Mortality models: comparison and application in old-age populations of selected economies

Name: Brian Jin-Wei Hu, FIA, FASSA, CFA

Affiliation: University of the Witwatersrand, Johannesburg, South Africa

E-mail: brianhu868@yahoo.com

Abstract:

This research examined which of the chosen five well-known extrapolative mortality models best captured the trends in old-age population mortality for different age groupings in four different economies. 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 economies. The research considered the goodness-of-fit of 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.

Keywords:

Mortality models, stochastic mortality projections, United Kingdom, Poland, Japan, Taiwan

1. Introduction

Improvement in 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 the British Actuarial Journal (2009). It has also been discussed in various international publications, such as in International Monetary Fund (2012) and World Economic Forum (2010, 2012). Understanding developments in old-age population mortality is important not only for governments with social security and welfare systems, but also for insurance companies and pension funds providing pension benefits. In an insurance context, a better understanding of the 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 Solvency II (European Insurance and Occupational Pensions Authority, 2011).

To date most studies on population mortality have looked at 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. The more commonly-used mortality models could be divided broadly into two regions in terms of their origination and usage: United Kingdom-Europe and Asia Pacific-North America. Studies originating from researchers in a particular region tend to use the prevalent model for that region.

This research considered five extrapolative mortality models that are well-known in the literature:

1. Lee-Carter (LC)
2. Renshaw-Haberman (RH)
3. Original Cairns-Blake-Dowd (CBD)
4. Booth-Maindonald-Smith (BMS) and
5. Hyndman-Ullah (HU)

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 the model’s early founding and history, it has been adopted internationally and appears to have become the starting model and the basis against which model improvements are compared. 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. As an 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.

As to date, most studies on mortality have looked at developed economies, this research aims to contribute to actuarial knowledge by also considering some emerging economies, as represented by Poland and Taiwan. The selected mortality models are applied in the old-age populations of these economies, and the results are compared against those in some developed economies, as represented by the United Kingdom and Japan. In the process, models from the two regions identified are applied across all four economies across Europe and Asia. This allows the usefulness and relevance of the models to be compared across economies, geographies and age groupings.

2. Data

The purpose of this research is to find models that explain trends in old-age male and female population mortality rates in selected developed and emerging economies in Europe and Asia, which are represented by:

a. The United Kingdom (European, developed), data available from 1947 to 2009

b. Poland (European, emerging), data available from 1958 to 2009
c. Japan (Asian, developed), data available from 1947 to 2009 and
d. Taiwan (Asian, emerging), data available from 1970 to 2009

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 economies. The underlying data were collected and compiled predominantly by the relevant national statistical agency of that economy, which lend credibility to the quality of the data.

Male and female mortality 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 UK data included data from England and Wales, Scotland and Northern Ireland. 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 economy level, in three age sub-groups of 65-69, 75-79 and 85-89, and in the 65-89 age group.

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 (2012) used age 65 as the reference point for projecting life expectancy in old age. 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.

As economies on the database 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, where improvements in mortality have started taking place. Therefore, the results and conclusions from this research are unlikely to be representative of all developed and emerging economies in general, because of their differing stages of economic development and socio-demographic backgrounds.

Nevertheless, the framework used in this research can be extended to other economies when longer histories of reliable population mortality data become available.

3. Mortality models considered

3.1. Expectation, explanation and extrapolative models

According to Booth and Tickle (2008), there are three main classes of models in the field of mortality modelling: expectation, explanation and extrapolative.

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. 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.

For these reasons, this research only considered the "extrapolative" class of models.

3.2. Lee-Carter (LC)

Of the five models, the Lee-Carter (LC) 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 LC model 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:

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

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:

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

3.3. Renshaw-Haberman (RH)

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 as:

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

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.

3.4. Original Cairns-Blake-Dowd (CBD)

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 (using the logistic function). 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, and is presented as:

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

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 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≥0, u<1, μ(t+s,x+u) = μ(t,x), where
μ(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) = μ(t,x)
b) q(t,x) = 1 – exp[-μ(t,x)]
= 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 this research, the 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.

3.5. Booth-Maindonald-Smith (BMS)

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 is presented as:

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

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 economies, 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.

