Institution: City University London

Unit of Assessment: 19 Business and Management Studies

Title of case study: Improved estimation of mortality and life expectancy for each constituent country of the UK and beyond

1. Summary of the impact

Graduated period life tables for men and women, based on the mortality experience of the population of England and Wales, have been published by the Office for National Statistics (ONS) using data from the 2001 Census. These tables are the sixteenth in a series known as the English Life Tables which are associated with decennial population censuses, beginning with the Census of 1841. Errors in crude census data owing to the small numbers of deaths involved, particularly in childhood and at very advanced ages, can be reduced by a statistical process of smoothing. A smoothing methodology developed at Cass Business School, City University London has been used in the latest ONS Decennial Life Tables. The tables show the increasing longevity of the population of England and Wales over a long period. The impact of this research is broad as life tables are used extensively in pensions planning, demography, insurance, economics and medicine. Life tables using this statistical smoothing methodology have also been prepared for Scotland, Northern Ireland, the Republic of Ireland and Canada.

2. Underpinning research

Life expectancy is the number of years that a person at a given age can expect to live, on average, in a given population. Published once every ten years, in association with the decennial population censuses, the ONS decennial life tables show age-specific death rates and life expectancy figures for England and Wales, beginning with the Census of 1841. The latest Decennial Life Tables, for 2000 to 2002, are based on crude death rates data centred on the census year 2001. A three-year period is needed to smooth most of the effect of the mortality experience of the census year being atypical of the general level of mortality at the beginning of the decade. In producing life tables, mortality rates during the first year of life and at the oldest ages may vary erratically owing to the small numbers of deaths involved and given errors present because no census is perfectly accurate or complete. Errors arising because of the small numbers of deaths and, to some extent, other types of error can be reduced by the process of smoothing these crude death rates.

For the latest ONS Decennial Life Tables (henceforth English Life Tables No 16, or ELT 16), the intention of this smoothing (or ‘graduation’) is to replace the crude rates by a series of graduated rates which, while forming a smooth progression over the whole age range, will still preserve the general shape of the mortality curve. In the past, various means of carrying out this smoothing were applied in constructing the ELT. For the current graduation, the methodology used was developed at Cass Business School by Vladimir Kaishev (at City since 2002, now Professor), Steven Haberman (at City since 1974, now Professor) and Dr Dimitrina Dimitrova (at City since 2004, now Lecturer). This follows a variable-knot spline regression approach (see next paragraph) and uses a weighted least squares version of the spline regression method proposed by Kaishev *et al.* (2006a, 2006b).

Fitting a smooth curve to a sample of noisy observations of a response variable which depends on one or more explanatory variables is one of the most intensively researched problems in statistics and is known as regression. The related literature is vast. One popular approach to solving the problem is to construct a least squares spline fit to the data. A spline curve of a fixed degree (usually two or three) is a piecewise polynomial function which consists of pieces of polynomials of the fixed degree, smoothly joined at some points called knots. Thus, a spline function is defined by its degree, by the number and location of its knots and by the coefficients of the spline basis functions. The graduation method proposed by the Cass academics determines the degree of the spline fit and also the number and position of its knots (respectively pieces of polynomials), according to an optimality criterion. Applying this method produces quadratic spline fits of the crude death rates, which are not overly parameterised and can be evaluated for any arbitrary age using a calculator.

A direct approach to spline regression is to assume that the degree of the spline and the number of knots are fixed (but unknown) and to find the knot locations which minimise the least squares distance between the noisy data points and the spline fit. However, this approach leads to a multi-extrema non-linear optimisation problem which is hard to solve especially for the large number of unknown parameters (knots and coefficients) often required when fitting wiggly dependences underlying the data. To overcome these difficulties, alternative approaches have been proposed, including step-wise knot inclusion/deletion strategies, Bayesian adaptive splines methodologies, adaptive genetic splines algorithms and penalised smoothing spline fitting methods. However, all have common limitations: they are computationally costly, may lead to over-fitting and do not estimate the degree of the spline.

The Cass academics developed a method called Geometrically Designed Splines (GeDS) which overcomes the difficulties and limitations of other approaches and which could automatically estimate simultaneously the degree of the spline (i.e., linear, quadratic cubic, etc.), the number of knots and their locations and the spline regression coefficients for large data sets; the wide range of signal-to-noise ratios and the various underlying functional dependencies between the response variable and the explanatory variables.

