Robust approach to life expectancy projection

Justyna Majewska and Grażyna Trzpiot

Abstract Mortality models most often are used to make projections of life expectancy. A good mortality model should satisfy some desirability criteria (Cairns et al (2008)). Models should be robust which means that parameter uncertainty should be low and small changes in the data should not result in significant changes in the estimates of the parameters and in their interpretation. Most of the existing mortality models are not robust against outliers due to wars, pandemics etc. or so called "longevity outliers". This paper is not the first attempt to deal with outliers in mortality data. Hyndman and Ullah (2007) used a combination of robust, nonparametric statistics and functional data analysis in developing a method for projection of age-specific mortality rates observed over time. While their objective was to identify and remove outliers, we highlight the necessity of incorporating them into projections in order to capture, in a more realistically way, perturbations that may occur in the future. The main contribution of this paper is to utilize a highly robust estimator to minimize the effect of outliers on point forecasts of life expectancy. We compare the point forecast accuracy and bias of seven stochastic models for life expectancy projection in the presence of outliers. Based on one-step ahead forecast errors we conclude that the Hyndman and Ullah method (2007) is the most accurate and

Justyna Majewska and Grażyna Trzpiot
University of Economics in Katowice, Department of Demography and Economic Statistics, 1 Maja 50, 40-287 Katowice, Poland
✉ justyna.majewska@ue.katowice.pl, grazyna.trzpiot@ue.katowice.pl

ARCHIVES OF DATA SCIENCE, SERIES A
(ONLINE FIRST) DOI 10.5445/KSP/1000058749/26
KIT SCIENTIFIC PUBLISHING ISSN 2363-9881
Vol. 2, No. 1, 2017

the least biased and in the Lee-Carter family of models the Lee-Carter (1992) and the Cairns-Blake-Dowd (2006) produce the most accurate point forecast of life expectancy when death rates across outlying years are replaced by highly robust estimates.

1 Introduction

Economic development, advances in medicine and healthcare, improvements in living conditions have resulted in a continuous increase in life expectancy during the last century all over the world. According to WHO (2015) in 2013 life expectancy at birth globally, for both sexes, was 71 years. Since 1990 life expectancy at birth has increased globally by 6 years. As people live longer, interest has shifted to the older generations. Nowadays, the global population aged 60 years can expect to live another 20 years on average, 2 years longer than in 1990.
Fundamental reforms in welfare policy have been taking place in many countries as a result of forecasts of an increasing elderly population. The population projections rely on age-specific projections of mortality rates. As mortality projections have become increasingly important, numerous models for mortality modelling and projection have been developed (e.g. Pollard (1987), Booth and Tickle (2008), Cairns et al (2008)). Because improvements in mortality projection may have an impact in guiding policy decisions regarding the allocation of current and future resources, the accurate modelling and projection of mortality rates and life expectancy are of growing interest to researchers (Hyndman and Ullah (2007)).

A wide variety of projection methods are in use, both between and within countries which produce different outcomes. Choosing the right model can easily become a challenge. The choice is usually supported with experts' opinions and knowledge, informed judgement, or assumptions on target levels of the life expectancy. Besides the observed past trends determine which method and historical period should be used. It is clear that in practice there does not exist a single "best" life expectancy prediction model for all countries. However, it is important to consider whether it is a good model or not. This requires a checklist of criteria against which a model can be assessed (Cairns et al (2008)). Robustness is one of the desired criteria. It means that parameter uncertainty should be low and small changes in the data should not result in significant changes in the estimates of the parameters and in the interpretation of them. Obviously, the presence of outliers has a serious effect on the modelling and projection of mortality and life expectancy.