3.6. Hyndman-Ullah (HU)

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. The model is presented as:

$$\begin{aligned}\text{Log } m(t,x) &= f(t,x) + \sigma(t,x)E(t,x) \\ &= \mu(x) + \sum_{i=1}^{n} B(t,i)D(x,i) + E_2(t,x) + \sigma(t,x)E(t,x) \\ &= \mu(x) + \sum_{i=1}^{n} B(t,i)D(x,i) + \sigma(t,x)E(t,x) + E_2(t,x)\end{aligned}$$

The HU model started off as modelling log mortality rates using a function f(t,x) that is observed with error. 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 the final line representing the model.

From Booth *et al* (2006, p294), the terms 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 economies 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.

4. Methodology

4.1. Computing language

Application of the five mortality models discussed 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 this author 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. The "Demography" package written by Hyndman (2011a) was used to fit the BMS and HU models. Once the parameters in the five models had been fitted, the "Forecast" package (Hyndman, 2011b) was used to make projections for all five models for consistency.

4.2. Assessing goodness-of-fit

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

Scenario A: Analysis of model-fitted mortality rates over the full period of data available for that economy, 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 from 1990 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 economy 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 economies 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 the Appendix. 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

4.2.1 Bayes' Information Criterion (BIC)

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

BIC = ll – 0.5 x p x log N

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. The BIC measure is appropriate across models that use the same number of years in the fitting process. As discussed in 3.5, the BMS model fits selectively only over years that meet the linearity of mortality decline assumption. Therefore, all else equal, 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.

4.2.2 Residuals

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

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

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

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

$$t = \frac{\overline{d}}{s_d / \sqrt{mn}}$$

where

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

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%).

4.2.3 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

$$\overline{d_A} = \frac{1}{mn}\sum_{x}\sum_{t}|\hat{m}(t,x) - m(t,x)|$$

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

$$t = \frac{\overline{d_A}}{s_{dA} / \sqrt{mn}}$$

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

4.3. 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 4.2). 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 4.2), 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 4.2). Within this period more reliance was placed on residuals and absolute residuals, with the BIC used only as an indicator, for reasons discussed in 4.2.1.

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, Scenario A was considered before B and C.

4.4. Application of selected model

Once a model has been selected for a gender in an economy, 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, after fitting the model over the full period of data available for that economy. 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 the Appendix), 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 economies 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 economies. 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 economy to economy.

For the purposes of interpretation, using the interesting rate parity argument, choosing this interest rate was similar to analysing the price of annuities denominated in British Pounds, on the different populations in different economies. 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 economy. 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.

Despite this, given that this research also aims to compare mortality projections across economies, the advantages with using a common interest rate outweighed the disadvantages. To bring more perspective into the analysis, results using other annual interest rates such as 0.5% and 2.5% were also presented.

5. Results

5.1. United Kingdom

The analysis began by considering the male population first. Following the steps described in Sections 4.2 and 4.3, the RH model emerged as the most suitable model. For the sake of keeping this paper concise, only the key results that emerged from the selection process were shown.

Table 1 shows the BIC statistics calculated for the five models, under the analysis for Scenario A in Section 4.2. 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. Periods which did not meet this assumption were not considered for fitting purposes by this model.

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

Table 1. BIC statistics, UK males

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 was shown in bold.

Moving onto Scenario B of Section 4.2, the five models were calibrated using the data from 1947 to 1989, and projections of mortality were made over 1990 to 2009. Table 2 shows the size of the residuals associated with the projections from the models, across the different age groupings. The residuals are shown in scientific notation (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 (26%)**	47	**16**	**-109**
CBD	146	71	139	239
BMS	100	**32**	93	169
HU	154	72	151	242

Table 2. 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.

Table 3 looks at the absolute residuals (x 10^{-4}), with the smallest absolute residual in each age grouping highlighted in bold.

