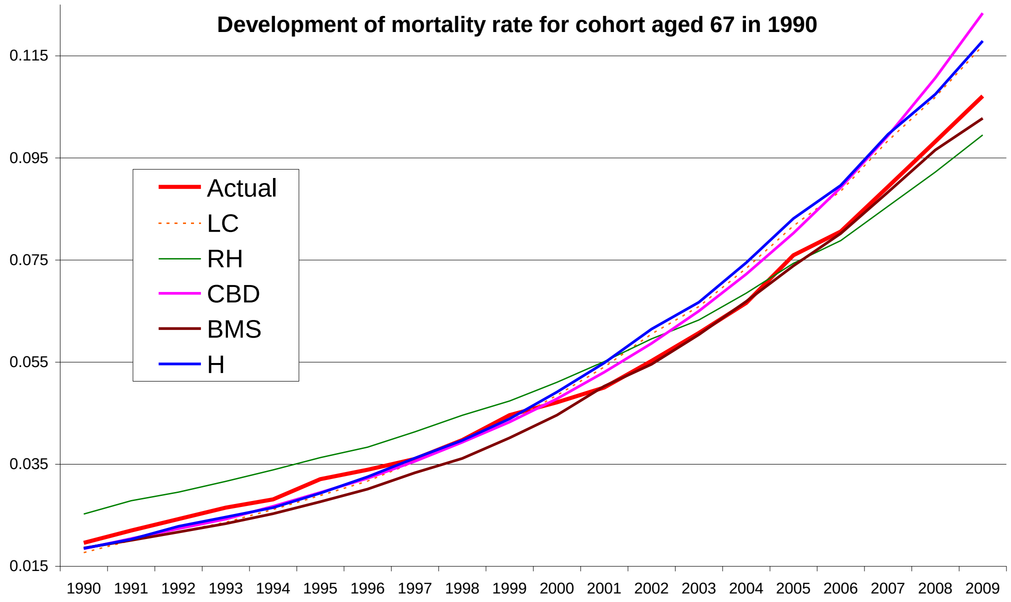
Japanese Males	**Forecast:1990-2009**			
Model	65-89	65-69	75-79	85-89
LC	39	36	13	74
RH	61	43	39	119
CBD	47	32	**12**	129
BMS	**37**	39	33	**45**
HU	39	**27**	15	82

Table 6.5. Absolute residuals and p-values (in brackets), 1990 – 2009, Japanese males

Although in Table 6.5 no p-value exceeded 5% (and p-values were hence not shown), the BMS model did better than the other four under the absolute residual criterion. In addition it was clearly the better model when the full period of data was considered, as from Tables 6.2 and 6.3. However, as in Table 6.4, the LC model appeared better when the residual criterion was used for the out-of-sample test. Then again, the BMS model can be regarded as being similar to the LC, with the main difference being the shorter time period used by the BMS model for fitting.

To test this further, mortality rates for a cohort population aged 67 in 1990 were projected using the five models, and compared against actual experience. This is shown in Figure 6.2.

Figure 6.2. Development of mortality rates for cohort aged 67 in 1990, Japanese males



Interestingly, in contrast to results for Polish pensioners (Figures 5.2 and 5.4) and for UK females (Figure 4.4), some models (especially the RH model in green after 2004) actually underestimated actual mortality (in red). It appeared that the BMS model (in brown) produced forecasts closest to actual experience.

Figure 6.2 showed that forecasts from the BMS model tracked actual experience quite closely. This bolstered the case for choosing BMS as the final model for Japanese males.

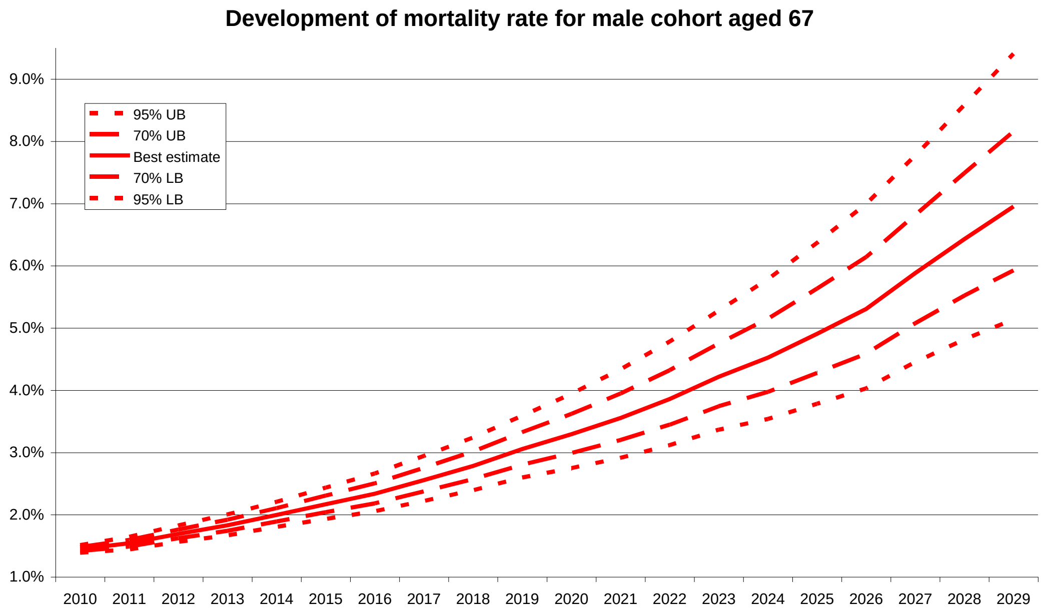
6.2.1.3 Forecast over the period 2010-2029, calibrated on 1947 - 2009

Having chosen the BMS model for Japanese males, Figure 6.3 considers the projected mortality rates for males aged 67 in 2010 over the period ending in 2029 showing the best estimate, and 70% and 95% confidence intervals.

Ten years into the projections in 2019, the best estimate mortality rate was 3.1% for a life aged 76. The 70% and 95% confidence levels were 2.8% to 3.3%, and 2.6% to 3.6% respectively. Twenty years into the projections in 2029, the width of the confidence intervals increased. The best estimate mortality rate was 7% for a life aged 86. The 70% interval was 5.9% to 8.2%, and 95%

interval 5.1% to 9.4%. The 70% and 95% intervals were not symmetric, with a larger band for higher mortality. As the BMS model also models log mortality, the explanation given in Appendix 4 holds in explaining the asymmetry observed in this instance.

Figure 6.3. Forecasted mortality rates for Japanese males, 2010-2029



LB = Lower Bound; UB = Upper Bound

Graphically the extent of uncertainty around Japanese male mortality was similar to those for UK females in Figure 4.5.

6.2.1.4 Quantification of impact of mortality

Similar to the exercise done for UK and Polish pensioners, the financial impact of uncertainty around longevity is quantified by considering the theoretical price to purchase the same 20-year level annuity introduced in 1.5.4. and 3.2.1.4.

Table 6.6 shows the price of the theoretical annuity for different mortality projection scenarios, and the percentage differences from the best estimate mortality scenario:

	95% LB	70% LB	BE	70% UB	95% UB
Annuity Price	13.9	13.7	13.5	13.2	13.0
% Difference	3.3%	1.8%	0.0%	-1.9%	-3.6%

LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

Table 6.6. Effect of mortality forecasts on annuity price, Japanese males

In absolute terms, the annuity prices were higher than those for UK (Table 4.6) and Polish (Table 5.6) males, reflecting the longer life expectancy for Japanese males. In terms of mortality uncertainty, results were less stable than for UK males but more stable than for Polish.

6.2.2 Females

6.2.2.1 Analysis over the full period, 1947 – 2009

The parameters of the five models were calibrated based on the full 63 years of mortality history across ages 65 to 89. However, due to its construction, the BMS model only used 45 years of history (from 1965 to 2009), but much more than the 22 years used by the BMS model for Japanese males in 6.2.1.1, and more than the period for UK and Polish pensioners of both genders. This suggests that Japanese females began to exhibit linear mortality improvements earlier and for a longer period than Japanese males, and UK and Polish pensioners of both genders.

Table 6.7 shows the BIC calculated for each model.

Model	LC	RH	CBD	BMS	HU
BIC	-25436	**-10746**	-16417	-13038	-14136

Table 6.7. BIC statistics, Japanese females

Interestingly, the RH model had the best BIC measure, even than the BMS model despite the latter’s advantage in being selective over the fitting period.

Table 6.8 considers the residuals (x10^{-4}) and associated p-values.

Japanese Females	**1947-2009**			
Model	65-89	65-69	75-79	85-89
LC	12	-2	**2 (0.21)**	49
RH	38	1	**0 (0.86)**	187
CBD	2	-2	**1 (0.32)**	15
BMS	**1 (0.07)**	**0 (0.11)**	**1 (0.06)**	**2 (0.14)**
HU	-5	-2	-4	-11

Table 6.8. Residuals and p-values (in brackets), 1947-2009, Japanese females

Similar to the results for males in Table 6.2, in Table 6.8 the BMS model produced residuals with p-values greater than 5% in all age groupings. All models other than HU achieved good fit in the 75-79 age sub-group.

Table 6.9 considers similar results for absolute residuals (x10^{-4}).

Japanese Females	**1947-2009**			
Model	65-89	65-69	75-79	85-89
LC	37	11	16	103
RH	47	**4**	7	210
CBD	18	5	13	50
BMS	17	6	**6**	49
HU	**14**	5	10	**33**

Table 6.9. Absolute residuals and p-values (in brackets), 1947-2009, Japanese females

No p-value exceeded 5%, and no model was convincingly the best if evaluated using the absolute residual criterion.

Out-of-sample mortality forecasts were then considered against actual historical mortality, over the 20 years 1990 to 2009.

6.2.2.2 Forecast over the period 1990 – 2009, calibrated on 1947 - 1989

The parameters of the models were calibrated to historical mortality rates over the 43 years from 1947 to 1989. The focus was to examine the accuracy of mortality forecasts over the following 20 years, compared to actual mortality over that same period from 1990 to 2009. For this reason, only residuals and absolute residuals were considered.

Table 6.10 considers the residuals ($x10^{-4}$). The associated p-values were not shown as they were all below 5%.

Japanese Females	**Forecast:1990-2009**			
Model	65-89	65-69	75-79	85-89
LC	47	-15	17	161
RH	58	86	69	**14**
CBD	46	-14	20	164
BMS	**11**	-13	-12	72
HU	35	**-10**	**7**	128

Table 6.10. Residuals and p-values (in brackets), 1990 – 2009, Japanese females

Table 6.11 shows similar results for absolute residuals ($x10^{-4}$).

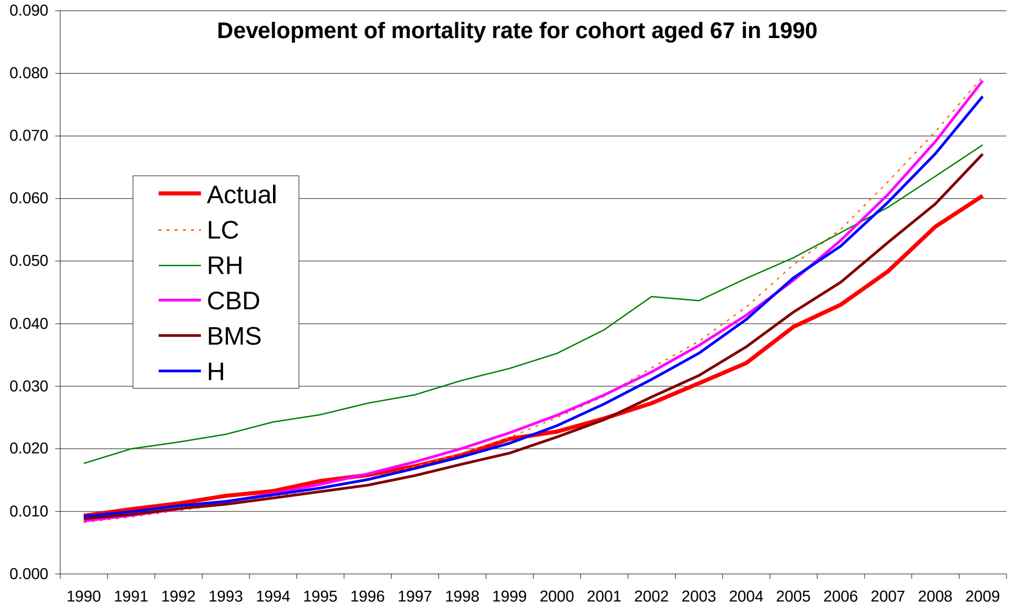
Japanese Females	**Forecast:1990-2009**			
Model	65-89	65-69	75-79	85-89
LC	57	15	18	162
RH	73	86	87	**43**
CBD	54	14	20	165
BMS	**29**	13	13	77
HU	44	**10**	**11**	131

Table 6.11. Absolute residuals and p-values (in brackets), 1990 – 2009, Japanese females

In Tables 6.10 and 6.11, whilst all p-values were less than 5%, the HU model did quite well compared to the other four. However, the BMS model still did well in the overall 65-89 age group.

Figure 6.4 shows the projected mortality rates for a cohort population aged 67 in 1990 using the five models, compared against actual mortality experience.

Figure 6.4. Development of mortality rates for cohort aged 67 in 1990, Japanese females



Whilst all models overestimated actual mortality rates (in red) over this period, the BMS model (in brown) was not far off from actual experience and was the closest of the models to actual experience.