Past life expectancy projections from official sources have generally underestimated the gains in life expectancy. As an example we present life expectancy projections (fig. 1) made in 2004 and 2013 for the 28 European Union member states (EU-28). Underestimation of life expectancy of both sexes is noticeable from 2005 to 2013. The assumption of fastest increase in average life expectancy in the next three decades results in one-year difference between these two projections in 2050. Such prediction failures for the EU-28 reduce the chance for an adequate anticipation of the need for additional investments in health and social services and pensions for the elderly. Table 1 presents

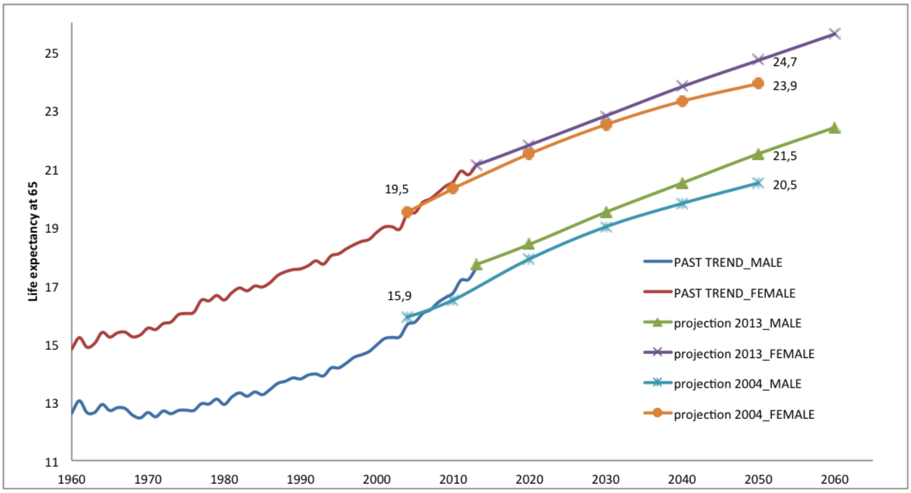

Fig. 1 Longevity trends (in years) at age 65 and projections for EU-28, 1960-2060
Data source: http://data.worldbank.org

how serious consequences may occur when life expectancy projections are inaccurate. Projections of the average life expectancy for women aged 65 in 2050 ($e_x = 23.9$ and $e_x = 24.7$) are the basis for calculating the ruin probabilities. The probability of "retirement ruin" is defined as the probability that a retirement (pension) plan is unsustainable. In short: under assumption of de-

Table 1 The probability of ruin for spending rates w (0.01, 0.02,....,0.1) for μ = 2.5% and σ=5%

e_x	0.01	0.02	0.03	0.04	0.05	**0.06**	0.07	0.08	0.09	0.1
23.9	0.005	0.036	0.097	0.183	0.284	**0.388**	0.489	0.582	0.663	0.732
24.7	0.005	0.039	0.121	0.204	0.326	**0.419**	0.520	0.607	0.743	0.769

fined contribution plan and randomness of capital S_t the probability of ruin: $P(PV_x > w) = P(\int_{x=0}^{T_x} e^{-(\mu t+\sigma B_t)}dt > w)$, where T_x is the future lifetime of an individual age x, μ is the drift modelling the trend of the capital investment, σ - the volatility modelling the diffusion of the capital investment. For details in calculation the probabilities of ruin see Trzpiot and Majewska (2015). The probability of ruin of a pensioners (here a woman at 65) can be interpreted in the following way: A woman with retirement age of 65, estimated life expectancy $e_x = 23.9$ and spending rate 0.06 (i.e. Euro 30 000 annually from a pension account of Euro 500 000) faces retirement ruin with probability 0.388. For $e_x = 24.7$ the probability of ruin is 0.419. So it is entirely clear that the underestimation of life expectancy might result in a significant higher probability of ruin.

In view of these facts, we want to examine the effect of using simple robust estimators in mortality models on the forecast accuracy of these models. As a result, we get information on the robustness of some existing stochastic mortality models from the Lee-Carter family. We use the outlier detection concept of Hyndman and Shang (2010) that is based on robust principal components analysis. Classical models are compared with the robust model of Hyndman and Ullah (2007). There is a difference in the meaning of robustness: The robust Hyndman-Ullah method assigns zero weight to outliers, while we use a very robust estimator to reduce the weights of outliers.