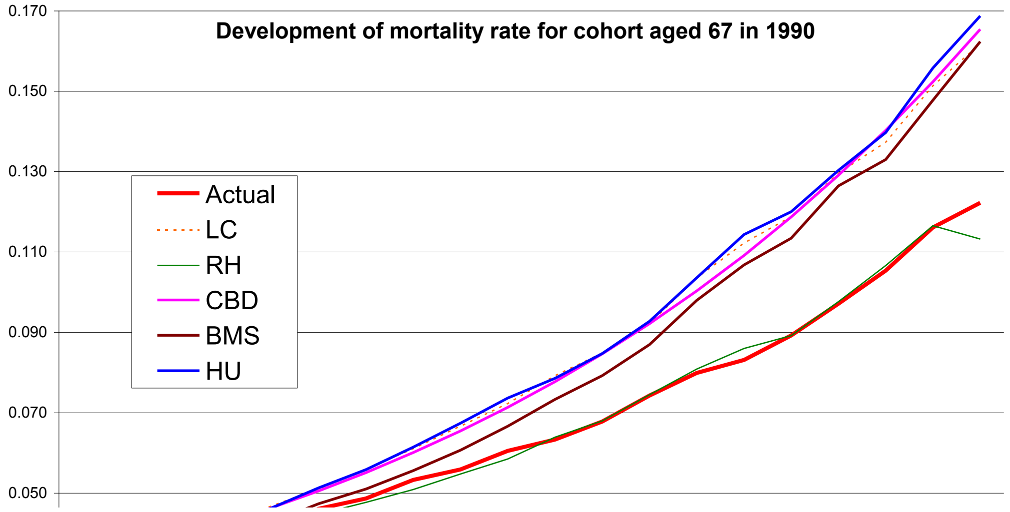
UK Males	**Forecast:1990-2009**			
Model	65-89	65-69	75-79	85-89
LC	134	67	146	167
RH	**54**	49	**32**	**116**
CBD	146	71	139	239
BMS	101	**33**	93	173
HU	154	72	151	243

Table 3. Absolute residuals and p-values (in brackets), 1990 – 2009, UK males

Similar to the results for residuals in Table 2, 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 all less than 5%.

Based on the analysis in Scenario C of Section 4.2, Figure 1 shows how the projections for age 67 compared against actual historical mortality from 1990 to 2009.

Figure 1. 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, but underestimated mortality in 2009. As Figure 1 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 surprisingly lighter mortality was due to the interaction of fitted and forecasted age, period and cohort parameters. With reference to section 3.3, 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 was largely why the lighter mortality occurred.

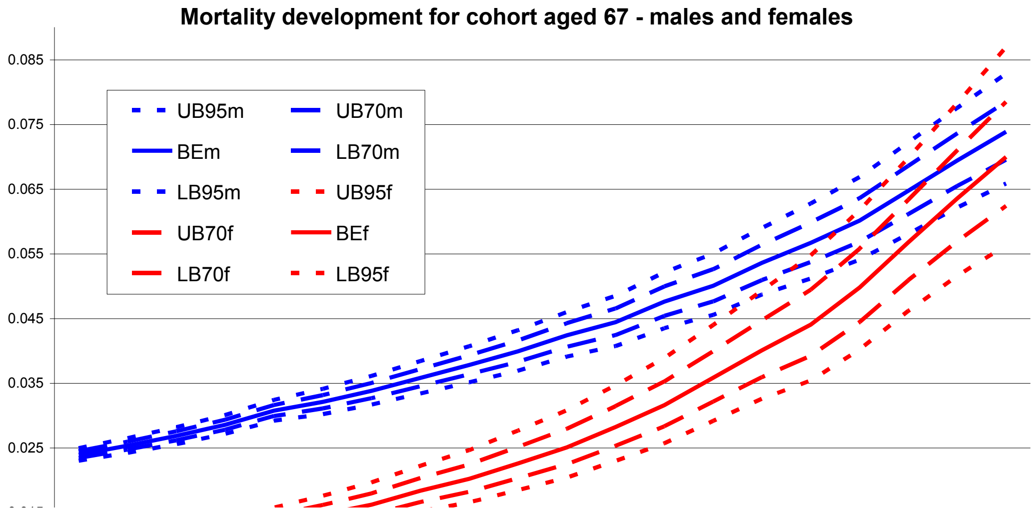
Despite the unusual projection for 2009, relative to the other four models which all overestimated mortality rates, the RH model tracked actual experience much more closely.

The eventual decision to choose the RH model for UK males was also consistent with the finding in Cairns *et al* (2007) of the existence of a cohort effect in male English and Welsh data.

A similar process was followed for UK females, and the BMS model emerged as the most appropriate.

With the selection of the RH model for males and BMS for females, projections of future mortality rates over the 20 years from 2010 were made for a 67-year-old. The best estimate, 70% and 95% confidence interval cases were considered. The projections are shown in Figure 2, which compared the best estimate development in cohort mortality for males (in blue) and females (in red), along with their 70% and 95% confidence intervals.

Figure 2. Comparison of developments in cohort mortality for males and females, for a 67-year old in 2010 in the UK



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 would begin to narrow, as males begin 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 2 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.