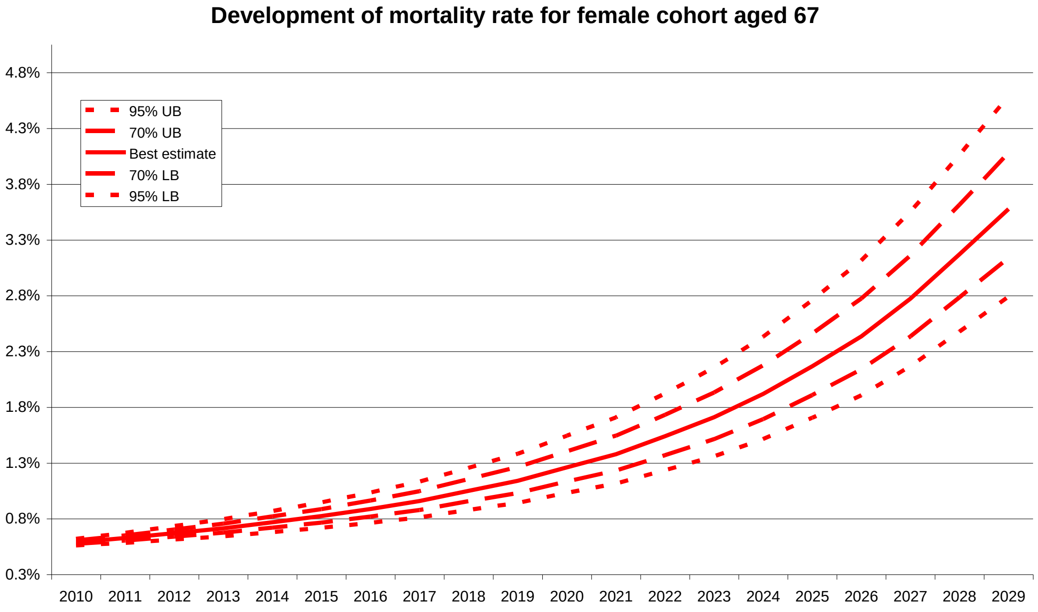
Although Tables 6.10 and 6.11 pointed towards the HU model as being a good choice for Japanese females (with BMS being a reasonable alternative), those tables were not conclusive due to the low p-values produced. Table 6.8 was more credible, as many of the p-values produced were significant statistically. In that table the BMS model performed the best. The suitability of the model was also supported by inspection in Figure 6.4, where the forecasts from the model tracked actual experience most closely than those from the other models. Hence, the BMS model was also chosen for Japanese females (along with Japanese males). This was the first time in this research that the same model was chosen for both genders in a country, possibly due to similarities in the linearity of mortality improvements experienced by both genders.

6.2.2.3 Forecast over the period 2010-2029, calibrated on 1947 - 2009

Having chosen the BMS model for Japanese females, Figure 6.5 shows projected mortality rates for the 20 years after 2009, in order to gauge the possible extent of developments in mortality.

Ten years into the projections in 2019, the best estimate mortality rate was 1.1% for a life aged 76. The 70% and 95% confidence levels were 1% to 1.2%, and 0.9% to 1.3% respectively. Twenty years into the projections in 2029, the width of the confidence intervals increased. The best estimate mortality rate was 3.5% for a life aged 86. The 70% interval was 3.1% to 4%, and 95% interval 2.7% to 4.5%. The 70% and 95% intervals were not symmetric, with a slightly larger room for higher mortality. As the BMS model also models log mortality, the explanation given in Appendix 4 holds in explaining the asymmetry observed in this instance.

Figure 6.5. Forecasted mortality rates for Japanese females, 2010-2029



LB = Lower Bound; UB = Upper Bound

Relative to Japanese males in Figure 6.3, the range of uncertainty around the best estimate mortality for females was narrower.

6.2.2.4 Quantification of impact of mortality

Table 6.12 shows the price of the theoretical annuity for different mortality projection scenarios, and the percentage differences from the base scenario:

	95% LB	70% LB	BE	70% UB	95% UB
Annuity Price	15.8	15.7	15.6	15.4	15.3
% Difference	1.5%	0.8%	0.0%	-0.9%	-1.8%

LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

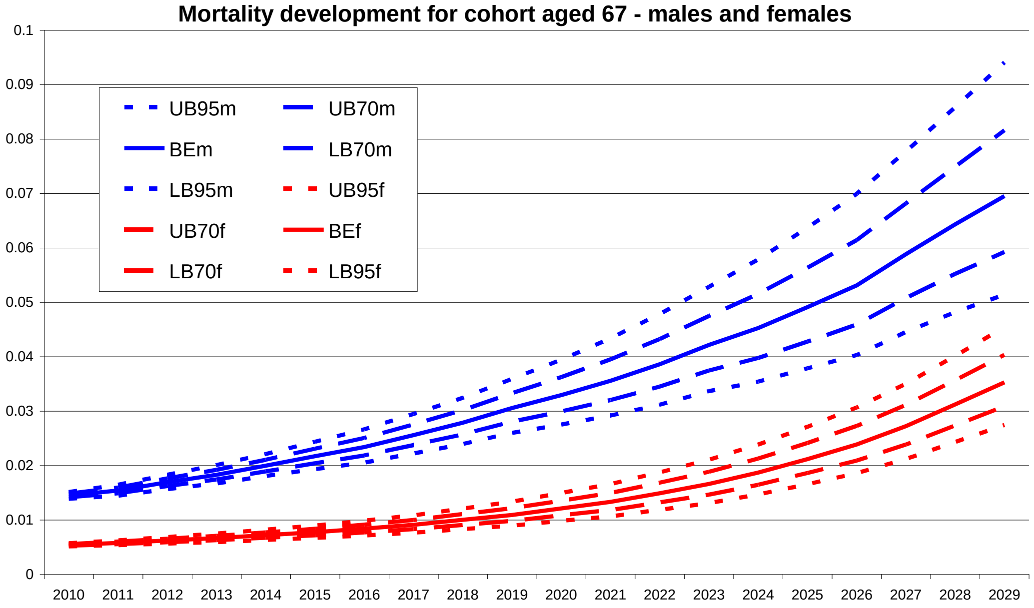
Table 6.12. Effect of mortality forecasts on annuity price, Japanese females

In absolute terms, compared to UK and Polish pensioners and Japanese males, Japanese females attracted the highest annuity prices, reflecting their longest life expectancies in old age out of the populations considered. The impact of mortality uncertainty on the annuity price was also the least out of the six populations considered to date. This also suggested that projections around Japanese female mortality were the most stable out of the six populations.

6.3 Conclusion

To summarise the results in 6.2.1.3 and 6.2.2.3, Figure 6.6 compared the best estimate development in cohort mortality for Japanese males (in blue) and females (in red), along with their 70% and 95% confidence intervals.

Figure 6.6. Comparison of developments in cohort mortality for males and females, for a 67-year old in 2010



LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

Unlike the results for UK pensioners in Figure 4.6, and somewhat similar to results for Poland in Figure 5.7, over the next 20 years, the gap between the best estimate mortality for Japanese males and females is not shown to narrow. The 95% upper bound for female mortality does not even cross the 95% lower bound for male mortality.

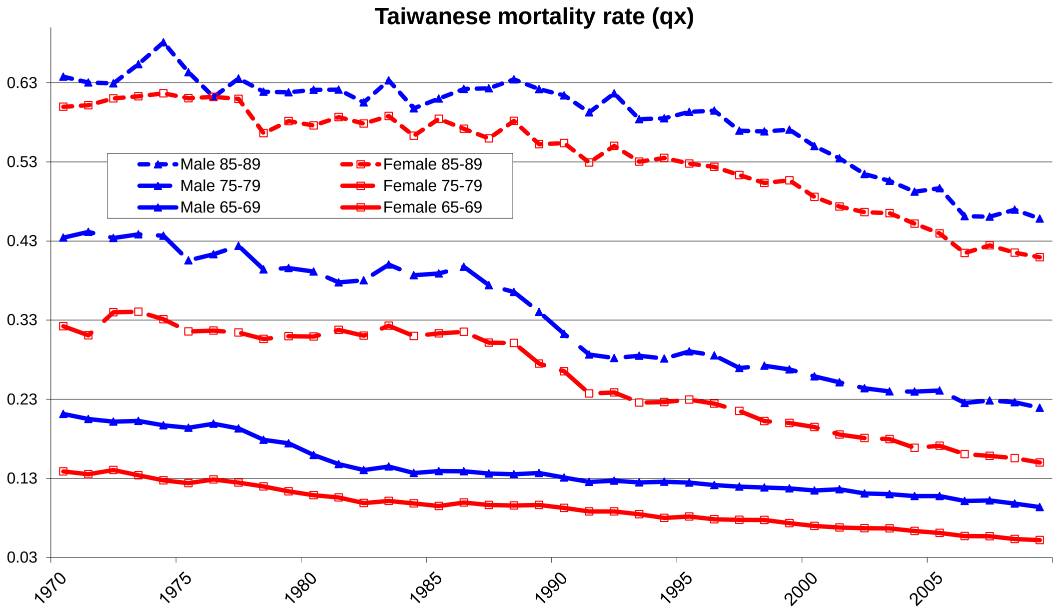
Figure 6.6 also showed graphically from the width of the bands that uncertainty around male mortality was wider than for females. This was because compared to females, the variation in historical male mortality rates had been greater.

CHAPTER SEVEN – TAIWAN

7.1 Introduction

The underlying data source was the Ministry of Interior (of the Republic of China situated on Taiwan). Data were available from 1970 to 2009. Figure 7.1 plots the historical mortality rates for both males and females over this period.

Figure 7.1. Mortality rates over time for both genders and different age sub-groups in Taiwan



For both genders in all three age sub-groups considered, the trend of decreasing rates of mortality was also clear over time, although not as strong as in Japan (in Figure 6.1). The improvement in mortality was more significant for the 75-79 and 85-89 age sub-groups.

7.2 Results and Discussion

The same approach set out in 4.2, 5.2 and 6.2 was used to analyse mortality rates for Taiwan.

7.2.1 Males

7.2.1.1 Analysis over the full period, 1970 – 2009

The parameters of the five models were calibrated based on the full 40 years of mortality history across ages 65 to 89. Despite its construction, the BMS model still used 39 years of history (from 1971 to 2009) for Taiwanese males, as it deemed this long sub-period to still have met the LC model's assumption of linearity of mortality improvement, as discussed in 2.1.5. This was the first time in this research that the BMS model has used almost the full period of data available to fit its parameters.

Table 7.1 shows the BIC calculated for each model.

Model	LC	RH	CBD	BMS	HU
BIC	-7267	**-5449**	-6927	-7385	-6212

Table 7.1. BIC statistics, Taiwanese males

As the BMS model used almost the full period to fit its parameters, it no longer had a structural advantage over the other models, as observed in both genders of UK and Polish pensioners, and in Japanese males. For Taiwanese males, the BMS model actually had the worst BIC measure. As shown in Table 7.1, the RH model had the best BIC measure.

Table 7.2 considers the residuals (x10^{-4}) and associated p-values.

Taiwanese Males	**1970-2009**			
Model	65-89	65-69	75-79	85-89
LC	14	-4	**-1 (0.84)**	61
RH	6	2	**0 (0.93)**	29
CBD	23	3	-34	132
BMS	**-4 (0.07)**	**-2 (0.12)**	**-4 (0.22)**	**-8 (0.23)**
HU	**-2 (0.21)**	**-1 (0.42)**	**-3 (0.39)**	**-3 (0.63)**

Table 7.2. Residuals and p-values (in brackets), 1970-2009, Taiwanese males

The BMS and HU models performed well, having p-values being greater than 5% in all four age groupings. In the 75-79 age sub-groups all models other than CBD fitted well.

Table 7.3 considers the absolute residuals (x10^{-4}) and associated p-values.

Taiwanese Males	**1970-2009**			
Model	65-89	65-69	75-79	85-89
LC	68	17	52	166
RH	33	**9**	**22**	88
CBD	65	13	57	176
BMS	56	16	51	130
HU	**29**	**9**	27	**61**

Table 7.3. Absolute residuals and p-values (in brackets), 1970-2009, Taiwanese males

Although all p-values were below 5% in Table 7.3 (and were hence not shown), the HU model generally produced the smallest absolute residuals.

Out-of-sample mortality forecasts are then considered against actual historical mortality, over the 20 years 1990 to 2009.

7.2.1.2 Forecast over the period 1990 – 2009, calibrated on 1970 - 1989

The parameters of the models were calibrated to historical mortality rates over the 20 years from 1970 to 1989. The focus was to examine the accuracy of mortality forecasts over the following 20 years, compared to actual mortality over that same period from 1990 to 2009. For this reason, only residuals and absolute residuals were considered.

Table 7.4 considers the residuals (x10^{-4}). Associated p-values were below 5% so were not shown.

Taiwanese Males	**Forecast:1990-2009**			
Model	65-89	65-69	75-79	85-89
LC	**96**	-60	169	269
RH	138	70	89	**150**
CBD	**96**	-35	**46**	358
BMS	100	-41	188	236
HU	114	**-31**	198	263

Table 7.4. Residuals and p-values (in brackets), 1990 – 2009, Taiwanese males

Table 7.5 considers the absolute residuals (x10^{-4}). The p-values were again all less than 5%.