This article is organized as follows: In section 2 we describe briefly the mortality projection models that are included in our comparison. In section 3 we analyse the dataset in terms of the presence of outliers and compare the point forecast accuracy of these methods. The evaluations focus only on life expectancy. Results and conclusions appear in the last section of the paper.

2 Stochastic mortality models

The majority of existing stochastic models can be presented in the age-period-cohort framework. The general model can be written as (Hunt and Blake (2014)):

$$\eta\left(E\left(\frac{D_{xt}}{E_{xt}}\right)\right) = \alpha_x + \sum_{i=1}^{N} f_i\left(x,\theta^{(i)}\right)\kappa_t^{(i)} + \gamma_{t-x}. \quad (1)$$

This equation has the following components:

- The link function η transforms the observed data into a form suitable for modelling (the raw data usually consists of death counts D_{xt} and exposures to risk E_{xt} at ages x and for years t).
- The static age function α_x captures the general shape of the mortality curve.
- N age/period terms $f_i\left(x,\theta^{(i)}\right)\kappa_t^{(i)}$, consisting of companion pairs of t period terms $\kappa_t^{(i)}$, which give the evolution of mortality t rates through time, and age functions $f_i\left(x,\theta^{(i)}\right)$ that determine which segments of the age range these trends affect.
- Cohort parameters γ_{t-x} determine the lifelong effects that are specific to different generations denoted by their year of birth.

Parametric age functions $f_i\left(x,\theta^{(i)}\right)$ take a specific functional form and are parameterised by a small number of variables $\theta^{(i)}$ over more general non-parametric age functions $\beta^{(i)}$.
We consider mortality rates q_{xt} defined as the underlying probability that an individual aged x at year t will survive until year $t+1$, and m_{xt} - the underlying death rate.

2.1 Lee-Carter family of models

The Lee Carter model under a Poisson setting *(LC model)* assumes a Poisson distribution of the deaths (Brouhns et al (2002)). The model structure proposed by Lee and Carter (1992) assumes that there is a static age function, α_x, a unique non-parametric age-period term ($N=1$), and no cohort effect:

$$\log\, m_{xt} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)}. \quad (2)$$

In order to project the evolution of mortality, only the time-varying index κ_t needs to be projected under the assumption of an ARIMA process.

Renshaw and Haberman (2006) generalised the Lee-Carter model by incorporating a cohort effect (*RH model*):

$$\log m_{xt} = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \beta_x^{(3)} \gamma_{t-x}^{(3)}. \tag{3}$$

Mortality projections for this model are derived using the time series projection of the estimated κ_t and γ_{t-x}, generated using univariate ARIMA processes under the assumption of independence between the period and the cohort effects. To estimate the model Renshaw and Haberman (2006) assumed a Poisson distribution of deaths.

Cairns, Blake and Down proposed a model structure with two age-period terms ($N = 2$) with pre-specified age-modulating parameters $\beta_x^{(1)} = 1$ and $\beta_x^{(2)} = x - \bar{x}$, no static age function and no cohort effect (Cairns et al, 2006). Thus, the *CBD model* is given by:

$$\text{logit} q_{xt} = \kappa_t^{(1)} + \kappa_t^{(2)} (x - \bar{x}) \tag{4}$$

where $\bar{x}$ is the average age in the data. Cairns et al (2006) obtain mortality projections by projecting the period effects $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ using a bivariate random walk with drift.

Cairns et al. (2009) extended the original CBD model by adding a cohort effect and a quadratic age effect (*CBD extension*):

$$\text{logit} q_{xt} = \kappa_t^{(1)} + \kappa_t^{(2)} (x - \bar{x}) + \kappa_t^{(3)} [(x - \bar{x})^2 - \sigma_x^2] + \gamma_{t-x} \tag{5}$$

where σ_x^2 is the average value of $(x - \bar{x})^2$.

2.2 Robust approach to mortality modelling

Hyndman and Ullah (2007) showed a particular version of a Lee-Carter methodology the so-called functional demographic model. They proposed a methodology to forecast age-specific mortality rates, based on the combination of functional data analysis, nonparametric smoothing and robust statistics. The approach allows for smooth functions of age, is robust to outliers, and provides a modelling framework easy to fit to constraints and other information (Hyndman and Ullah (2007)).