To achieve this, instead of following the conventional approaches to least squares spline regression estimation, the Cass academics took a novel geometric approach consisting of two stages (Kaishev *et al.* 2006a, 2006b). First, they constructed a linear spline fit, driven by the data, which captured the geometrical shape of the underlying functional dependence. Second, in order to smooth this linear fit but follow its shape and therefore the shape of the data, they appropriately 'attached' to it a spline curve of higher degree and optimally estimated knot locations, so that it minimised the least square error. This method has been thoroughly tested and verified in several simulated and real data applications, including data from materials science and demography (Kaishev *et al.* (2007) and Dimitrova *et al.* (2008, 2013)). The results show that it is extremely numerically efficient, requires the initial setting of only two input parameters and therefore is semi-automatic and produces simultaneously linear quadratic and higher order spline fits, the best of which is selected as the final model. This is the only method which is capable of estimating the degree of the spline fit, the knot locations and spline coefficients.

In applying GeDS to the context of smoothing the English Life Tables No 16 and other national mortality datasets, the Cass academics produced quadratic spline fits of the crude mortality data, which have also been extrapolated up to the limiting age of 120. As mentioned above, the estimated mortality curves are not overly parameterised and can be evaluated easily for any arbitrary age, producing a smooth life table and related life expectancy statistics.

3. References to the research

Dimitrova D.S., Kaishev V.K., & Penev S.I. (2008). GeD spline estimation of multivariate Archimedean copulas. *Computational Statistics and Data Analysis*, 52(7), 3570-3582.

Dimitrova D.S., Haberman S., & Kaishev V.K. (2013). Dependent Competing Risks: Cause Elimination and its Impact on Survival. *Insurance: Mathematics and Economics*, Forthcoming.

Kaishev V.K., Dimitrova D.S., Haberman S., & Verrall R. (2006). *Geometrically designed, variable know regression splines: asymptotics and inference*. (Statistical Research Paper No. 28). London: Faculty of Actuarial Science & Insurance, City University London.

Kaishev V.K., Dimitrova D.S., Haberman S., & Verrall R. (2006). Geometrically designed, variable knot regression splines: variation diminish optimality of knots. (Statistical Research Paper No. 29). London: Faculty of Actuarial Science & Insurance, City University London.

Kaishev V.K., Dimitrova D.S., & Haberman S. (2007). Modelling the joint distribution of competing risks survival times using copula functions. *Insurance: Mathematics and Economics*, 41(3), 339-361.

The research was published in journals that apply a stringent peer-review process prior to accepting articles for publication and was supported by grants from The Leverhulme Trust and the Institute and Faculty of Actuaries.

4. Details of the impact

A distinctive feature of the actuarial research at Cass lies in the strength of the relationships with the actuarial profession and with the Government Actuaries Department (GAD), the non-ministerial department which provides actuarial analyses to governments and organisations in the UK public sector. In 2008, GAD brought together clients, industry figures and stakeholders to discuss future longevity improvements. This was one of their 'hot topics' [1]. The event was attended by statisticians from the Office for National Statistics (ONS), Cass Business School and many other academics and professional actuaries. It was at this event that the ONS statisticians first observed the statistical smoothing technique, Geometrically Designed Splines (GeDS), developed by the Cass team. This information was communicated to the ONS staff working on the English Life Tables 16 which led to a successful collaboration between the ONS and Cass, resulting in the GeDS being used as the smoothing methodology [2], [3]. As Adrian Gallop of ONS and GAD explains: "*The GeDS smoothing methodology provides an efficient way of graduating mortality rates, including extrapolation to the oldest ages and has a particular advantage in that mortality rates at non-integer ages can be readily determined*" [4].

The ELT 16 provides a valuable time series which can be used to monitor trends in mortality in England and Wales over a long period of time. These period life tables show the increasing longevity of the population of England and Wales over a long period and can be compared with the experience of other countries. One way of illustrating the reductions in death rates is to show the increase in expectation of life at birth, shown in Life Expectancy Tables. ONS Life Expectancy Tables for England and Wales are constructed at various ages for mortality based on the years 1910 to 1912 and at twenty year intervals thereafter until 1970 to 1972, together with those for 1980 to 1982, 1990 to 1992 and 2000 to 2002. The ELTs are used in calculating historical rates of mortality improvement over long periods from 1911 [3].

Further, when mortality estimates are combined with fertility data, migration data and trend-based assumptions, life tables can be used to make population estimates (the number of people who were usually resident in an area at the mid-year point) and population projections (a picture of the population as it may develop in future years) [5]. The ONS population projections use mortality rates estimates in each calendar year in the period 1961 to 2007. These have been graduated using a method similar to that used for graduating the English Life Tables No 16.

ELT 16 and related life expectancy tables are available to the public and businesses for use free of charge through the ONS website. Web metrics provided by the ONS contain information on the number of web visits and downloads of the English Life Tables No 16 and show between 280 and 490 downloads per month in the period November 2012 to June 2013 [6]. Following the publication of the English Life Tables based on data for 2000 to 2002, similar life tables employing the same graduation methodology were prepared by the ONS on behalf of the General Register Office for Scotland [7] and Northern Ireland Statistics and Research Agency [8].