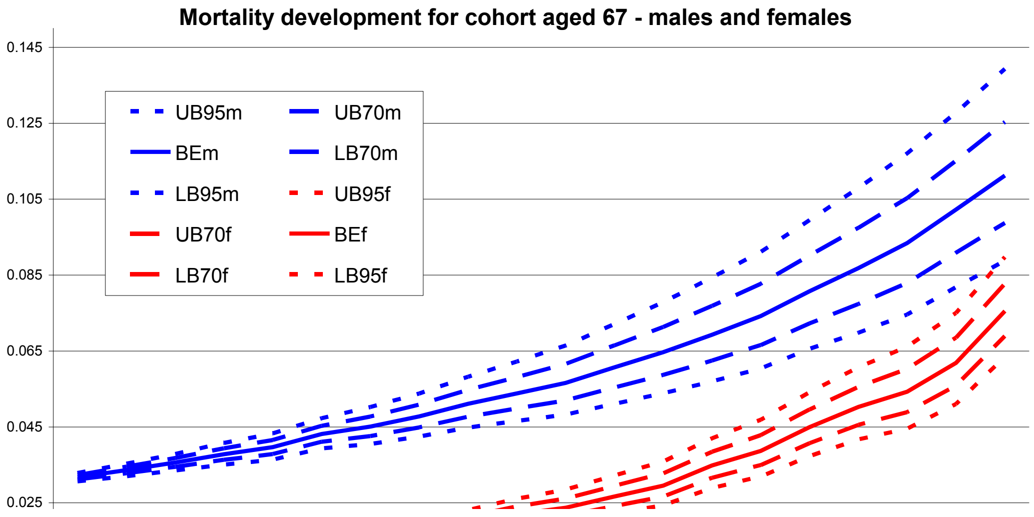
Interestingly, the convergence of UK male and female mortality over time was also mentioned in Mayhew (2012). He found that the gap between male and female life expectancy at age 30 in England and Wales was closing, and men could catch up with females by 2030. This was thought to be attributable to declining 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.

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. Using an interest rate of 1.5%, the percentage difference in the theoretical annuity price mentioned in Section 4.4 between the best estimate and 95% confidence interval mortality levels was at around 2% and 3% respectively for UK males and females. This suggests that the financial impact of mortality uncertainty was not too drastic (at least compared to other economies, as will be discussed in Section 5.5), even with interest rates at historic lows.

5.2. Poland

Following the steps set out in Sections 4.2 and 4.3, the BMS model was chosen for Polish males, and RH model for females. As for the UK, projections of future mortality rates over the 20 years from 2010 were made for a 67-year-old. The best estimate, 70% and 95% confidence interval cases were considered. The projections are shown in Figure 3.

Figure 3. Comparison of developments in cohort mortality for males and females, for a 67-year old in 2010 in Poland



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

Unlike the results for UK pensioners shown in Figure 2, over the following 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.

Figure 3 also showed graphically from the width of the bands that uncertainty around male mortality was wider than for females.

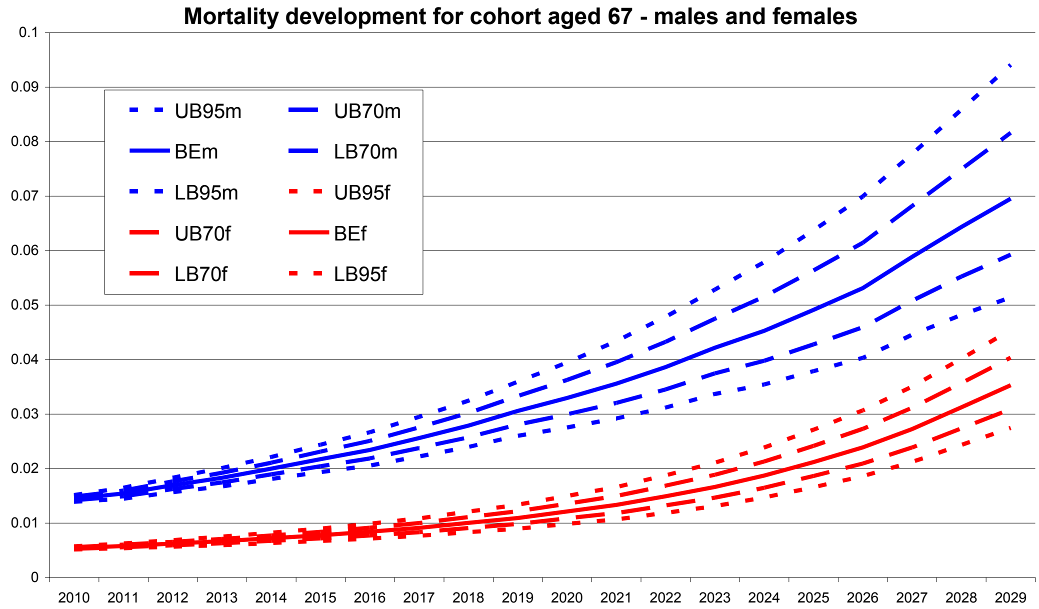
Comparing Figure 2 with Figure 3, relative to UK populations, there is less smoothness in the projected mortality rates for Polish populations. This is possibly due to the following reasons. Firstly, looking at historical mortality rates, since 1950 the UK populations have experienced steady declines 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 introduced more variability into 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.