This method (HUrob) utilizes the RAPCA algorithm (Hubert et al (2002)),

which stands for Reflection Algorithm for Principal Component Analysis, to obtain projection-pursuit estimates of principal components and their associated scores. Note that for spectral data, e.g. $n = 50$, $p = 1000$, the algorithm reduces the 1000-dimensional original data set to 49 dimensions. If $p \geq n$ the RAPCA method starts by reducing the data space to the affine subspace spanned by the n observations (it is done by a singular value decomposition). The main step of the RAPCA algorithm is to search for the direction in which the projected observations have the largest robust scale. To make the algorithm computationally feasible, the collection of directions to be investigated are restricted to all directions that pass through a highly robust location estimator (the L1-median). Having found the first direction, the data are reflected such that the first eigenvector is mapped onto the first basis vector, then the data are projected onto the orthogonal complement of the first eigenvector. This is simply done by omitting the first component of each (reflected) point. Doing so, the dimension of the projected data points can be reduced by one and consequently, the computations do not need to be done in the full dimensional space (Rousseeuw et al (2006)). The integrated squared error provides a measure of the accuracy of the principal component approximation for each year. Outlying years would result in a large integrated squared error (see Hyndman and Ullah (2007) for details). By assigning zero weight to outliers, the HUrob method can be used to model and forecast mortality rates without possible the influence of outliers.
In a broad analysis conducted by Shang et al (2011) the HUrob method provided more accurate forecasts of life expectancy than the simpler methods in section 3.2. A very important finding of their analysis is the fact that adopting robust estimation procedures minimises the effect of outliers on point forecasts. The approach taken into the consideration in this study, is based on a classical robust method where outliers are not excluded from the dataset. A good estimation method should be able to recover the underlying "normal" mortality varying pattern across age and year, with minimal effects of the noises including the outliers. Hence, death rates in the outlying years are replaced by a highly robust location estimator the spatial median.

3 Empirical analysis

The data sets were taken from the Human Mortality Database (HMD, 2015). For the analysis we have chosen UK female and male mortality data for age group 55-100 between 1922 and 1990. We restricted the age range to 55-100 as the CBD model and the extension of CBD model have been particularly designed to fit higher ages in the first part, identification of potential outliers in the development of age-specific mortality rates of the analysed population was conducted. In section 3.2, we compare the point forecast accuracy and bias of the seven stochastic models for life expectancy projection.

3.1 Identification of outliers

In order to identify potential outliers we use the approach proposed by Hyndman and Shang (2010). It treats mortality data as a time series of functions which are the realizations of the data on the functional space. Then these curves are visualized using functional equivalents of boxplots (or bagplots), and on this basis outliers in the observed curves are identified. The functional highest density region (HDR) boxplot displays the modal curve (i.e., the curve with the highest density), and the inner and outer regions. The functional HDR boxplot provides an additional advantage in that it can identify unusual "inliers" that fully in sparse regions of the sample space (Hyndman and Shang (2008)). Figures 2-5 show the functional highest density region (HDR) boxplots for two periods, 1922-1960 and 1961-1990, separet for males and females. Figures 2-5 were produced with the R package *rainbow* (Shang and Hyndman (2009)).

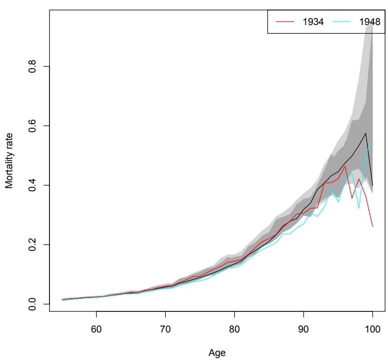

Fig. 2 The functional HDR boxplot [UK, 1922-1960, males, age: 55-100]

