Report of Honours Research Project conducted under support of an APRA Brian Gray Scholarship

Principal Components Modelling of Mortality

Michael Jesse O'Neill

School of Finance and Applied Statistics, Australian National University

17/12/2004

Author: Michael Jesse O'Neill

Reference No. MJO

Background

Descriptive mortality models involve functions which are used to describe patterns of mortality during the life of a cohort. Each cohort comprises a generational group. These models are used to predict future mortality of different cohorts, in to order inform the financial policies of insurance companies and the social policies of governments. Prevailing theories of mortality focus on mortality related to the ageing process or the underlying age-structure of mortality ('age effects') and mortality relating to the environment or time trends in mortality ('period effects'). However, mortality can also be affected by the life histories of different generational groups which make up the population. Most models fail to take account of these so-called 'cohort-effects'. The main reason for this is the difficulty in objectively identifying and modelling these effects using existing methods. Cohort effects are the least visible effects, being masked by more dominant age and period effects. There have been attempts to model the three effects together, but these have proven to be less informative and less statistically sound. The method developed in this research project uses an alternative approach for modelling age, period and cohort effects. It is able to prioritise the modelling of dominant age and period effects, and then address the more subtle cohort effects.

Aims and significance of the project

The method of modelling used in this research project is called Principal Components Analysis (PCA). It involves a mathematical procedure which transforms the mortality data into two uncorrelated principal component functions: the mortality function across all ages (the age-dependent factors of mortality) and the coefficient function across all years (the time-dependent factors of mortality). The first principal component is the most significant, describing about 90% of the mortality fluctuations which have occurred over the period of the study. The succeeding principal components are of decreasing significance. PCA mortality models are becoming more widely used in mortality modelling. This project demonstrates the advantages of PCA methods in describing the mortality profile of a cohort.

Cohort effects are clearly secondary to age and period effects. Cohort effects cannot be completely isolated from age and period effects. The three effects are confounded. However, cohort effects should not be discounted. If models can be adapted to fit these more subtle effects, then very accurate mortality forecasts may be possible for specific cohorts. The development of such models is supported by demographic literature.

The PCA model developed in this research project is called the 'age-period-cohort PCA' or 'APC-PCA' model. It can be distinguished from all existing mortality models by its use of PCA to model age, period and cohort effects together, with corresponding principal components mortality functions and coefficient functions. It involves age-, time- as well as cohort-dependent factors of mortality.

Research approach

Male mortality data from the Netherlands mortality database from 1850 to 1991 (Tabeau *et al*, 1994) was used for the development and testing of the new APC-PCA model.

Firstly, various PCA models were applied and critically evaluated. They were compared on the basis of how much variation they explain. Conclusions were reached about the preferred method. Compared to standard PCA methodology, the Lee and Carter (1992) model and the Booth-Maindonald-Smith model (Booth *et al*, 2002) are driven by ease of forecasting rather than accuracy. The Wilmoth (1990) model is driven by the desire to explain cohort effects. The De Jong and Tickle (2003) model represents an attempt to simplify PCA methodology. All of these models depart from standard PCA methodology in ways that are not favourable.

The consequences of using the different PCA approaches were analysed. Conclusions were drawn about the most appropriate way to apply PCA in order to obtain the best fit to the data and to most clearly highlight the profile of mortality over time. Standard methodology was preferred, and adopted to develop the APC-PCA model.

New methods for removing observational error in the data were also applied and evaluated (Ramsay and Silverman, 1997; Hyndman, 2003; de Jong and Tickle, 2003). However, the raw data was preferred for the purposes of the APC-PCA model, due to the risk of prematurely discarding the more subtle cohort effects as observational error.

The Netherlands male mortality data was very suitable for the development of the new APC-PCA model, given that it has been accurately recorded over a period of 142 years for the entire population. The data was particularly useful for testing the robustness of the model, since significant cohort effects occurred in the Netherlands during the period of the mortality study.

When PCA was applied to the Netherlands data, the age, period and cohort mortality effects became very clear. The effects were easily interpretable when the principal component functions were plotted. In particular, contour plots were used to aid in the interpretation of the model fit. Horizontal, vertical and diagonal patterns proved to be related to age, period and cohort effects, and interaction effects, respectively. The effects identified by PCA were strongly attributable to historical events, in particular World Wars I and II.

The greatest difficulty encountered in development of the APC-PCA method was fitting the model. The cohort-dependent mortality factors could not simply be fitted alongside the age-dependent and time-dependent factors, since they had to be fitted using PCA in the age and cohort dimensions. Difficulty in application of PCA in the age and period dimensions as well as the age and cohort dimensions was overcome by using an expectation maximization algorithm (EM algorithm). An EM algorithm is an iterative procedure used to determine the most appropriate parameters for the model, given the observed data. An EM algorithm was successfully developed to allow application of PCA in the age and cohort dimensions. As a result, the APC-PCA model could be successfully fitted to the data, first by performing PCA to determine the age-dependent and time-dependent factors, and then by performing PCA to determine the cohort-dependent factors with the assistance of the EM algorithm.