5.3. Japan

Unlike the UK and Poland, in the case of Japan, one model was found as the best for both males and females. It was the BMS model.

Figure 4 shows the projections of future mortality rates over the 20 years from 2010 for a 67-year-old.

Figure 4. Comparison of developments in cohort mortality for males and females, for a 67-year old in 2010 in Japan



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

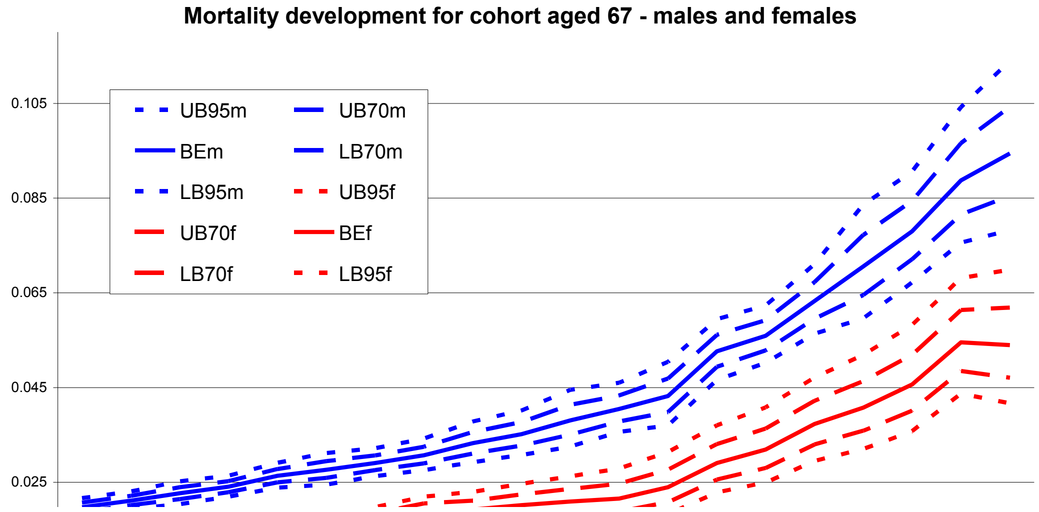
Unlike the results for UK pensioners in Figure 2, and somewhat similar to results for Poland in Figure 3, over the next 20 years, the gap between the best estimate mortality for Japanese males and females is not expected to narrow. The 95% upper bound for female mortality does not even cross the 95% lower bound for male mortality over this projection period.

Figure 4 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 in Japan.

5.4. Taiwan

In Taiwan, the HU model was chosen for males, and BMS for females. Figure 5 showed the projections of future mortality rates over the 20 years from 2010 for a 67-year-old.

Figure 5. Comparison of developments in cohort mortality for males and females, for a 67-year old in 2010 in Taiwan



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

Unlike the results for UK pensioners in Figure 2, but similar to results for Poland (Figure 3) and Japan (Figure 4), 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 economies, due to the less than regular shape of historical mortality rates in Taiwan.