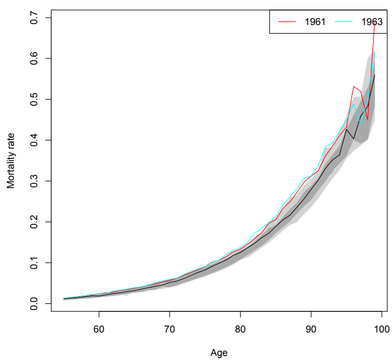

Fig. 3 The functional HDR boxplot [UK, 1961-1990, males, age: 55-100]

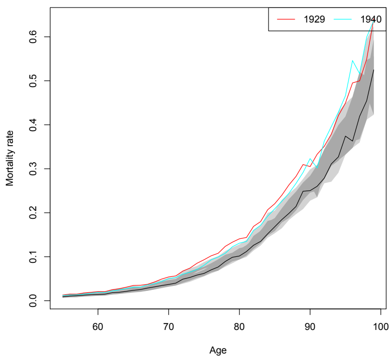

Fig. 4 The functional HDR boxplot [UK, 1922-1960, females, age: 55-100]

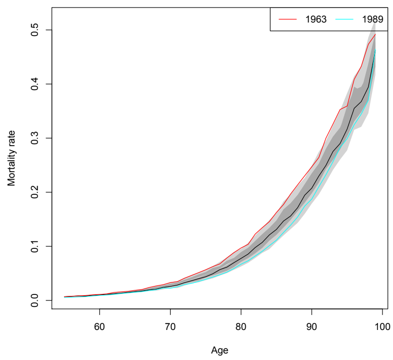

Fig. 5 The functional HDR boxplot [UK, 1961-1990, females, age: 55-100]

In figures 2- 5 the dark grey regions show the 50% HDR and the light grey regions show the 95% HDR. Curves outside these regions are identified as outliers. In each of these periods outliers clearly exist in the data and we seek to identify them. For male the population, in the first period the years 1934 and 1948 (in fig. 2) are outliers, in the second period the outliers (inf fig. 3) are 1961 and 1963. For the female population - in the first period the outlying years are 1929 and 1940 (see fig. 4). Becerra et al (2006) found strong evidence for power laws in casualty distributions for all disasters (natural and violent conflict), both globally and by continent except for North America and non-EU Europe. Moreover, their power-law findings for casualties in natural disasters fit in with established results for whole wars, terrorist events and events in individual modern wars. In the second period the outliers are 1963 and 1989 (see fig. 5).In addition, we observe significantly higher mortality rates in the age groups 85-100 years. Hence, we need methods that are not sensitive to the presence of outliers.

3.2 Projection of mortality and life expectancy

We have tested the following models: The classical LC model given by (3), the RH model (4), the CBD model (5), the extension of the CBD model (6), the HUrob model described in section 2.2, and the modification of the LC and

CBD models where the death rates across outlying year were replaced by their median.
All models were calibrated using two samples in order to determine their ability to make accurate projections. In the first test the models were calibrated with population mortality data using the period 1922 to 1960. Mortality rates were volatile during this period and the data also include the impact of World War II. We carried out stochastic projections from 1961 to 2011 for the ages of 55 to 100. The projections were then compared against the actual mortality experienced during the period to test the projection accuracy of the models. The comparison is is implemented as follows : Using the data in the fitting period, we compute one-step-ahead point projections, and determine the projection errors by comparing the projections with the actual out-of-sample data. Then, we increase the fitting period by one year, and compute one-step-ahead projections, and calculate the projection errors. In the second test the models were calibrated for the data for the period from 1961 to 1990. The mortality rates during this period were relatively smooth and show improvement over time. Projections were carried out from 1991 to 2011. Both test results were then compared and the life expectancy was calculated only for the age of 65.
To measure point forecast accuracy, we used the mean absolute forecast error (MAFE) and the mean forecast error (MFE). The MAFE is the average of absolute errors, and measures forecast precision, regardless of sign. The MFE is the average of errors and is a measure of bias. These measures are often used to evaluate forecasts of log mortality rates and life expectancyc (Shang et al, 2011; Shang, 2015).
Table 3 provides summaries of the point forecast accuracy based on the MAFEs for one-step-ahead forecasts of mortality rates averaged over different ages and years in the projection period for UK males. Table 2 shows the corresponding MFEs for one-step-ahead forecasts. Table 5 presents the MAFEs for one-step-ahead life expectancy point forecasts averaged over years in the projection period for males. Corresponding MFEs for one-step-ahead point forecasts of life expectancy are shown in Table 4.