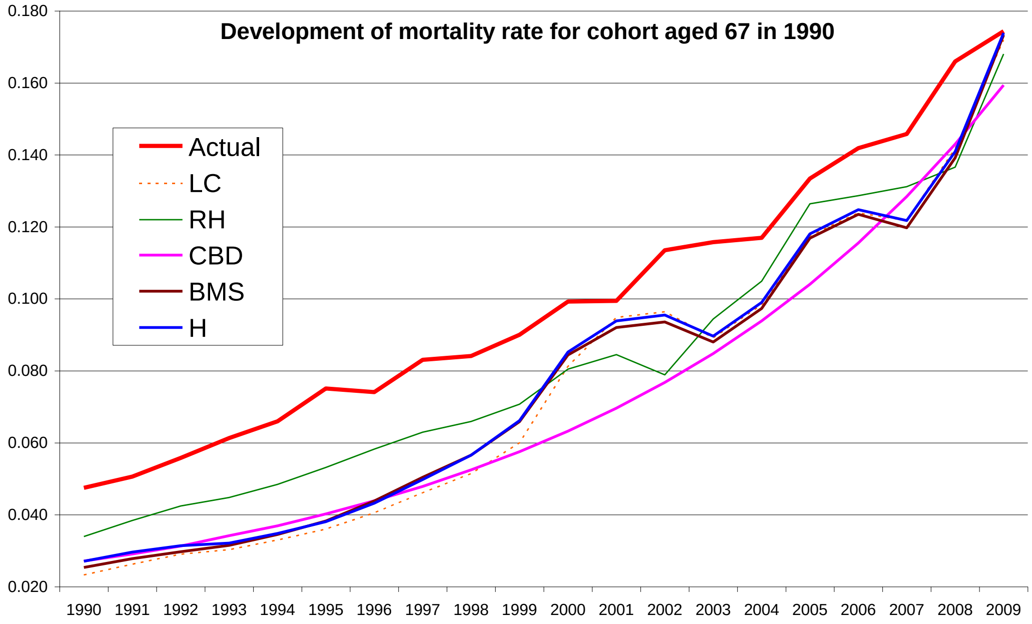
Taiwanese Males	**Forecast:1990-2009**			
Model	65-89	65-69	75-79	85-89
LC	144	60	173	280
RH	160	70	129	**214**
CBD	**118**	35	**51**	358
BMS	134	41	188	254
HU	139	**32**	198	269

Table 7.5. Absolute residuals and p-values (in brackets), 1990 – 2009, Taiwanese males

In Tables 7.4 and 7.5, no model was convincingly the best, particularly as none of the p-values exceeded 5%.

As before, mortality rates for a cohort population aged 67 in 1990 were projected using the five models, and compared against actual experience. The results are shown in Figure 7.2.

Figure 7.2. Development of mortality rates for cohort aged 67 in 1990, Taiwanese males



Interestingly, in contrast to results for other populations, all models underestimated actual mortality (in red). Graphically, the HU (in blue) and BMS (in brown) model seemed to have tracked actual mortality closest, followed by the RH model.

As an aside, compared to the other countries, actual cohort mortality rates for Taiwanese males did not always increase over time with age. In some years for some ages, the rate stayed flat compared to the prior age. In some years for some ages, the rate even decreased. The occasional jaggedness in actual experience affected the pattern of forecasted mortality rates too, except that of the CBD model. In the developed economies UK and Japan where mortality improvements came into effect much earlier and were generally more stable, this jaggedness did not occur. In Poland, another emerging economy, occasionally the actual cohort mortality rate also stayed flat from one age to the next, as observed in Figure 5.2.

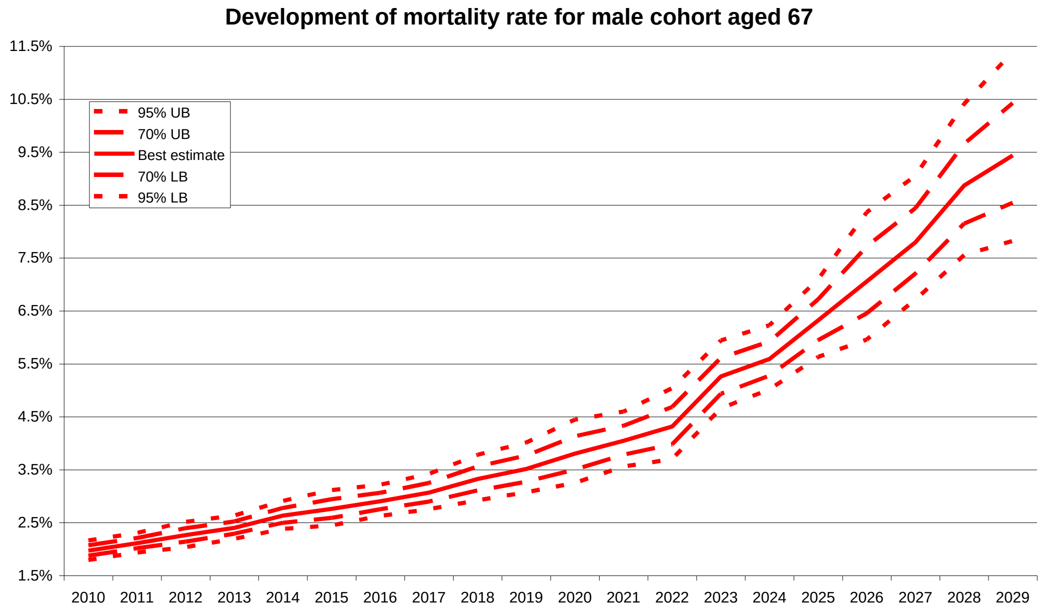
Based largely on Tables 7.2 and 7.3, the HU model appeared to be better than BMS. The HU model also did better under the BIC in Table 7.1. For these reasons, the HU model was chosen for Taiwanese males, although the BMS model might well have been a possible alternative. This did not differ much from the model choices in Wang and Liu (2010), which was the BMS model for males and LC for females when the period of fitting was just over 30 years, and the HU model for both genders for 20 years or less. However, strictly that finding was of limited relevance to this research as it considered mortality rates across all ages, not just in the 65-89 group.

7.2.1.3 Forecast over the period 2010-2029, calibrated on 1970 - 2009

Having chosen the HU model for Taiwanese males, Figure 7.3 shows projected mortality rates for males aged 67 in 2010 over the period ending in 2029, showing the best estimate, and 70% and 95% confidence intervals.

Ten years into the projections in 2019, the best estimate mortality rate was 3.5% for a life aged 76. The 70% and 95% confidence levels were 3.3% to 3.8%, and 3.1% to 4% respectively. Twenty years into the projections in 2029, the width of the confidence intervals increased. The best estimate mortality rate was 9.4% for a life aged 86. The 70% interval was 8.5% to 10.4%, and 95% interval 7.8% to 11.4%. The 70% and 95% intervals were not symmetric, with a larger room for higher mortality. As the HU model also models log mortality, the explanation given in Appendix 4 holds in explaining the asymmetry observed in this instance.

Figure 7.3. Forecasted mortality rates for Taiwanese males, 2010-2029



LB = Lower Bound; UB = Upper Bound

Graphically, Figure 7.3 indicated that the extent of uncertainty around the best estimate mortality projection was similar between Taiwanese and Japanese males (Figure 6.3).

7.2.1.4 Quantification of the impact of mortality

Table 7.6 shows the price of the theoretical annuity for different mortality projection scenarios, and the percentage differences from the base scenario:

	95% LB	70% LB	BE	70% UB	95% UB
Annuity Price	13.1	12.9	12.7	12.5	12.3
% Difference	3.1%	1.7%	0.0%	-1.7%	-3.4%

LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

Table 7.6. Effect of mortality forecasts on annuity price, Taiwanese males

As seen in Polish and Japanese males, for Taiwanese males the distribution of annuity prices was not symmetrical, with higher variation in the event of extreme mortality at the 95% upper bound level. This was possibly because in moving towards very old ages, there was a greater chance of people dying than living, as there is a limit to how long people can live. In absolute terms, the annuity prices were lower than for Japanese males (Table 6.6), but higher than for UK (Table 4.6) and Polish (Table 5.6) males, reflecting the different life expectancies in the populations. In terms of mortality uncertainty, results were less stable than for UK males, similar to Japanese males, and more stable than for Polish males.

7.2.2 Females

7.2.2.1 Analysis over the full period, 1970 – 2009

The parameters of the five models were calibrated based on the full 40 years of mortality history across ages 65 to 89. However, due to its construction, the BMS model only used 24 years of history (from 1986 to 2009), markedly less than the 39 years used by the BMS model for Taiwanese males in 7.2.1.1. This suggested that Taiwanese males began to exhibit linear mortality improvements earlier and for a longer period than Taiwanese females.

Table 7.7 shows the BIC for each model.

Model	LC	RH	CBD	BMS	HU
BIC	-5984	**-5273**	-6118	-3477	-5898

Table 7.7. BIC statistics, Taiwanese females

As the BMS model used only about half of the full history available, its structural advantage re-appeared, giving it the best measure under BIC. When the BMS model was excluded, the RH model showed the best score under the BIC criterion.

Table 7.8 considers the residuals (x10^{-4}) and associated p-values.

Taiwanese Females	**1970-2009**			
Model	65-89	65-69	75-79	85-89
LC	4	**-1 (0.17)**	**0 (0.92)**	17
RH	3	**1 (0.05)**	**0 (0.95)**	14
CBD	9	3	-22	63
BMS	**-1 (0.3)**	**0 (0.43)**	**-2 (0.31)**	**-1 (0.39)**
HU	**-2 (0.12)**	**-1 (0.39)**	**-2 (0.3)**	**-3 (0.52)**

Table 7.8. Residuals and p-values (in brackets), 1970-2009, Taiwanese females

From Table 7.8, as for Taiwanese males in Table 7.2, the BMS and HU models both did well, with associated p-values being greater than 5% in all four age groupings. In the 65-69 and 75-79 age sub-groups, all models other than CBD fitted well.

Table 7.9 considers the absolute residuals (x10^{-4}). The associated p-values were all less than 5%, so they were not shown.

Taiwanese Females	**1970-2009**			
Model	65-89	65-69	75-79	85-89
LC	33	10	37	66
RH	23	6	**16**	59
CBD	38	9	40	94
BMS	**20**	**4**	26	**44**
HU	21	6	19	46

Table 7.9. Absolute residuals and p-values (in brackets), 1970-2009, Taiwanese females

Based on Table 7.9, no model was convincingly the best if evaluated using the absolute residual criterion, although the BMS model performed reasonably there, as did the HU model, due to the differences between the two models being very small.

Out-of-sample mortality forecasts were then compared against actual historical mortality, over the 20 years 1990 to 2009.

7.2.2.2 Forecast over the period 1990 – 2009, calibrated on 1970 - 1989

The parameters of the models were calibrated to historical mortality rates over the 20 years from 1970 to 1989. The focus was to examine the accuracy of mortality forecasts over the following 20 years, compared to actual mortality over that same period from 1990 to 2009. For this reason, only residuals and absolute residuals are considered.

Table 7.10 considers the residuals ($x10^{-4}$). Associated p-values are not shown as they were below 5%.

Taiwanese Females	**Forecast:1990-2009**			
Model	65-89	65-69	75-79	85-89
LC	**100**	-16	206	170
RH	171	139	265	**129**
CBD	132	10	**86**	361
BMS	108	**-1**	219	178
HU	117	5	222	193

Table 7.10. Residuals and p-values (in brackets), 1990 – 2009, Taiwanese females

Table 7.11 considers the absolute residuals ($x10^{-4}$). Associated p-values were not shown as they were below 5%.

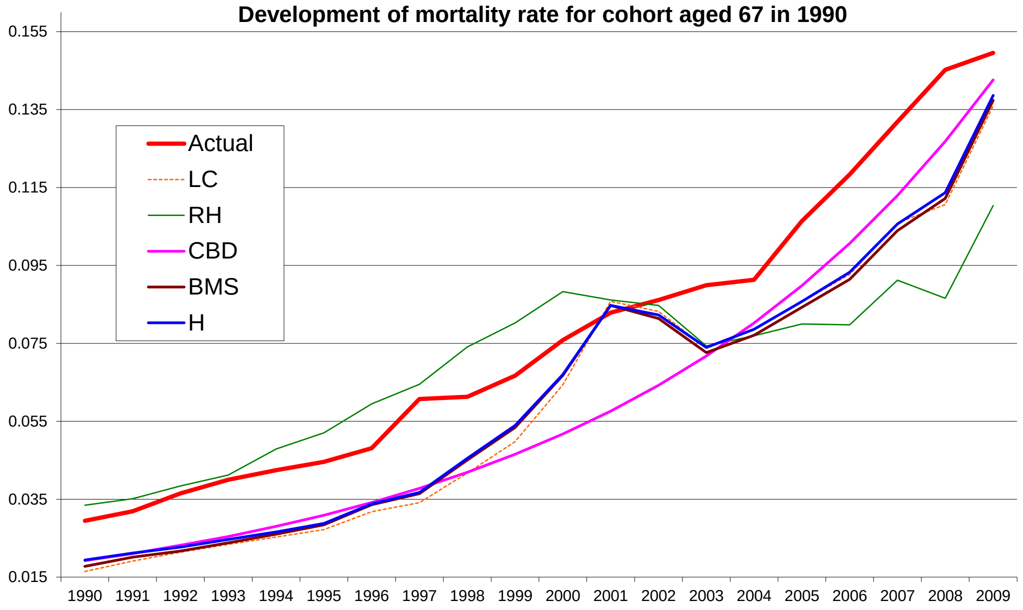
Taiwanese Females	**Forecast:1990-2009**			
Model	65-89	65-69	75-79	85-89
LC	**111**	17	207	179
RH	179	139	270	**154**
CBD	133	10	**90**	361
BMS	112	**7**	221	186
HU	120	**7**	223	199

Table 7.11. Absolute residuals and p-values (in brackets), 1990 – 2009, Taiwanese females

From Tables 7.10 and 7.11, no model was clearly the best, particularly as none of the associated p-values exceeded 5%.

Figure 7.4 shows the projected mortality rates for a cohort population aged 67 in 1990 across the five models, compared against actual experience.

Figure 7.4. Development of mortality rates for cohort aged 67 in 1990, Taiwanese females



Unfortunately unlike Figure 7.2, Figure 7.4 did not contribute towards helping to find the right model. The forecasts from the BMS and HU models were very similar. Although the forecast from the CBD model was close towards the end, in the early years it had severely underestimated mortality. The RH model was promising initially but also underestimated mortality severely in the latter half.