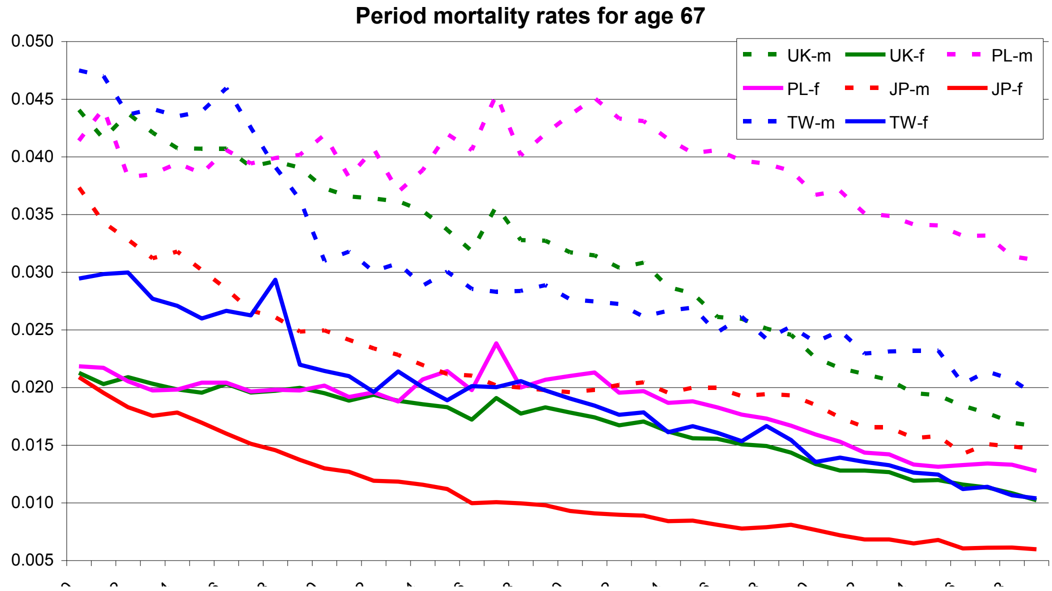
As for Japan, Figure 5 also showed graphically from the width of the bands that uncertainty around male mortality is wider than for females. This was because as in Japan, the variability in historical mortality rates was greater for Taiwanese males than females.

Comparing Figure 4 with Figure 5, relative to the 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 introduced more variability into 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.

5.5. Comparison of historical mortality rates across economies

Based on data available from the HMD, figure 6 summarised the mortality rates for males and females aged 67 in 1970, over the period 1970 to 2009, across both genders in the four economies considered. The year 1970 was the starting point because that was when sufficient mortality data for all four economies became available in this study.

Figure 6. Period mortality rates for age 67, across all eight populations



For all four economies, 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 group. 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, due to a significant drop in the late 1970s (when economic growth began to take off) and continual declines thereafter, by 2009 it was similar to the level for UK males. In contrast, Polish males actually experienced 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.

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.

5.6. Comparison of projected mortality rates across economies

To summarise the findings in Sections 5.1 to 5.4, the model chosen for each gender in each economy is shown in Table 4:

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

Table 4. Summary of models chosen for the eight populations

From Table 4, 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). 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 original 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 original 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. On the other hand the basic version of the CBD model was probably not complex enough to have captured enough of the relevant 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 suggested 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 best 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). In these latter two populations, 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 occurred 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 economies. 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.

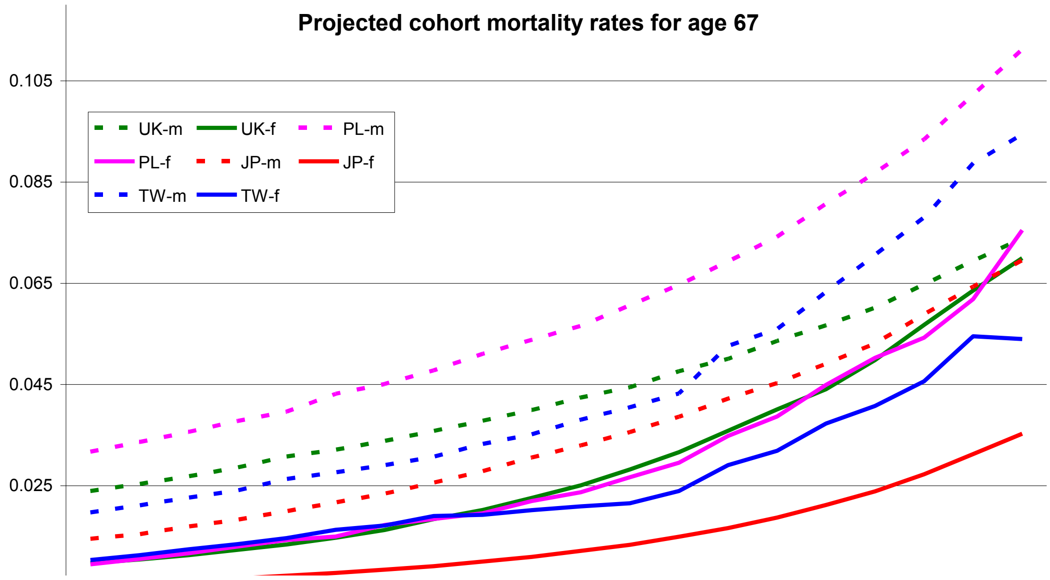
In contrast, in the emerging economies (Poland and Taiwan) where mortality improvements were more variable, the BMS model was best for only two of the four populations, and was not particularly preferable for either gender.