Table 2 MFEs for one-step-ahead point forecasts of male log mortality rates by method. The mean is taken over ages and years in the projection period

Mortality model	MFE for 1961-2000	MFE for 2001-2011
LC	-0.431	-0.119
LC (Median)	-0.171	-0.013
RH model	-0.251	-0.064
CBD	0.859	0.537
CBD (Median)	0.245	0.329
CBD extension	-0.017	0.010
HUrob	0.000	0.000

Table 3 Point forecast accuracy of male log mortality rates by method, as measured by the MAFE for one-step-ahead forecasts. The mean is taken over ages and years in the projection period

Mortality model	MAFE for 1961-2000	MAFE for 2001-2011
LC	0.388	0.208
LC (Median)	0.183	0.110
RH model	0.205	0.157
CBD	0.327	0.183
CBD (Median)	0.158	0.075
CBD extension	0.282	0.125
HUrob	0.102	0.025

Table 4 MFEs for one-step-ahead point projections of female life expectancy by method. The mean is taken over ages and years in the projection period.

Mortality model	MFE for 1961-2000	MFE for 2001-2011
LC	-0.570	0.493
LC (Median)	-0.275	0.294
RH model	0.270	0.161
CBD	-0.213	-0.300
CBD (Median)	-0.201	-0.126
CBD extension	0.303	0.159
HUrob	0.001	0.001

Table 5 Point forecast accuracy of female life expectancy by method, as measured by the MAFE for one-step-ahead forecasts. The mean is taken over years in the projection period.

Mortality model	MAFE for 1961-2000	MAFE for 2001-2011
LC	0.615	0.693
LC (Median)	0.314	0.302
RH model	0.380	0.580
CBD	0.632	0.429
CBD (Median)	0.201	0.196
CBD extension	0.212	0.195
HUrob	0.147	0.162

4 Results and conclusions

Based on the information provided in tables 2-5 it is clear that the HUrob method tends to perform better than the other LC family methods. The HUrob method achieves the best point forecast accuracy for male life expectancy. Among the robust improvements of classical methods the CBD method performs best for males. The LC (Median) provides smaller errors than classical methods. The CBD (Median) method is superior among the LC methods. We also notice small errors for the RH method and the LC (Median). In general, there is a clear association between differences in the age patterns in forecast errors and differences in the size and sign of forecast errors in life expectancy. The HUrob method and the extensions of CBD methods tend to overestimate life expectancy, while the other LC methods underestimate life expectancy. Overall, the of our investigation findings regarding point forecasts of life expectancy rates indicate that the HUrob method is the most accurate and the least biased. In general, the HUrob method leads to smaller MAFE values than the LC family methods. This result was expected, and the same conclusions were obtained by Shang et al (2011). This is due to two factors. First, the smoothing of mortality rates means that the observational error is treated separately from dynamic changes over time. Second, the additional principal components allow more complicated dynamics to be modelled, rather than the restriction to simple age-specific time trends that result from a single principal component. Regarding the direction of bias, the HUrob method shows a tendency towards a one-step-ahead overestimation of the life expectancy for males. In contrast, the LC family methods exhibit a strong tendency towards a one-step-ahead underestimation of the life expectancy.

Comparisons between HUrob and LC methods with a robust estimator demonstrate the effect of adopting robust estimation procedures to minimise the effect of outliers on point forecasts. The most interesting finding is that a highly robust estimator considered in classical and simple models provides smaller forecast errors. The robust methods perform well in terms of accuracy and bias, both with regard to mortality rates and life expectancy. In each case, the reduction in forecast accuracy due to robust estimation is greatest when the forecasting period is longer.