As an aside, similar to males, Taiwanese females also experienced ages when cohort mortality remained flat or even decreased. The occasional jaggedness in actual experience also affected the shape of forecasted mortality rates, except of the CBD model. Throughout this research the CBD model has consistently produced mortality forecasts which are smooth and increasing with age.

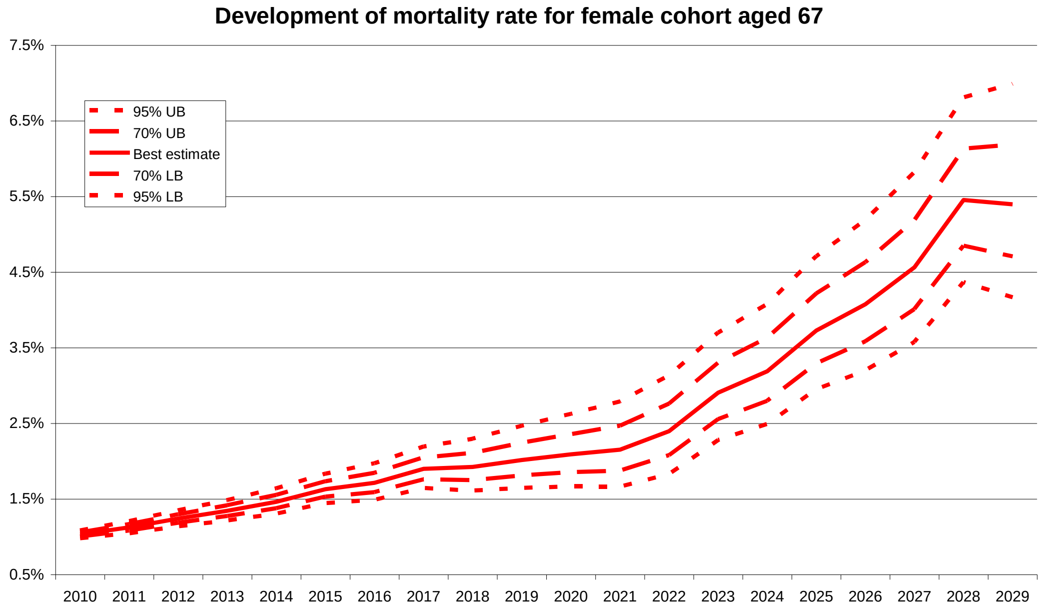
Based largely on the results in Tables 7.8 and 7.9, and partly the better BIC measure in Table 7.7, the BMS model was chosen for Taiwanese females, although this time the HU model could have been an appropriate alternative. It was interesting that for Taiwanese pensioners, the best two available models for both males and females were the BMS and HU models, though no model fitted very well.

7.2.2.3 Forecast over the period 2010-2029, calibrated on 1970 - 2009

Having chosen the BMS model for Taiwanese females, Figure 7.5 shows projected mortality rates for females aged 67 in 2010 over the period ending in 2029, showing the best estimate, and 70% and 95% confidence intervals.

Ten years into the projections in 2019, the best estimate mortality rate was 2% for a life aged 76. The 70% and 95% confidence levels were 1.8% to 2.2%, and 1.6% to 2.5% respectively. Twenty years into the projections in 2029, the width of the confidence intervals increased. The best estimate mortality rate was 5.4% for a life aged 86. The 70% interval was 4.7% to 6.2%, and 95% interval 4.2% to 7%. The 70% and 95% intervals were not symmetric, with a larger room for higher mortality. As the BMS model also models log mortality, the explanation given in Appendix 4 holds in explaining the asymmetry observed in this instance.

Figure 7.5. Forecasted mortality rates for Taiwanese females, 2010-2029



LB = Lower Bound; UB = Upper Bound

Graphically, the range of uncertainty around the best estimate mortality was less for females than for males in Figure 7.3. However, from the output of the model, unique to Taiwanese females, a marginal decline in mortality rate (of less than 0.2%) was expected from 2028 to 2029 under the best estimate and mortality lower bound scenarios.

7.2.2.4 Quantification of the impact of mortality

Table 7.12 shows the price of the theoretical annuity for different mortality projection scenarios, and the percentage differences from the base scenario:

	95% LB	70% LB	BE	70% UB	95% UB
Annuity Price	14.8	14.6	14.4	14.2	14.0
% Difference	2.6%	1.4%	0.0%	-1.5%	-3.0%

LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

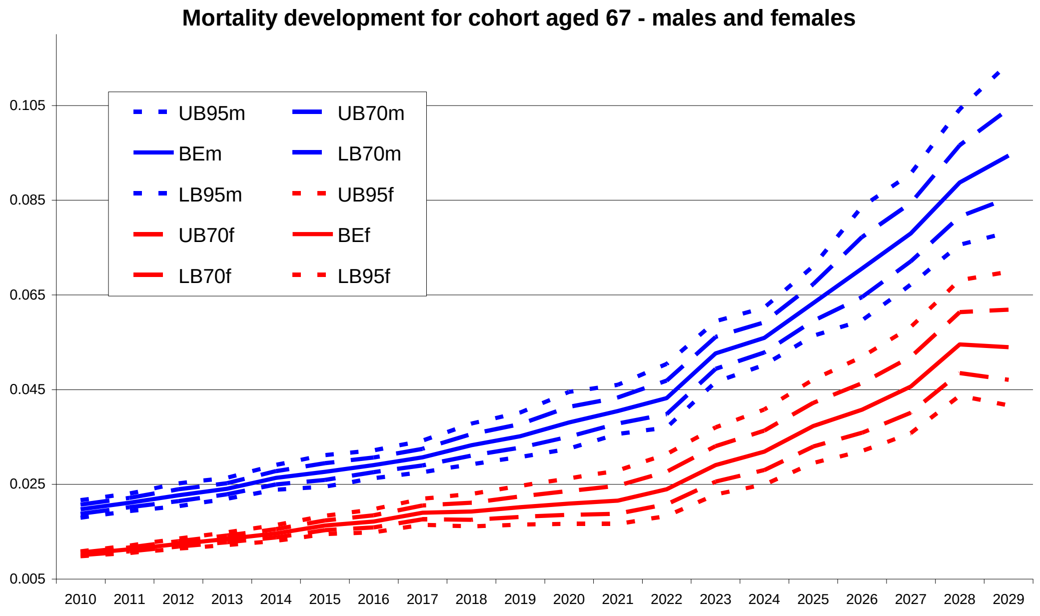
Table 7.12. Effect of mortality forecasts on annuity price, Taiwanese females

In absolute terms, the annuity prices were lower than for Japanese females (Table 6.12), and similar to those for UK (Table 4.12) and Polish females (Table 5.12), reflecting the different life expectancies in the various populations. In terms of mortality uncertainty, results were less stable than for Japanese females, similar to Polish females, and slightly more stable than for UK females.

7.3 Conclusion

To summarise the results in 7.2.1.3 and 7.2.2.3, Figure 7.6 compared the best estimate development in cohort mortality for Taiwanese males (in blue) and females (in red), along with their 70% and 95% confidence intervals between the ages of 67 and 86.

Figure 7.6. Comparison of developments in cohort mortality for males and females, for a 67-year old in 2010



LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

Unlike the results for UK pensioners in Figure 4.6, but similar to results for Poland (Figure 5.6) and Japan (Figure 6.6), the gap between the best estimate mortality for males and females is not expected to narrow, and may in fact widen. The 95% upper bound for female mortality does not even cross the 95% lower bound for male mortality. The forecasted rates were also more jagged compared to the other countries, due to the shape of actual rates discussed in Figures 7.2 and 7.4.

As for Japan, Figure 7.6 also showed graphically from the width of the bands that uncertainty around male mortality is wider than for females. This was because as in Japan, the variability in mortality rates was greater for Taiwanese males than females.

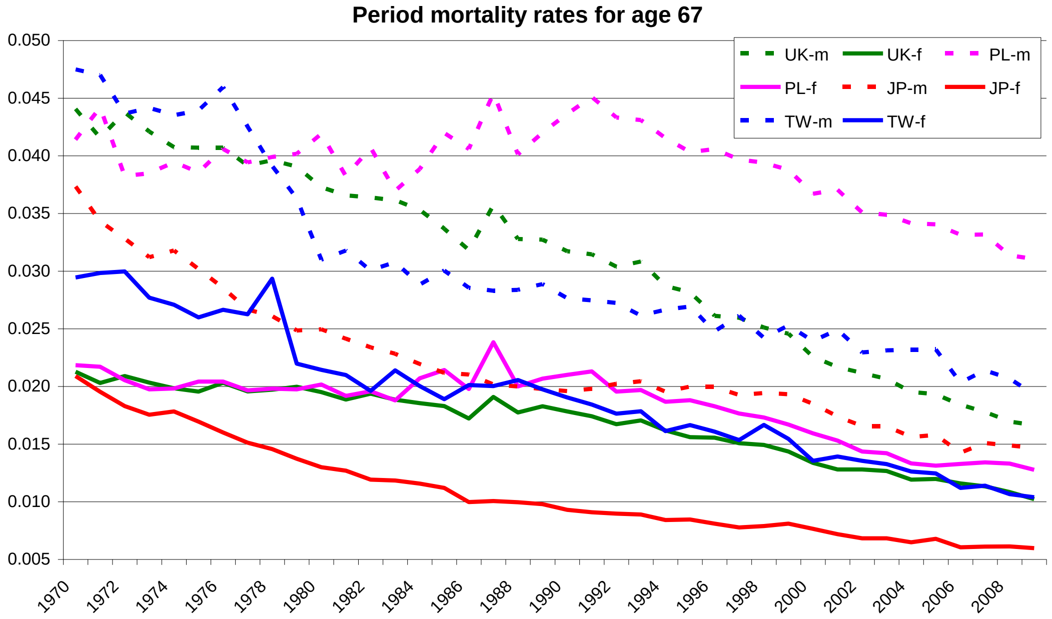
Comparing Figure 6.6 with Figure 7.6, relative to Japanese populations, there is more jaggedness in the projected mortality rates for Taiwanese populations. This is possibly due to the following reasons. Firstly, looking at historical mortality rates, for the 65-69 age sub-group, since the late 1950s the Japanese populations have experienced steady declines in mortality. Looking at Taiwanese mortality rates since 1970, mortality rates had in fact increased in some years. In the case of the 65-69 age sub-group, only after the late 1970s did mortality rates continue to decline. This naturally introduces more variability in the historical mortality pattern in Taiwan compared to Japan, which fed through into the projected mortality rates. Secondly, Japan had more years of data (63 years), compared to the 40 years for Taiwan, which gave the underlying models more years of data to fit, contributing to a smoother result.

CHAPTER EIGHT – GENERAL DISCUSSIONS AND CONCLUSION

8.1 Inter-country comparison of trends

Based on the available data from the HMD, figure 8.1 summarised the mortality rates for males and females aged 67 over the period 1970 to 2009, across both genders in the four countries considered. The year 1970 was the starting point because that was when sufficient mortality data for all four countries became available in this study.

Figure 8.1. Period mortality rates for age 67, across all eight populations



For all four countries, unsurprisingly male mortality rates remained higher than for their female counterparts. In 2009, the period mortality rate for Japanese males (the lowest amongst males) remained higher than that for Polish females (the highest amongst females).

Amongst males, Japanese males consistently showed declines in mortality and the lowest mortality in the study. UK male mortality also continued to decline, but not as rapidly as for the Japanese. Whilst Taiwanese males had the highest mortality rate in 1970, this dropped significantly in the late 1970s when economic growth began to take off. Mortality rates continued to decline thereafter, and by 2009 it was similar to the level for UK males. In contrast, Polish males actually experienced periods of higher mortality until the early 1990s (after the fall of the Berlin Wall and transition to a free-market economy began), and only thereafter were continual declines in mortality experienced over the period to 2009.

Amongst females, again, Japanese females consistently showed declines in mortality, and the lowest mortality in the group. UK female mortality also continued to decline, but not as rapidly as for the Japanese. Whilst Polish females had similar mortality rates as UK females in 1970, by 2009 Polish rates were noticeably higher, with the mortality improvement having really started only from the early 1990s (similar to Polish males). In contrast, in 1970 Taiwanese females had the highest mortality rate, but with mortality improvements over time, by 2009 their rates were very similar to those for UK females.

In summary, over the four decades since 1970, for the two developed economies, Japan and the UK, mortality improvement has been uninterrupted and steady. In contrast, for the two emerging economies Taiwan and Poland, mortality improvement began quite rapidly after the start of major economic reforms, which were in the late 1970s and early 1990s respectively.

For completeness, Figures 8.2 and 8.3 showed the period mortality rates for those aged 77 and 87 over the period 1970 to 2009.