The advantages and limitations of the new method were investigated extensively in this project. The model was found to fit the data well in most cases. Residual effects were very small with no clear structure, suggesting that the model picked up all important features of the Netherlands' mortality experience.

Key results and implications

The need for development of powerful descriptive mortality models is emphasised in this project. A new method of mortality modelling called age-period-cohort PCA or APC-PCA is advanced. Age, period and cohort effects in mortality have been modelled using PCA and the EM algorithm. The method is very good at describing the mortality data, due to its flexibility and its capacity to model all three effects together. The success of the APC-PCA approach is evidenced by its description of the most significant features in the mortality data. Significant cohort effects have been modelled, resulting in a better fit to the Netherlands historical mortality experience.

The new APC-PCA method is endorsed in this project. In some cases it proved to be more powerful than existing methods of mortality modelling. Importantly, the APC-PCA method modelled cohort effects in addition to dominant age and period effects. Even with confounded age, period and cohort effects, the new method was able to reveal the relative significance of cohort effects.

The method developed in this project was used to predict future mortality based on past mortality experience (time series prediction). There is clear scope for extending the method further, and developing different ways of forecasting, which build in a margin of forecasting error.

Implications for prudential supervision

Cohort effects have been identified in international mortality data, associated with events such as World War I. However, smaller cohort effects may exist generally in aggregate populations, being masked by more dominant age and period effects. Cohort analysis could be extended to analyse health issues, mortality data by cause of death and specific risks such as veteran's pension risks, in the Australian population.

APRA supervises a wide variety of superannuation funds and insurance providers. Legislative supervision is aided by powerful modelling of risk through data collection and analysis. Actuarial techniques need to be developed further to arrive at models which balance parsimony and descriptive power. The gradual development of more robust and accurate models for mortality forecasting will help to further APRA's goals. With a better understanding of the process of ageing experienced by various cohorts in the Australian population, policy makers will be better positioned to maintain social support and to ensure the safety of superannuation funding.

Further analysis of cohort effects using methods like APC-PCA should inform APRA policy in drafting future standards for actuarial analysis and management of risk. It may also improve public awareness of the true cost of reasonable pension provision. Ultimately, improved actuarial modelling in the areas of superannuation and life insurance could result in public product awareness and improved performance of supervised institutions.

Conclusions

This project applies PCA methodology and develops a new approach to mortality modelling. PCA methods are used to understand historical mortality experience in greater detail. The power and flexibility of existing PCA methodology is further enhanced by the inclusion of cohort mortality factors to model cohort effects. The cohort factors are modelled after the dominant age and period factors. This process is facilitated by the use of an EM algorithm. The new APC-PCA method outlined in this research project can be used to effectively describe mortality data in terms of age, period, as well as cohort effects.

Publications arising from honours project

O'Neill M. *Principal Components Modelling of Mortality*. Honours Thesis, Faculty of Economics and Commerce, Australian National University, 2004.

O'Neill M. Principal Components Modelling of Mortality – a new approach to Age-Period-Cohort effects. Submitted for publication in *The North American Actuarial Journal*.

O'Neill M. Principal Components Modelling of Australian Mortality. In preparation.

O'Neill M. Principal Components Modelling of Mortality – the importance of pre-smoothing. In preparation.

O'Neill M. Principal Components Modelling of Mortality – forecasting Age-Period-Cohort effects. In preparation.

References

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

de Jong P, Tickle L. Forecasting mortality rates and related actuarial values. Research Paper, Department of Actuarial Studies, Macquarie University: Sydney, 2003; to appear; available at http://www.actuary.mq.edu.au/research_papers/index.shtml.

Hyndman R. Be smooth and robust: a new approach to age-specific demographic forecasting. Seminar at the Australian National University: Canberra, 2003; available at http://www-personal.buseco.monash.edu.au/~hyndman/.

Lee R, Carter L. Modelling and forecasting US mortality (with discussion). *Journal of the American Statistical Association* 1992; **87(419)**:659-675.

Ramsay J, Silverman B. *Functional Data Analysis.* Springer-Verlag New York Inc: New York, 1997.

Tabeau E, Willekens F, v and van Poppel F. 1994. “Mortality in the Netherlands: the Data Base.” Report no. 36, Netherlands Interdisciplinary Demographic Institute, The Hague, The Netherlands.

Wilmoth J. Variation in vital rates by age, period, and cohort. In *Sociological Methodology,* Clogg C (ed.). Basil Blackwell: Oxford, 1990; **20:**295-335.