This comparative analysis was limited to the most popular stochastic mortality models. Indication of the best model was beyond the aim of this study. Nevertheless, this analysis suggest to conduct a survey with different mortality models, length of historical and forecast periods. It must be also acknowledged that outliers identified in the case of the UK in the period 1922-1960 are specific, due to World War II. However, there is no doubt that any outlier may affect the predicted trend of mortality rates and the resulting mortality and life expectancy projections could be biased.

References

Becerra O, Johnson N, Meier P, Restrepo J, Spagat M (2006) Natural disasters, casualties and power laws: a comparative analysis with armed conflict. Proceedings of the Annual Meeting of the American Political Science Association URL `http://irevolution.files.wordpress.com/2011/07/apsa-paper-2006-harvard.pdf`

Booth H, Tickle L (2008) Mortality modelling and forecasting: a review of methods. Annals of Actuarial Science 3(1-2):3–43, DOI 10.1017/S1748499500000440

Brouhns N, Denuit M, Vermunt JK (2002) A poisson log-bilinear regression approach to the construction of projected lifetables. Insurance: Mathematics and Economics 31(3):373 – 393, DOI 10.1016/S0167-6687(02)00185-3

Cairns AJG, Blake D, Dowd K (2006) A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. Journal of Risk and Insurance 73(4):687–718, DOI 10.1111/j.1539-6975.2006.00195.x

Cairns AJG, Blake D, Dowd K (2008) Modelling and management of mortality risk: a review. Scandinavian Actuarial Journal 2008(2-3):79–113, DOI 10.1080/03461230802173608

Hubert M, Rousseeuw PJ, Verboven S (2002) A fast method for robust principal components with applications to chemometrics. Chemometrics and Intelligent Laboratory Systems 60(1):101 – 111, DOI 10.1016/S0169-7439(01)00188-5, fourth International Conference on Environ metrics and Chemometrics held in Las Vegas, NV, USA, 18-20 September 2000

Hunt A, Blake D (2014) A general procedure for constructing mortality models. North American Actuarial Journal 18(1):116–138, DOI 10.1080/10920277.2013.852963

Hyndman R, Shang HL (2008) Bagplots, Boxplots and Outlier Detection for Functional Data, Physica-Verlag HD, Heidelberg, pp 201–207. DOI 10.1007/978-3-7908-2062-1_31

Hyndman RJ, Shang HL (2010) Rainbow plots, bagplots, and boxplots for functional data. Journal of Computational and Graphical Statistics 19(1):29–45

Hyndman RJ, Ullah S (2007) Robust forecasting of mortality and fertility rates: A functional data approach. Computational Statistics and Data Analysis 51(10):4942–4956, DOI 10.1016/j.csda.2006.07.028

Lee RD, Carter LR (1992) Modeling and forecasting U.S. mortality. Journal of the American Statistical Association 87(419):659–671, DOI 10.2307/2290201

Pollard J (1987) Projection of age-specific mortality rates. Population Bulletin of the United Nations 21-22:55–69

Renshaw A, Haberman S (2006) A cohort-based extension to the Lee-Carter model for mortality reduction factors. Insurance: Mathematics and Economics 38(3):556 – 570, DOI 10.1016/j.insmatheco.2005.12.001

Rousseeuw PJ, Debruyne M, Engelen S, Hubert M (2006) Robustness and outlier detection in chemometrics. Critical Reviews in Analytical Chemistry 36(3-4):221–242, DOI 10.1080/10408340600969403

Shang H, Hyndman R (2009) R package: Rainbow: Rainbow plots, bagplots and boxplots for functional data. https://CRANR-projectorg/package=rainbow

Shang HL (2015) Statistically tested comparisons of the accuracy of forecasting methods for age-specific and sex-specific mortality and life expectancy. Population Studies 69(3):317–335, DOI 10.1080/00324728.2015.1074268, pMID: 26452306

Shang HL, Booth H, Hyndman R (2011) Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. Demographic Research 25(5):173–214, DOI 10.4054/DemRes.2011.25.5

Trzpiot G, Majewska J (2015) Modelling longevity risk. In: Trzpiot G (ed) Modelowanie wielowymiarowych struktur danych i analiza ryzyka, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, chap 5, pp 41–49