Figure 8.2. Period mortality rates for age 77, across all eight populations

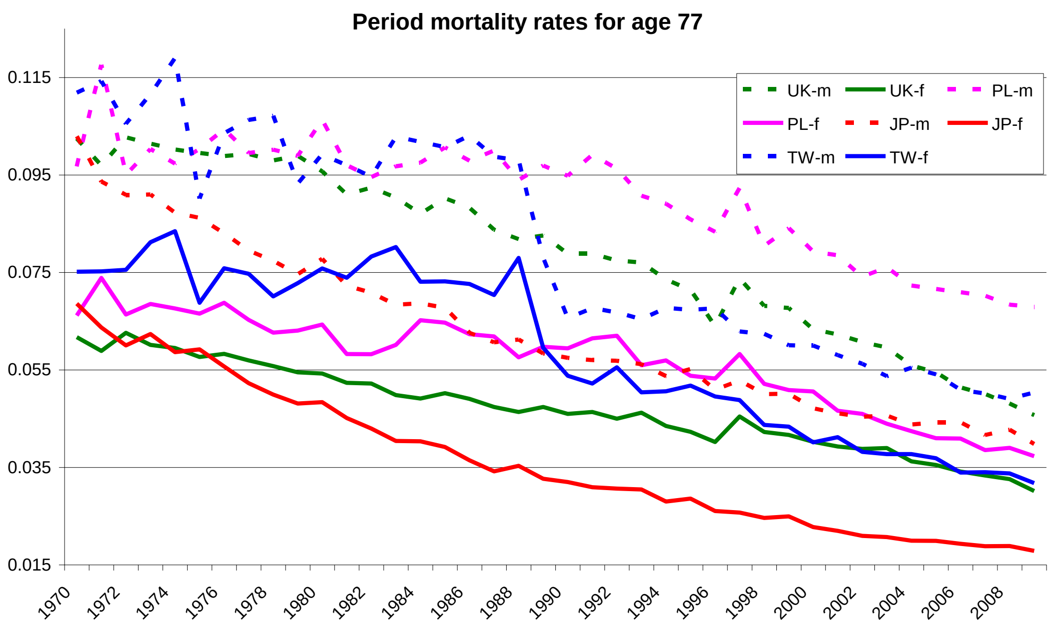
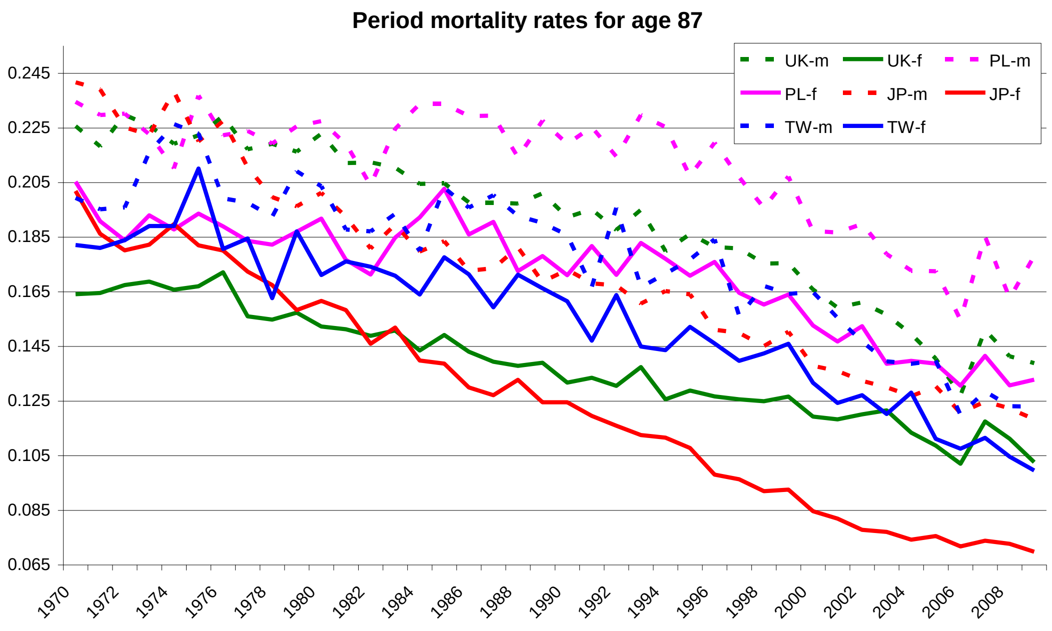


Figure 8.3. Period mortality rates for age 87, across all eight populations



Compared to those aged 67, for older ages 77 and 87, the difference in historical period mortality rates between males and females began to narrow. This development made sense, because as people began to approach the limit of human life expectancy, they were more likely to die and the mortality rate differences between genders decreased.

Again, for Japan and the UK, the comments for the 67-year-olds remained relevant for the 77- and 87-year-olds. Japan again exhibited the lowest mortality rate for all age groups, across males and females. For Taiwan, the rapid improvement in mortality over the four decades to UK levels remained apparent. The exception was with Poland, where the 77- and 87-year-olds began to experience mortality improvements earlier (from the mid-1980s) than the 67-year-olds (from the 1990s).

In all four countries, for the three ages considered, there were similarities in the patterns of mortality improvements of males and females.

8.2 Comparison of efficacy of models

To summarise the findings in Chapters 4 to 7, the model chosen for each gender in each country is shown in Table 8.1:

Model chosen	Male	Female
UK	RH	BMS
Poland	BMS	RH
Japan	BMS	BMS
Taiwan	HU	BMS

Table 8.1. Summary of models chosen for the eight populations

From Table 8.1, there was no single model that was best for all populations. Of the eight populations studied, the BMS model was found to be most appropriate for five of them (including both males and females in Japan, but was less convincing for Polish males). This was followed by the RH model, which worked well in two European populations. The HU model worked best for Taiwanese males. The well-known LC model, and the basic version of the CBD model used in this study, did not come across as the best model in any of the populations being studied.

The LC model did well in fitting to the full history of data in UK females and Polish males (where the BMS model was ultimately chosen), but fared less well when its forecasts were compared to actual experience. The LC model might not have worked well because it assumed mortality improvements to have been linear for all years (unlike the BMS model which was selective on which years met this assumption). Where the LC model appeared promising at the start, it was later outdone by the BMS model consistently, which was designed as an improvement over the LC model in the first place.

The basic version of the CBD model did well in fitting to the full history of data for UK males, and produced reasonable forecasts for UK females. Outside the UK, it did not appear to work as well. Interestingly, the CBD model was the one that consistently produced cohort mortality forecasts that were smooth and increased with age, which was observed in 7.2.2. Closer-fitting models sometimes displayed less smoothness, particularly when Taiwanese males and females were considered. The design of the basic version of the CBD model may have contributed to the smoothness, and it is repeated below from Equation 2.3 for convenience:

$\text{Logit } q(t,x) = P_t^{(1)} + (x - \bar{x})P_t^{(2)} + E(t,x)$

All else equal, the term $(x - \bar{x})$ contributes towards a higher logit value (and hence a higher mortality rate) as age x increases. None of the other four mortality models has a similar structural term that contributes towards a higher mortality rate as age increases. Hence, this aspect of the design of the basic version of the CBD model may explain why projected cohort mortality rates increased with age in every population considered.

On the other hand the basic version of the CBD model was probably not complex enough to have captured enough of the factors that affect mortality. For this reason in Cairns *et al* (2007), more complex versions of the model were introduced, which had additional cohort and quadratic period terms to capture other factors that may affect mortality. As the scope of this research was to consider some of the more commonly used mortality models internationally, the intention was not to investigate variants of a particular type of model. For this reason, only the original CBD model was considered. More advanced CBD models can be left for future research.

The RH model did not appear to work well outside UK males and Polish females, which suggests that no significant cohort effect has emerged in the other populations, which further suggests that across populations the drivers of mortality can be different. As a reference, the significance of the cohort effect in English and Welsh males was found in Cairns *et al* (2007).

Although the BMS model was already chosen as the most appropriate in five of the eight populations (in addition to being a close second for Taiwanese males), it also worked relatively well in UK males and Polish females (where the RH model was chosen). Whilst the forecasted mortality rates from the BMS model were not as close to actual experience as those produced by the RH model, the BMS model did well in fitting to the full history of data in those two populations. Besides being a clear model for Japan, it seemed to do reasonably well in every population considered in this study.

Where the BMS model did well, the HU model sometimes did well too, as seen in Taiwanese females and males. The HU model also fitted well to the full history of data in Polish males and females. However, it did not do as well in Japan and the UK. In this study, the HU model appeared to have done better in the emerging economies, represented by Taiwan and Poland for the purpose of this research. This may be because as its design incorporates a large number of parameters, the HU model is better structured to capture the greater variability of mortality rates that pertains to emerging economies.

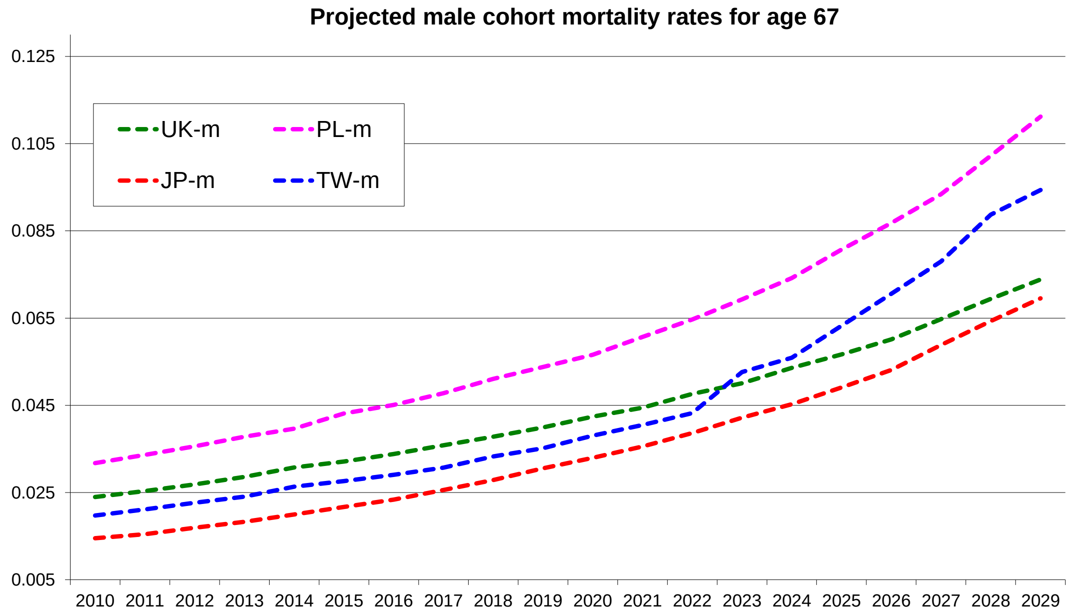
Considered differently, amongst the four populations in the developed economies (UK and Japan), the BMS model was most appropriate for three of them, including for the female populations of both countries. This may be because the rate of mortality improvement in developed economies has during the last quarter of the twentieth century stabilised to a level of being close to linear, which could be captured well by the BMS model.

In contrast, in the emerging economies (Poland and Taiwan) where mortality improvements were more variable, the BMS model was best for only two of the four populations, and was not particularly preferable for one gender.

Lastly, from a geographical perspective, in the European populations (UK and Poland), it was either the BMS or RH model that worked well. In the Asian populations (Japan and Taiwan), it was mainly the BMS, followed by the HU model.

Figure 8.4 summarised the projected best estimate cohort mortality rates for males aged 67 in 2010, across the four countries in this study. The steepness of the gradient reflects the speed at which mortality rates increase with age.

Figure 8.4. Projected male cohort mortality rates for age 67, across four countries

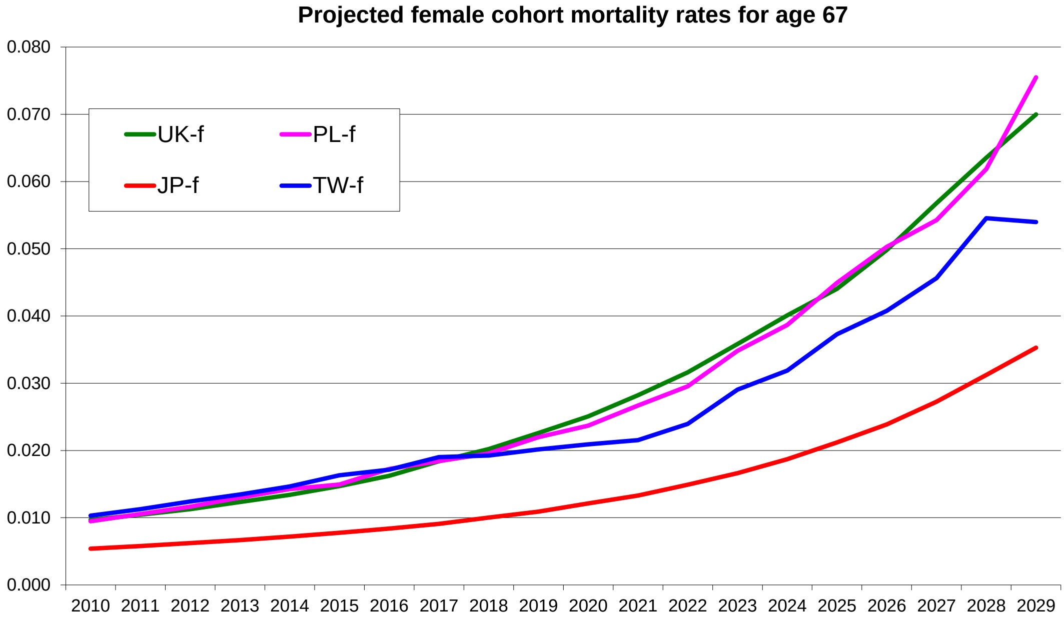


Japanese males were expected to experience the lowest mortality rates throughout the projection period, and Polish males the highest. Although Taiwanese males aged 67 started with lower mortality than UK males, by 2023 when the survivors are aged 80, they are expected to experience higher mortality than UK survivors aged 80. It could be seen that the gradients for developed economies the UK and Japan were less steep than for the emerging economies Poland and Taiwan. This meant that the rate of increase in male mortality rates in old age was less in developed than emerging economies.

The progression of cohort mortality rates for the developed economies Japan and the UK were also close to linear, with increases in mortality rates being fairly steady. Whilst the progression of mortality rates for the emerging economies Taiwan and Poland were initially also not far from linear, after about ten years the rates began to increase quite sharply. This suggests that the mortality rates of the very old (beyond age 80) were expected to be higher in the emerging economies in this study. Before age 80, the mortality rate could even be lower than in developed economies (as in Taiwan versus the UK).

Figure 8.5 summarised the projected best estimate cohort mortality rates for females aged 67 in 2010, across the four countries in this study.