The impact of this research has been extended internationally as life data have been subject to the smoothing process to meet the needs of the demography division of Statistics Canada, the Canadian government office concerned with national statistics and the Central Statistics Office, Republic of Ireland.

In Ireland, the Cass researchers advised Kevin McCormack, the Senior Statistician in charge of the Social Analysis Division in the Central Statistics Office. Mr McCormack is presently engaged in a project to develop a more statistically accurate cubic-spline graduation method and to apply it to the latest Irish crude mortality rates from 2005 to 2007. Period Life Tables have been produced by the Irish Central Statistics office on fifteen occasions from 1926 to 2005 to 2007. On each occasion the Kings 1911 formula for Osculatory Interpolation was used to graduate the crude mortality rates. The shift now to using cubic-spline graduation of Irish crude mortality rates therefore represents a major shift in statistical policy [9].

Statistics Canada recently produced its latest set of national and sub-national life tables for the period 2005 to 2007 using the Cass smoothing methodology to find both the optimal number of knots and their positions on age distribution in the context of constructing life tables [10]. The Canadian work is important as it confirms the accuracy of the Cass methodology, as explained in the report, *Methods for Constructing Life Tables for Canada*: "*An evaluation of the method by*

Kaishev et al (2009) was made for two regions, Canada as a whole and Newfoundland and Labrador. The number and the position of the knots given by the Kaishev et al model were very close to those established empirically"[11].

The impact of the research is broad since life tables are extensively used in pensions planning, demography, insurance, economics and medicine. For example, in order to price annuity and life insurance products and ensure the solvency of insurance companies through adequate reserves, actuaries must develop projections of future life expectancy and make risk assessments related to insured lives [12]. Life tables are a key tool for the work of pensions and life insurance actuaries. A key risk managed by actuaries is whether the progressive reductions in mortality rates seen in life tables will continue; this is often referred to as longevity risk. For instance, the general population life tables serve as a benchmark for pension schemes where they are used to produce reference survival statistics and are compared with within-scheme mortality experience.

Increasing longevity greatly affects government policy, because as longevity increases in the UK, people work for longer. On average, a healthy 65 year-old male can now expect to live for another 21 years, a 65 year-old female for another 24 years. Government legislation was introduced to encourage more people to save for their retirement through the purchase of annuities [13]. The Government consultation cites the English Life Tables 16 and ONS life expectancy tables which use the smoothing methodology designed by the Cass academics.

The advantage of the Geometrically Designed Splines method over existing statistical smoothing methods is in its intrinsic geometric nature. This leads to its extreme numerical efficiency and allows for fast smoothing of large real data samples from any field of real life application. In recognition of the enormous potential of the method, requests for implementation of the GeDS software have already been received from Intel Corporation and advanced instrument and sensor manufacturer, Aerodyne Research.

5. Sources to corroborate the impact

1. Government Actuary's Department Annual Report 2008-09, p.11, available on request.
2. Kaishev, Vladimir K; Haberman, Steven and Dimitrova, Dimitrina S (2009). Spline graduation of crude mortality rates for the English Life Table No. 16. In: English Life Tables No.16 (2000-02) Methodology, Office for National Statistics.
3. Office for National Statistics (2009). Decennial Life Tables 2000-02, Population Trends, Volume 136, No. 1, pp. 108-111.
4. Emailed testimony from GAD and ONS, available on request.
5. Office for National Statistics National Population Projections 2008-based: Chapter 7 – Mortality, Series PP2 No. 27. Editor: Steve Rowan. [Released 21 October 2009].
6. Data may be verified by the Demographic Analysis Unit - fertility, families, ageing and mortality, Population Statistics Division (PSD) Office for National Statistics UK.
7. General Register Office for Scotland (2012) Scottish Decennial Life Tables, 2000-02.
8. Northern Ireland Statistics and Research Agency (2012) Decennial Life Tables for Northern Ireland (2000–02).
9. McCormack, Kevin (2011). Graduation of Crude Mortality Rates for the Irish Life Tables, presented at the 33rd Conference on Applied Statistics in Ireland, 2013, Kildare, Ireland (15th-17th May), p. 16. Full presentation is available on request.
10. Government of Canada (2013). Methods for Constructing Life Tables for Canada, Provinces and Territories, Statistics Canada, ISBN 978-1-100-21511-2, published March 2013.
11. Corroboration may be sought from the Demography Division, Statistics Canada.
12. Groupe Consultatif Actuariel Européen Position Paper (2011). Use of age & disability as rating factors in insurance, published December 2011.
13. HM Treasury (2010). Removing the requirement to annuitise by age 75, The Outcome of a UK Government Consultation, published 15th July 2010.