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 7 summarised the projected best estimate cohort mortality rates for males and females aged 67 in 2010, across the eight populations in the four economies in this study.

Figure 7. 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 economies. The UK was the only economy where male rates appeared to be converging towards those of females, as discussed in section 5.1.

As both males and females approached the natural limit to human life expectancy, the eventual increases in female mortality rates needed to be faster to compensate for their lower starting levels, 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 were 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 7, the developed economies, which had longer and less variable histories of mortality, 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.

5.7. Comparison of theoretical annuity prices and stability of mortality forecasts

Table 5 is a summary of the prices of the theoretical 20-year level annuities described in 4.4, for the eight populations under different mortality scenarios. As all were calculated at 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 5. Comparison of theoretical annuity prices across eight populations, at an interest rate of 1.5% for a 67-year-old

From Table 5, 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, UK and Polish males.

It was interesting to note that for males and females, the Asian populations were generally expected to live for longer than the European ones in this study.

From Table 5, 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 6 is a summary of the percentage annuity 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 6. 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 7 is similar to Table 6 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 7. Comparison of variation to annuity prices across eight populations, at an interest rate of 0.5%

Compared to Table 6, 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 6.

Table 8 is similar to Tables 6 and 7 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. Comparison of variation to annuity prices across eight populations, at an interest rate of 2.5%

Compared to Table 6, 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 6.

Table 9 considered the percentage price change in the annuity under different interest rates. The price of the annuities under the best estimate, 95% lower bound and 95% upper bound mortality scenarios were considered. The prices calculated under the 0.5% and 2.5% interest rates were compared against that using the 1.5% rate.

There is an inverse relationship between the rate of interest and the annuity price. To illustrate this, in the example given by Table 9, 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 9. 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 9, at a starting 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% - (-8.9%)), whilst that for Polish males would be smaller at 16.2% (=8.9% - (-7.3%)).

Table 10 builds on Table 9 by considering the range of fluctuations in the annuity price for all populations, which is calculated 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 10. Fluctuation in annuity price with rates ranging from 0.5% to 2.5%

From Table 10, the order of the sizes of fluctuations was very similar to the order of the sizes of annuity prices in Table 5. This supported 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.

6. Conclusion

Although only four economies were chosen for this study, many interesting insights were gained, particularly for the 65-69 age sub-group, which 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, tends to be the first-time buyers of annuity products, and by default have the most number of years to live.

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

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 economy considered where male mortality rates are expected to converge towards female rates over the forecast period.

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.

7. Room for further research

This research was intended as a starting point for applying different stochastic mortality models to populations across different continents in different stages of economic development. For now, only four economies were selected, but it would be straightforward to extend the type of analysis conducted in this research to other economies, 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, although the results are not expected to change noticeably, given that the choice of 20 years was applied consistently across economies, genders and models.

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

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-economy 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-economy 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

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. (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., 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

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.

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

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

International Monetary Fund (2012). 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

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

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

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

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 (2012). Mortality, 2010-based NPP reference volume. Downloaded from www.ons.gov.uk

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

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

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

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.

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

Appendix. Difference between period and cohort mortality

For illustration, the Appendix looks at 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	$q_{67,2010,1943}$	*$q_{67,2011,1944}$*	*$q_{67,2012,1945}$*	*$q_{67,2013,1946}$*	*$q_{67,2014,1947}$*
68		**$q_{68,2011,1943}$**			
69			**$q_{69,2012,1943}$**		
70				**$q_{70,2013,1943}$**	
71					**$q_{71,2014,1943}$**