Figure 8.5. Projected female cohort mortality rates for age 67, across four countries

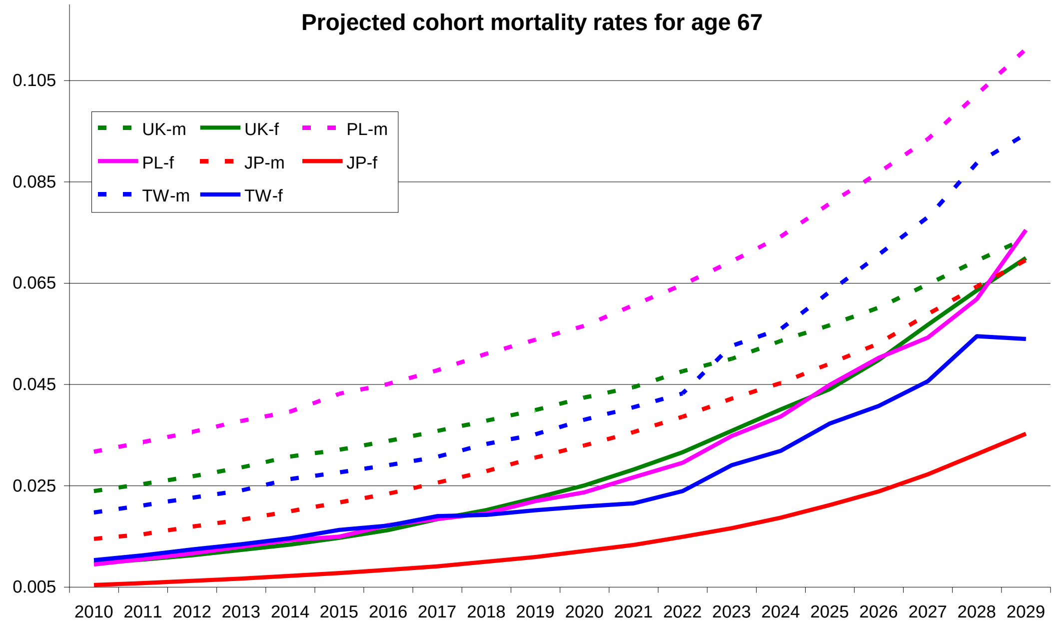


For females, again Japan was expected to experience the lowest mortality rates throughout the projection period. Similar to its male counterpart, the rate of increase in female old-age mortality with age was the least for Japan. Taiwan was expected to experience the next lowest rates in general (after looking beyond the first eight years), which is an emerging economy and another Asian country. Perhaps surprisingly, UK and Polish female mortality rates (not far from the rates for Japanese males) were expected to be very similar over the next two decades, despite their difference in economic development.

Except for Japanese females over the first ten years, none of the progression in female cohort mortality could be regarded as being close to linear. This was largely a function of the low base of mortality rate from which all four countries started (which was not more than 1% at age 67). As females age, inevitably the associated mortality rates need to increase to reflect the reality of there being a limit to human life. Hence the increases in mortality rates appeared quadratic, even exponential, relative to the low base at age 67. The exception in Japan was caused by the increases in mortality rates being minimal, until the individual approached age 80, and thereafter increases began to pick up.

Figure 8.6 combined the output in Figures 8.4 and 8.5 in one graph, including males and females.

Figure 8.6. Projected cohort mortality rates for age 67, across eight populations



Cohort mortality rates of males were generally expected to be higher than of females throughout the projection period. The exception was with Polish female rates exceeding Japanese and UK males towards the end, and UK females converging towards Japanese male rates.

Towards the end of the projection, the gap between UK male and female rates was also noticeably small, especially compared to that of the other three countries. The UK was the only country where male rates appeared to be converging towards those of females, as discussed in section 4.3.

Relative to females, the progression of male cohort mortality rates were closer to being linear. This was because the starting mortality rates for males were higher. As both males and females approached the natural limit to human life expectancy, the increases in female rates needed to be faster to compensate for their lower starting rates, and those for males could be more gradual. This largely explained the “close-to-linearity” shape of male rates.

Whilst the projections revealed interesting trends, it should be noted that these are based on extrapolations of historical trends, using the model deemed most appropriate for that population. For developed economies UK and Japan, more years of data were used in the fitting process (63 years, from 1947 to 2009). In contrast, for the emerging economies, fewer years of data were available, with 52 years for Poland (from 1958 to 2009) and 40 years for Taiwan (from 1970 to 2009). In addition to differences in data availability between the developed and emerging economies, developed economies also began to experience steady continual mortality improvements earlier. In contrast, in the emerging economies mortality rates even increased in some years, before mortality improvements (which developed economies had experienced many years earlier) began to take hold. Consistent mortality improvements in emerging economies appeared as economic growth accelerated and standards of living improved. The different phases of economic and mortality development encapsulated in the datasets for emerging economies meant that there was more variability in the pattern of mortality rates for the emerging economies, compared to their developed counterparts. This variability fed through into the projections from the mortality models, explaining why there was more jaggedness in the mortality projections for emerging economies.

From Figure 8.6, the developed economies with longer and less variable histories of mortality data were associated with smoother projections with less jaggedness. At the same time some measure of caution is needed in the interpretation of these projections. If the current population is materially different from the population on which the extrapolation was done (for example in terms of economic wealth, access to healthcare and nutrition, genetics/ethnicity), then the quality of the extrapolation may be undermined.

8.3 Comparison of theoretical annuity prices and stability of mortality forecasts

Table 8.2 is a summary of the prices of theoretical 20-year level annuities described in 1.5.4, for the eight populations under different mortality scenarios. As all were calculated on an interest rate of 1.5%, assuming payment of one pound as income each year, differences in this theoretical price are closely linked to differences in the life expectancy of a 67-year-old.

Annuity Prices	**95% LB**	**70% LB**	**BE**	**70% UB**	**95% UB**
UK males	12.6	12.5	12.3	12.2	12.1
Poland males	11.6	11.4	11.1	10.9	10.6
Japan males	13.9	13.7	13.5	13.2	13.0
Taiwan males	13.1	12.9	12.7	12.5	12.3
UK females	14.7	14.5	14.3	14.0	13.8
Poland females	14.6	14.5	14.3	14.1	13.9
Japan females	15.8	15.7	15.6	15.4	15.3
Taiwan females	14.8	14.6	14.4	14.2	14.0

LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

Table 8.2. Comparison of theoretical annuity prices across eight populations, at an interest rate of 1.5%

From Table 8.2, Japanese females are expected to enjoy the longest life expectancy, followed by Taiwanese females. Expectations are similar for UK and Polish females, with there being more variability for the former.

Any of the female populations considered is expected to live for longer than all the male populations.

Within males, the same pattern continued, with Japanese males expected to live for the longest, followed by Taiwanese males. UK males are close behind, followed by Polish males.

It was interesting to note that for males and females, the Asian populations were expected to live for longer than the European ones in this study. As an observation, the Asian populations were also genetically and culturally more homogeneous, being less affected by immigrations and emigrations compared to the European populations. Tsuda (2001) noted that for many decades in the Twentieth Century, Japan was the only advanced industrial country that did not rely on unskilled foreign labour, largely because of its insistence on ethnic homogeneity and refusal to accept unskilled foreign workers. However, over time this may change as Japan grapples with maintaining economic growth for a rapidly ageing society. With the labour force expected to decrease by 10% in 25 years, Japan may need to attract 600 000 immigrants each year simply to maintain the current size of its workforce. Ma (2007) described the history of immigration to Taiwan over the last four hundred years, and how the overwhelming majority of the population in Taiwan can trace their origins back to the south-eastern part of mainland China at some point in history.

While the role of ethnicity and genetics is beyond the scope of this research, as an after-thought, Murray *et al* (2006) found that in the United States, Asian Americans (born in the United States) generally enjoyed the lowest mortality rates and highest life expectancies relative to other ethnic groups. This applied to both males and females. The authors did caution that the observed disparities in life expectancy cannot be explained by race, income or access to healthcare alone.

In contrast to the relative homogeneity of the Asian populations, the UK has become a more multi-cultural country in recent history. Migration Watch UK (2006) commented that there was relatively little migration into the UK until the 1950s, when immigration from the rest of the Commonwealth picked up, at a rate of 500 000 people each decade. Until 1983, this inflow was counterbalanced by emigration out of the UK. Since then the UK has seen a net inflow of migrants into the country. The ethnic population from the rest of the Commonwealth was estimated at 7% of the population of England and Wales in 2006.

More recently, Cangiano (2012) reported that half of the increase of the UK population between 1991 and 2010 (2.4 million people) was due to net immigration into the UK. It was noted that without net immigration, Scotland's population would stagnate over the next twenty years and eventually decrease. Since 1998, net immigration (of around 200 000 people a year) has contributed more to population growth in the UK than births. During this period, the increase in population from net migration exceeded births by about 40 000 each year. In the 2011 Census published by the Office for National Statistics (2012b), it was noted that 13% of residents (7.5 million people) in England and Wales are born outside the UK. Of these, just over half had arrived in the last 10 years ending in March 2011, exceeding the combined number of immigrants who entered the UK between 1940 and 2000. Amongst the foreign-born residents, the top three countries of birth are India, Poland and Pakistan, together accounting for 23% of foreign-born residents. Of all non-UK countries, over the last decade, Poland showed by far the largest percentage increase as a country of birth, with a nine-fold rise over the last decade following its accession to the European Union in 2004, accounting for about 8% of foreign-born residents (580 000 people) in 2011. All these developments point to a noticeable change in the UK genetic pool over time. In terms of what effect this would have on future mortality, this is something to be left for future researchers.

On the other hand, Dustmann, Frattini and Rosso (2012) documented that since the late 1990s, Poland has experienced a high increase in emigration. By 2007 working-age (age 15 to 65) emigrants accounted for over 2% of the total registered working-age Polish population. Most of the recent emigrants reside in the UK, followed by Germany. The flow of emigrants was partly attributable to Poland becoming a member of the European Union in 2004, giving its citizens the right to work in other European Union countries. It also remains to be seen if the emigration has a selection effect on the genetic pool of those remaining in Poland, and to what extent this may affect future mortality developments. Again, this is beyond the scope of this research.

From Table 8.2, based on the lower annuity prices expected for UK males and females relative to their Taiwanese counterparts, perhaps contrary to perception, populations from developed economies are not necessarily expected to outlive some populations from emerging economies.

Table 8.3 is a summary of the relative price differences from the best estimate mortality case, across the eight populations.

% Price Difference	**95% LB**	**70% LB**	**70% UB**	**95% UB**
UK males	2.1%	1.1%	-1.1%	-2.1%
Poland males	4.4%	2.4%	-2.4%	-4.6%
Japan males	3.3%	1.8%	-1.9%	-3.6%
Taiwan males	3.1%	1.7%	-1.7%	-3.4%
UK females	2.9%	1.6%	-1.7%	-3.3%
Poland females	2.5%	1.3%	-1.4%	-2.8%
Japan females	1.5%	0.8%	-0.9%	-1.8%
Taiwan females	2.6%	1.4%	-1.5%	-3.0%

LB = Lower Bound; UB = Upper Bound

Table 8.3. Comparison of variation to annuity prices across eight populations, at an interest rate of 1.5%

In general, less variability is expected around female mortality projections than males. Between developed and emerging economies, the developed economies had annuity prices that were less variable, largely due to the contribution from Japanese females and UK males.

The interest rate of 1.5% used was then varied to assess how the theoretical annuity prices and percentage price differences would change. For simplicity, the rate was decreased and increased by 1%, and alternative rates of 0.5% and 2.5% were used.

Table 8.4 shows the annuity prices in the eight populations calculated using an interest rate of 0.5%.

Annuity Prices	**95% LB**	**70% LB**	**BE**	**70% UB**	**95% UB**
UK males	13.7	13.6	13.5	13.3	13.2
Poland males	12.6	12.4	12.1	11.8	11.5
Japan males	15.2	15.0	14.7	14.4	14.2
Taiwan males	14.3	14.1	13.9	13.6	13.4
UK females	16.1	15.9	15.6	15.3	15.1
Poland females	16.0	15.9	15.6	15.4	15.2
Japan females	17.4	17.3	17.2	17.0	16.8
Taiwan females	16.2	16.0	15.8	15.5	15.3

LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

Table 8.4. Comparison of theoretical annuity prices across eight populations, at an interest rate of 0.5%

Not surprisingly, by lowering the pricing interest rate from 1.5% to 0.5%, the absolute prices of annuities increased for all populations. The order of price magnitude across the populations had also not changed from Table 8.2. The percentage price difference between using an interest rate of 1.5% and 0.5% is shown in Table 8.8.

Table 8.5 is similar to Table 8.3 in considering the variation to annuity prices from the best estimate case under different mortality scenarios, but at an interest rate of 0.5%.

% Price Difference	**95% LB**	**70% LB**	**70% UB**	**95% UB**
UK males	2.2%	1.2%	-1.2%	-2.2%
Poland males	4.7%	2.5%	-2.6%	-4.9%
Japan males	3.5%	1.9%	-2.0%	-3.9%
Taiwan males	3.3%	1.8%	-1.8%	-3.5%
UK females	3.1%	1.7%	-1.8%	-3.5%
Poland females	2.6%	1.4%	-1.5%	-2.9%
Japan females	1.6%	0.9%	-1.0%	-1.9%
Taiwan females	2.7%	1.5%	-1.6%	-3.1%

LB = Lower Bound; UB = Upper Bound

Table 8.5. Comparison of variation to annuity prices across eight populations, at an interest rate of 0.5%

Compared to Table 8.3, under different mortality scenarios, the percentage change in annuity price from the best estimate scenario had increased slightly, due to the effect of the lower pricing interest rate. However, the order in terms of size of difference across the various populations had not changed from Table 8.3.

Table 8.6 shows the annuity prices in the eight populations calculated using an interest rate of 2.5%.

Annuity Prices	95% LB	70% LB	BE	70% UB	95% UB
UK males	11.6	11.5	11.4	11.2	11.1
Poland males	10.7	10.5	10.3	10.1	9.8
Japan males	12.7	12.6	12.4	12.1	11.9
Taiwan males	12.0	11.9	11.7	11.5	11.3
UK females	13.4	13.3	13.1	12.8	12.6
Poland females	13.4	13.2	13.1	12.9	12.7
Japan females	14.4	14.3	14.2	14.1	14.0
Taiwan females	13.5	13.3	13.2	13.0	12.8

LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

Table 8.6. Comparison of theoretical annuity prices across eight populations, at an interest rate of 2.5%

Not surprisingly, by increasing the pricing interest rate from 1.5% to 2.5%, the absolute prices of annuities decreased for all populations. The order of annuity price magnitude across the populations has also not changed from Table 8.2 or Table 8.4. The percentage price difference between using an interest rate of 1.5% and 2.5% is discussed in Table 8.8.

Table 8.7 is similar to Tables 8.3 and 8.5 in considering the variation to annuity prices from the best estimate case under different mortality scenarios, but at an interest rate of 2.5%.

% Price Difference	95% LB	70% LB	70% UB	95% UB
UK males	1.9%	1.0%	-1.1%	-2.0%
Poland males	4.2%	2.2%	-2.3%	-4.3%
Japan males	3.1%	1.7%	-1.8%	-3.4%
Taiwan males	3.0%	1.6%	-1.7%	-3.2%
UK females	2.8%	1.5%	-1.6%	-3.2%
Poland females	2.3%	1.3%	-1.3%	-2.6%
Japan females	1.4%	0.8%	-0.9%	-1.7%
Taiwan females	2.4%	1.3%	-1.4%	-2.8%

LB = Lower Bound; UB = Upper Bound

Table 8.7. Comparison of variation to annuity prices across eight populations, at an interest rate of 2.5%

Compared to Table 8.3, under different mortality scenarios, the percentage change in annuity price from the best estimate scenario had decreased slightly, due to the effect of the higher pricing interest rate. However, the order of magnitude of difference across the various populations had also not changed from Table 8.3 or Table 8.5.

Table 8.8 considers the percentage price change in the annuity price due to different interest rates. The price of the annuities under the best estimate, 95% lower bound and 95% upper bound mortality scenarios are considered. The price calculated under the 0.5% and 2.5% interest rates are compared against that using the 1.5% rate.

There is an inverse relationship between the rate of interest and the annuity price. The annuity price is higher when the rate is lower, and the price is lower when the rate is higher. This is because when the rate is lower, the return on investment is lower, more money is needed upfront to meet the same stream of cashflow payments and the price charged by the insurance company is higher. Similarly when the rate is higher the price charged is lower.

To illustrate this, in the example given by Table 8.8, the annuity price using a 0.5% interest rate is 9.4% more expensive than using 1.5%, under the best estimate mortality scenario for UK males. When a rate of 2.5% is used, the price is 7.7% cheaper than using 1.5%. Similarly, the annuity price using a 0.5% interest rate is 9.1% more expensive than using 1.5%, under the 95% lower bound mortality scenario for UK males. When a rate of 2.5% is used, the price is 8.1% cheaper than using 1.5%.

% change in prices	**95% LB, 0.5%**	**BE, 0.5%**	**95% UB, 0.5%**	**95% LB, 2.5%**	**BE, 2.5%**	**95% UB, 2.5%**
UK males	9.1%	9.4%	8.7%	-8.1%	-7.7%	-8.1%
Poland males	9.0%	8.9%	8.4%	-7.6%	-7.3%	-7.1%
Japan males	9.6%	9.1%	8.9%	-8.4%	-8.5%	-8.2%
Taiwan males	9.3%	9.2%	8.8%	-8.1%	-8.0%	-8.0%
UK females	9.6%	9.2%	9.2%	-8.7%	-8.7%	-8.4%
Poland females	9.9%	9.3%	9.2%	-8.4%	-8.7%	-8.5%
Japan females	10.4%	10.0%	10.0%	-8.7%	-8.9%	-8.7%
Taiwan females	9.7%	9.7%	9.3%	-8.8%	-8.5%	-8.5%

LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

Table 8.8. Comparison of variation to annuity prices across eight populations, at an interest rate of 0.5% and 2.5%, relative to 1.5%

From Table 8.8, at a low interest rate level of 1.5%, annuity prices in all populations are more sensitive to further decreases in interest rates than increases. This could be seen in the percentage increase in price being greater than the decrease in all mortality scenarios, despite an equal 1% change in interest rates in either direction. Populations with higher annuity prices (for example Japanese and Taiwanese females) are also more sensitive than other populations to movements in interest rates, exhibiting greater percentage price fluctuations in response to interest rate movements. For example, the range of fluctuation in annuity price under the best estimate mortality scenario for Japanese females would be 18.9% (=10% to (-8.9%)), whilst that for Polish males would be smaller at 16.2% (=8.9% to (-7.3%)).

Table 8.9 considers the range of fluctuation in price for all populations, which is derived by subtracting the percentage price change under the 2.5% rate scenario from that under 0.5%. This was done for the best estimate, 95% lower bound and upper bound mortality scenarios.

Fluctuation in price	**95% LB**	**BE**	**95% UB**
UK males	17.2%	17.1%	16.8%
Poland males	16.6%	16.1%	15.6%
Japan males	18.0%	17.6%	17.1%
Taiwan males	17.4%	17.1%	16.8%
UK females	18.3%	17.9%	17.6%
Poland females	18.3%	18.0%	17.7%
Japan females	19.1%	18.9%	18.7%
Taiwan females	18.5%	18.2%	17.8%

LB = Lower Bound; BE = Best Estimate; UB = Upper Bound

Table 8.9. Fluctuation in annuity price with rates ranging from 0.5% to 2.5%

From Table 8.9, the order of size of fluctuations was almost identical to the orders of the size of annuity prices in Tables 8.2, 8.4 and 8.6. This supports the observation that populations with longer life expectancies attracting higher annuity prices are also more sensitive to movements in interest rates. In addition, across all populations the fluctuation was greatest under the 95% lower bound mortality scenario, and least under the 95% upper bound scenario. This is because at high levels of life expectancy (represented by the 95% lower bound mortality scenario), more payments are expected from the annuity product, and interest rates play a greater effect on the size of the annuity price.

8.4 Conclusion

Although only four countries were chosen for this study, many interesting insights were gained, particularly for the 65-69 age sub-group, for which forward-looking projections outside the data availability period were made. Arguably, this sub-group remains of most interest to actuaries involved in the pricing of annuities or calculations of capital requirements. This is because this group is the largest five-year age sub-group by the number of people, and tends to be the first-time buyers of annuity products.

In general, the BMS model tended to work well, particularly for females, in Asian countries or in developed economies. This was followed by the RH model which worked well in European countries.

Whilst it was well-known that females are expected to outlive males, it was interesting that even females in the emerging economies selected were expected to outlive males in the developed economies selected. Variations in female mortality are also expected to be less than for males. The UK appeared to be the only country where male mortality rates are expected to converge towards female rates over the forecast period in the age range considered.

Intriguing was the result that both males and females of the two Asian retiree populations selected were expected to outlive their European counterparts. Annuity prices for the Asian populations were therefore also higher than for the European ones. Surprisingly, although the UK is a developed economy, its old-age population is not necessarily expected to outlive that of an emerging economy like Taiwan.

Although retirees in some emerging economies may outlive those in some developed economies, for the very old (survivors beyond age 80), mortality rates are still expected to be lower in the developed economies.

In terms of uncertainty around mortality, in developed economies the magnitude of this is expected to be less than in emerging economies. When expressed financially as a price variation around the best estimate annuity price, at extreme mortality scenarios the variation was on average around 3% for the developed economies, and on average close to 3.5% for emerging economies. Between Asia and Europe, the variation was just above 3% for Europe and under 3% for Asia. Between males and females, the variation was higher on average for males at close to 3.5%, and lower for females at around 2.5% on average.

Populations with longer life expectancies (which attract higher annuity prices) are also more sensitive than other populations with shorter life expectancies to changes in interest rates, particularly to further decreases in rates. For these other populations, if mortality rates turn out to be lower and life expectancies longer than expected, then interest rates will also play a bigger role in the magnitude of the annuity price. This is because a longer lifespan extends the length of cashflows to be paid under an annuity. When this stream of cashflows becomes longer, the annuity price to charge becomes more sensitive to the return on investment or interest rate attainable in the market.

An environment marked by low (and possibly lower) interest rates and improving (and possibly faster improving) mortality rates would therefore demand careful attention on the pricing and reserving of annuity products and benefits.

8.5 Room for further research

This research was intended as a starting point in applying different stochastic mortality models to populations across different continents in different stages of economic development. For now, only four countries were selected, but it would be straightforward to extend the type of analysis conducted in this research to other countries, when data of sufficient quality for them become available.

This research had only considered the original basic form of the CBD model. Extended versions were described in Cairns *et al* (2007), which could be tried at a later stage. Hence, for future research, it would be useful to consider variants of the CBD model on similar datasets.

In assessing the adequacy of out-of-sample forecasts, a forecast window of 20 years was chosen for consistency across the four economies. The choice was also due to the limited availability of data in the emerging economies, and the need to have the forecast period being no longer than the fitting period for statistical reasons. It is possible to change this to a different number of years, and varied by population.

REFERENCES

A'Hearn, B., Baten, J. and Crayen, D. (2006). Quantifying quantitative literacy: age heaping and the history of human capital. Working paper 996, Universitat Pompeu Fabra, November 2006. Downloaded at http://www.econ.upf.edu/en/research/onepaper.php?id=996

Alho, J. and Spencer, B. (2005). *Statistical Demography and Forecasting*. Springer. New York.

BBC (2005). Short history of immigration, BBC News. Downloaded at http://news.bbc.co.uk/hi/english/static/in_depth/uk/2002/race/short_history_of_immigration.stm

BBC (2009). Females less physically active, BBC News. Downloaded at http://news.bbc.co.uk/2/hi/health/7811398.stm

BBC (2011). Compulsory retirement age at 65 fully abolished, 1 October 2011. Downloaded at http://www.bbc.co.uk/news/business-15127835

BBC (2012). Polish MPs approve plans to raise retirement age to 67, 11 May 2012. Downloaded at http://www.bbc.co.uk/news/world-europe-18032956

Bekaert, G., Wei, M. and Xing, Y. (2002). Uncovered interest rate parity and the term structure. National Bureau of Economic Research working paper 8795, downloaded from www.nber.org/papers/w8795, February 2002

Bell, W. (1997). Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. *Journal of Official Statistics*, 13(3), 279-303

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

Booth, H., Tickle, L. and Smith, L. (2005). Evaluation of the variants of the Lee-Carter method of forecasting mortality: a multi-country comparison. *New Zealand Population Review, Special issue on stochastic population projections*, 31(1), 13-37

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, 289-310

Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: a review of methods. *Annals of Actuarial Science* 3 I/II 3-44

British Actuarial Journal. (2009), 15, Supplement 1. Institute and Faculty of Actuaries.

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

Bunting, M. (2001). Haven't we been here before? The Guardian, 23 May 2001. Downloaded from http://www.guardian.co.uk/uk/2001/may/23/immigration.immigrationandpublicservices11

Cairns, A. (2007). LifeMetrics software user guide, version 2. Code written by Professor Andrews Cairns, Heriot-Watt University, 2007. Available from JP Morgan LifeMetrics team, www.jpmorgan.com/lifemetrics

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

Cairns, A., Blake, D., Dowd, K., Coughlan, G., Epstein, D., Ong, A. and Balevich, I. (2007). A quantitative comparison of stochastic mortality models using data from England & Wales and the United States. The Pensions Institute. Abridged version published on *NAAJ* (2009) V13(1) 1-35

Cangiano, A. (2012). The impact of migration on UK population growth. The Migration Observatory at the University of Oxford, January 2012. Available at www.migrationobservatory.ox.ac.uk

Canudas-Romo, V. (2010). Three measures of longevity: time trends and record values. *Demography* 47(2) 299-312

Currie, I. (2006). Smoothing and forecasting mortality rates with P-splines. Talk given at the Institute of Actuaries, April 2007, available at http://www.ma.hw.ac.uk/~iain/research/talks.html

Currie, I., Durban, M. and Eilers, P. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, 4, 279-298

European Insurance and Occupational Pensions Authority (2011). EIOPA Report on the fifth Quantitative Impact Study (QIS5) for Solvency II. Downloaded from https://eiopa.europe.eu

Faculty of Actuaries and Institute of Actuaries (2002). Formulae and tables for examinations of the Faculty of Actuaries and the Institute of Actuaries.

Financial Services Board (2012). Solvency Assessment and Management, Second South African Quantitative Impact Study, Technical Specifications. Downloaded from www.fsb.co.za

Furukawa, T. (1996). Mathematical Theories on Longevity. Asakura bookstore, Tokyo, Japan.

Government Actuary’s Department (2001). National population projections: review of methodology for projecting mortality. Government Actuary’s Department, United Kingdom.

Hogg, R. and Tanis, E. (2001). *Probability and statistical inference*. Sixth Edition. Prentice Hall, New Jersey.

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org (data downloaded on 19 March 2012).

Hyndman, R. (2011a). “Demography” package user guide. Code written by Professor Rob Hyndman, Monash University, 2011, available from http://robjhyndman.com/software/demography

Hyndman, R. (2011b). “Forecast” package user guide. Code written by Professor Rob Hyndman, Monash University, 2011, available from http://robjhyndman.com/software/forecast

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

Hyndman, R., Booth, H. and Yasmeen, F. (2011). Coherent mortality forecasting: the product-ratio method with functional time series models, Working paper 01/11. Monash University.

International Monetary Fund (2012a). Global Financial Stability Report, Chapter 4: The Financial Impact of Longevity Risk, April 2012.
Downloaded from http://www.imf.org/external/pubs/ft/gfsr/2012/01/index.htm

International Monetary Fund (2012b). World Economic Outlook: Coping with High Debt and Sluggish Growth, October 2012. Downloaded at http://www.imf.org

International Monetary Fund (2012c). World Economic Outlook Database October 2012. Downloaded at http://www.imf.org/external/pubs/ft/weo/2012/02/weodata/index.aspx

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

Lee, R. and Carter, L. (1992). Modelling and forecasting US mortality. *Journal of the American Statistical Association*, 87, 659-675

Li, S. (2010). Pricing longevity risk with the parametric bootstrap: a maximum entropy approach. *Insurance: Mathematics and Economics*, 47(2), 176-86

Li, S. and Chan, W. (2005). Outlier analysis and mortality forecasting: the United Kingdom and Scandinavian countries. *Scandinavian Actuarial Journal*, 3, 187-211

Li, S. and Chan, W. (2011). Simultaneous prediction intervals: an application to forecasting US and Canadian mortality. Presented at the “Living to 100” Symposium at Orlando, Florida, organised by the Society of Actuaries, January 2011

Li, S., Hardy, M. and Tan, K. (2009). Uncertainty in mortality forecasting: an extension to the classical Lee-Carter approach. *ASTIN* bulletin 39(1) 137-64

Li, S., Hardy, M. and Tan, K. (2010). Developing mortality improvement formulas: the Canadian insured lives case study. *NAAJ* 14(4).

Li, S., Chan W. and Cheung, S. (2011). Structural changes in the Lee-Carter mortality indexes: detection and implications. *NAAJ* 15(1).

Ma, Y.J. (2007). The Seven Waves of Immigration in Taiwan. China Times, 14 June 2007. Downloaded at http://www.zonaeuropa.com/20070614_1.htm

Matysiak, A. and Nowok, B. (2007). Stochastic forecast of the population of Poland, 2005 – 2050. *Demographic Research*, 17, 301-338

Mayhew, L. (2012). Research conducted for the UK's Office of National Statistics. News report from the BBC in 2012, http://www.bbc.co.uk/news/health-17811732

McNown, R. and Rogers, A. (1992). Forecasting cause-specific mortality using time series methods. *International Journal of Forecasting*, 8(3), 413-432

Migration Watch UK (2006). The history of immigration to the UK. Briefing paper 6.1, downloaded at www.migrationwatchuk.org

Murray, CJL., Kulkarni, SC., Michaud, C., Tomijima, N., Bulzacchelli, MT., Iandiorio, TJ. and Ezzati, M. (2006). Eight Americas: investigating mortality disparities across races, counties, and race-counties in the United States. PLoS Med 3(9): e260.

O'Connell, A. (2011). Longevity Bulletin Issue 02 – November 2011. Institute and Faculty of Actuaries. Available from longevitybulletin@actuaries.org.uk.

Office for National Statistics (2012a). Mortality, 2010-based NPP reference volume. Downloaded from www.ons.gov.uk

Office for National Statistics (2012b). 2011 Census: key statistics for England and Wales, March 2011. Downloaded from www.ons.gov.uk

Ozeki, M. (2005). Application of mortality models to Japan. *Presented at the Living to 100 and Beyond Symposium, sponsored by the Society of Actuaries, Florida, 2005*.

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

Shang, H., Booth, H. and Hyndman, R. (2011). Point and interval forecasts of mortality rates and life expectancy: a comparison of ten principal component methods. *Demographic Research* 25, 173 - 214

Shaw, C. (2007). Fifty years of United Kingdom national population projections: how accurate have they been? Population trends 128, Summer 2007. Office of National Statistics, www.ons.gov.uk

Sweeting, P. (2011). A trend-change extension of the Cairns-Blake-Dowd model. *Annals of Actuarial Science* 5(2), 143 - 162

Tabeau, E., Ekamper, P., Huisman, C. and Bosch, A. (2001). Predicting mortality from period, cohort or cause-specific trends: a study of four European countries. In (E. Tabeau, A. van den Berg Jeths and C. Heathcote, eds.) *Forecasting mortality in developed countries: insights from a statistical, demographic and epidemiological perspective* (159-187). Kluwer Academic Publishers, Dordrecht.

The China Post (2008). Retirement age to be raised to 65, 19 April 2008, downloaded from http://www.chinapost.com.tw/taiwan/national/national%20news/2008/04/19/152582/Retirement-age.htm

Tsuda, T. (2001). Reluctant hosts: the future of Japan as a country of immigration. University of California, San Diego. Downloaded at http://migration.ucdavis.edu/rs/more.php?id=39_0_3_0

Wang, C. and Liu, Y. (2010). Comparisons of mortality modelling and forecasting – empirical evidence from Taiwan. *International Research Journal of Finance and Economics*. Issue 37, 46 – 55.

Wang, J., Huang, H., Yang, S. and Tsai, J. (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.

Warnock, E. (2012). Japan bill to raise retirement age passes Lower House. The Wall Street Journal, Asia, 2 August 2012. Downloaded from http://online.wsj.com/article/SB10000872396390443687504577564800240852924.html

World Economic Forum (2010). Longevity Round Table, 25 August 2010. Downloaded from http://www.ageing.ox.ac.uk/files/WEF%20Longevity%20Rountable%20Summary%20-%20Final-2.pdf

World Economic Forum (2012). Global Population Ageing: Peril or Promise? Global Agenda Council on Ageing Society. 2012. Downloaded from http://www.weforum.org/reports/global-population-ageing-peril-or-promise

World Health Organisation (2008). The World Health Report 2008. Downloaded from http://www.who.int/whr/2008/en/

APPENDICES

Appendix 1. Format of mortality data from the HMD

Appendix 1 shows the format of the mortality data used in the study. It takes the form of a matrix, with the rows being ages and columns the years. In this example the UK male period mortality rates derived from the HMD for ages 65 to 89, over the years 1947 to 1956 were shown.

	1947	**1948**	**1949**	**1950**	**1951**	**1952**	**1953**	**1954**	**1955**	**1956**
65	0.035	0.034	0.038	0.039	0.041	0.038	0.037	0.037	0.038	0.037
66	0.037	0.035	0.038	0.039	0.042	0.039	0.038	0.039	0.039	0.039
67	0.042	0.040	0.043	0.043	0.048	0.044	0.045	0.043	0.045	0.045
68	0.047	0.044	0.046	0.046	0.052	0.047	0.049	0.048	0.047	0.049
69	0.053	0.049	0.051	0.051	0.057	0.053	0.053	0.053	0.053	0.052
70	0.056	0.052	0.056	0.054	0.060	0.055	0.056	0.057	0.057	0.057
71	0.058	0.054	0.058	0.059	0.065	0.059	0.059	0.060	0.061	0.061
72	0.068	0.063	0.068	0.070	0.076	0.069	0.069	0.066	0.070	0.070
73	0.075	0.068	0.074	0.075	0.080	0.074	0.077	0.075	0.075	0.077
74	0.082	0.075	0.083	0.082	0.092	0.080	0.080	0.082	0.083	0.084
75	0.085	0.078	0.089	0.090	0.097	0.087	0.087	0.087	0.091	0.090
76	0.100	0.084	0.095	0.098	0.111	0.099	0.098	0.095	0.100	0.101
77	0.111	0.097	0.100	0.106	0.117	0.104	0.101	0.105	0.108	0.107
78	0.124	0.110	0.116	0.113	0.131	0.118	0.117	0.116	0.119	0.119
79	0.134	0.122	0.135	0.132	0.141	0.131	0.128	0.129	0.135	0.132
80	0.146	0.123	0.138	0.140	0.152	0.135	0.135	0.133	0.140	0.138
81	0.150	0.132	0.144	0.148	0.162	0.144	0.145	0.147	0.149	0.149
82	0.175	0.150	0.169	0.171	0.182	0.169	0.167	0.164	0.168	0.164
83	0.194	0.168	0.183	0.187	0.202	0.180	0.178	0.185	0.184	0.189
84	0.216	0.184	0.202	0.209	0.227	0.204	0.198	0.201	0.208	0.203
85	0.230	0.194	0.225	0.222	0.237	0.216	0.212	0.209	0.227	0.226

86	0.262	0.207	0.238	0.249	0.272	0.242	0.239	0.236	0.246	0.242
87	0.284	0.236	0.259	0.261	0.298	0.258	0.251	0.263	0.266	0.260
88	0.280	0.236	0.280	0.275	0.310	0.272	0.271	0.280	0.288	0.283
89	0.310	0.256	0.304	0.310	0.333	0.307	0.290	0.294	0.322	0.312

Appendix 1.1. Format of mortality data used for this research

Appendix 2. Difference between period and cohort mortality

Appendix 2 discusses how the progression of period and cohort mortality for a 67-year-old in 2010 differs. Period mortality tracks the horizontal development over time of successively “younger” 67-year-olds born in later years. Cohort mortality tracks the diagonal development of the same individual over time.

In terms of notation, $q_{x,t,c}$ denotes the mortality rate of an individual aged x in time t born in year c. Developments in period mortality are *italicised*. Developments in cohort mortality are given in **bold**.

	2010	**2011**	**2012**	**2013**	**2014**
67	$\mathrm{q}_{\mathrm{67,2010,1943}}$	$q_{67,2011,1944}$	$q_{67,2012,1945}$	$q_{67,2013,1946}$	$q_{67,2014,1947}$
68		$\mathbf{q}_{\mathbf{68,2011,1943}}$			
69			$\mathbf{q}_{\mathbf{69,2012,1943}}$		
70				$\mathbf{q}_{\mathbf{70,2013,1943}}$	
71					$\mathbf{q}_{\mathbf{71,2014,1943}}$

Appendix 2.1. Difference in development of period and cohort mortality for a 67-year-old in 2010

Appendix 3. Details on parameters of forecasted UK male mortality rates under the RH model

Appendix 3 shows details on the actual parameter values relevant to the forecasted mortality rates for UK males under the RH model, for age 85 in 2008 and age 86 in 2009, as illustrated in Figure 4.2.

To recapitulate, the mortality rate of age 85 in 2008 is described in (2.2) as

$m(2008,85) = \exp(A_{85}^{(1)} + A_{85}^{(2)}P_{2008}^{(2)} + A_{85}^{(3)}C_{1923}^{(3)})$

$= \exp(-1.58346 + 0.042799*(-13.227) + 0.024503*0.00003)$

$= \exp(-2.14957)$

$= 0.116535$

In comparison,

$m(2009,86) = \exp(A_{86}^{(1)} + A_{86}^{(2)}P_{2009}^{(2)} + A_{86}^{(3)}C_{1923}^{(3)})$

$= \exp(-1.46859 + 0.052343*(-13.5654) + 0.031538*0.00003$

$= \exp(-2.17864)$

$= 0.113195$

From above, the reason for m(2009,86) being less than m(2008,85) was due to the second term, with $A_{86}^{(2)}P_{2009}^{(2)}$ being more negative than $A_{85}^{(2)}P_{2008}^{(2)}$. The year 2009 was on its own expected to experience a lower rate of mortality than year 2008. As the associated age interaction term for 2009 ($A_{86}^{(2)}$) was also bigger than for 2008 ($A_{85}^{(2)}$), this resulted in the second term of m(2009,86) being noticeably more negative, explaining the decline in mortality from age 85 to 86.

Appendix 4. Explanation of asymmetry at 95% confidence level

Appendix 4 explains why at the 95% confidence level, the mortality rate at the upper bound departs more from the best estimate rate than at the lower bound. This is due to the modelling of log mortality rate and the effect of the exponential function in calculating the mortality rate.

Appendix 4.1 summarises from section 4.2.1.3, the mortality rates (to the first decimal point) for a life aged 86 in 2029, at the best estimate, 70% lower, 70% upper, 95% lower and 95% upper levels. The differences from the best estimate rate are also given.

	Best Estimate	**70% lower**	**70% upper**	**95% lower**	**95% upper**
Rate	7.4%	7.0%	7.8%	6.6%	8.3%
Difference	0.0%	-0.4%	0.4%	-0.8%	0.9%

Appendix 4.1. Summary of forecasted mortality rates for UK male life aged 86 in 2029

From equation (2.2), the RH model models the log mortality rate. Appendix 4.2 shows the log mortality rates at the levels discussed.

	Best Estimate	**70% lower**	**70% upper**	**95% lower**	**95% upper**
Rate	-2.60535	-2.66601	-2.54469	-2.72007	-2.49063
Difference	0.00000	-0.06066	0.06066	-0.11472	0.11472

Appendix 4.2. Forecasted log mortality rates for UK male life aged 86 in 2029

It can be seen that when log mortality rates are considered, there is no visible asymmetry from the best estimate rate at the 70% or 95% confidence levels. However, in moving from log mortality to mortality by raising to the exponent, the less negative log mortality rates become proportionally larger. This is shown by Appendix 4.3, when mortality rates with four decimal points are considered.

	Best Estimate	**70% lower**	**70% upper**	**95% lower**	**95% upper**

Rate	7.3877%	6.9529%	7.8498%	6.5870%	8.2858%
Difference	0.0000%	-0.4348%	0.4620%	-0.8007%	0.8980%

Appendix 4.3. Forecasted mortality rates for UK male life aged 86 in 2029 at three decimal point level

When rounded to the first decimal point, the rates above correspond to Appendix 4.1. It can be seen that when more decimal points are considered, there is asymmetry at both the 70% and 95% confidence levels. As the difference between the upper and lower bound log mortality rates is greater at the 95% than 70% level, the asymmetry of the forecasted mortality rates becomes visible at the 95% level. This is because only one decimal point was shown in the study.