Fuzzy Mortality Model Based on the Algebra of Oriented Fuzzy Numbers

Andrzej Szymański*, Agnieszka Rossa*

1 Introduction

Since the introduction of the Lee–Carter model in [5] proposed to forecast the trend of age-specific mortality rates, a range of mortality models have been presented with modeling target defined as the probability of death, the age-specific mortality rate or the force of mortality. The main difficulty in the use of the Lee–Carter model and its stochastic modifications is due to the assumed homogeneity of the random term. However, this property is not confirmed by the analysis of the real-life data. The problem prompted search for solutions that could do without this assumption. One of the possible options is to set research in the framework of the fuzzy numbers. This line of thinking was adopted by Koissi and Shapiro [2], where empirical observations and the model's parameters of the Lee-Carter model were treated as fuzzy symmetric triangular numbers. The Koissi–Shapiro model involves however some problems with parameters' estimation, which arise from the necessity to find the minimum of a criterion function with a max-type operators incorporated in it. Such an optimization task cannot be solved using standard algorithms. One approach to such a problem can be the Banach algebra of oriented fuzzy numbers (OFN), developed by Kosiński with co-authors [3, 4]. In this paper the fuzzy mortality model obtained by applying the OFN algebra to the Koissi–Shapiro model will be presented. Prediction accuracy of the proposed model with analogous results obtained with the use of the Lee–Carter mortality model will be also discussed.

2 The Koissi–Shapiro model

One of the most interesting generalizations of the Lee–Carter model (LC) referring to the algebra of fuzzy numbers is the fuzzy Lee–Carter model introduced by Koissi and Shapiro [2]. This version of the mortality model assumes fuzzy representation of the log-central death rates as well as model's parameters. Such an approach allows taking account of uncertainty involved in mortality rates and entering a random term into the fuzzy structure of the model.

* Department of Demography, Institute of Statistics and Demography, University of Łódź, Poland, e-mail: anszyman@interia.pl; agnieszka.rossa@uni.lodz.pl

This concept builds on the assumption that the real log-central rates of mortality $\ln m_{x,t}$ are not exactly known for each age group x and time period t, thus the role of the explanatory variable is played by log-central mortality rates which are fuzzy numbers.

It is well-known that death statistics are subject to reporting errors of several kinds. They may be reported for incorrect year, area, or assigned statistics that are incorrect, e.g. age. Moreover, the midyear population data that serve as the denominators of mortality rates are also the subject of errors. It is regarded as the population at July 1 and is assumed to be the point at which half of the deaths in the population during the year have appeared. Such an estimate can be underestimated or overestimated. For these reasons, fuzzy representation of the mortality rates seems to be justified.

Koissi and Shapiro created fuzzy death rates by converting log-central mortality rates $\ln m_{x,t}$ into symmetric, triangular fuzzy numbers $Y_{x,t}$ expressed by ordered pairs $(y_{x,t}, e_{x,t})$, i.e.

$$Y_{x,t} = (y_{x,t}, e_{x,t}), \quad x = 0, 1, \ldots, X, \ t = 1, 2, \ldots, T, \tag{2.1}$$

where $y_{x,t} = \ln m_{x,t}$ are the crisp central values and "fuzziness parameters" $e_{x,t}$ serve as spreads of fuzzy numbers $Y_{x,t}$.

The Koissi–Shapiro model is defined as

$$Y_{x,t} = A_x \oplus (B_x \odot K_t), \quad x = 0, 1, \ldots, X, \ t = 1, 2, \ldots, T, \tag{2.2}$$

where $A_x = (\alpha_x, s_{A_x}), B_x = (\beta_x, s_{B_x}), K_t = (\kappa_x, s_{K_t})$ are fuzzy triangular symmetric numbers with central values $\alpha_x, \beta_x, \kappa_x$ and spreads $s_{A_x}, s_{B_x}, s_{K_x}$, respectively, while $\oplus$, $\odot$ are the addition and multiplication operators of fuzzy numbers (see [2] for details).

Koissi and Shapiro assumed that unknown model's parameters $\alpha_x, \beta_x, \kappa_x$ and $s_{A_x}, s_{B_x}, s_{K_x}$ can be estimated minimizing a criterion function based on the so-called Diamond distance [1]. Two components S_1, S_2 of the criterion function $S = S_1 + S_2$ can be written as

$$\begin{aligned} S_1(\alpha_x, \beta_x, \kappa_t) = \sum_{x=0}^{X} \sum_{t=1}^{T} \big[& 3\alpha_x^2 + 3(\beta_x \kappa_t)^2 + 3y_{x,t}^2 + 6\alpha_x \beta_x \kappa_t \\ & - 4\alpha_x y_{x,t} - 4\beta_x \kappa_t y_{x,t} + 2e_{x,t}^2 \big], \\ S_2(\beta_x, \kappa_t, s_{A_x}, s_{B_x}, s_{K_t}) = 2\sum_{x=0}^{X} \sum_{t=1}^{T} \big[& (\max\{s_{A_x}, |\beta_x| s_{K_t}, |\kappa_t| s_{B_x}\})^2 \\ & - 2e_{x,t} \max\{s_{A_x}, |\beta_x| s_{K_t}, |\kappa_t| s_{B_x}\} \big]. \end{aligned} \tag{2.3}$$

This concept poses however a major problem in estimation, since the expression $\max\{s_{A_x}, |\beta_x| s_{K_t}, |\kappa_t| s_{B_x}\}$ appearing twice in S_2 prevents the use of standard non-linear optimization methods. In the next section a fuzzy mortality model simplifying the estimation procedure is proposed.

3 The modified Koissi–Shapiro model

In the modified fuzzy version of the Lee–Carter model, mortality rates and model's parameters are represented by means of oriented fuzzy numbers. The new model will be termed the Extended Fuzzy Lee–Carter model (EFLC).

It is assumed that the log-central death rates have the OFN representation $\vec{Y}_{x,t} = (f_{Y_{x,t}}, g_{Y_{x,t}})$ with functions $f_{Y_{x,t}}, g_{Y_{x,t}}$ defined for $u \in [0,1]$ as

$$\begin{aligned} f_{Y_{x,t}}(u) &= y_{x,t} - e_{x,t}(1-u), \\ g_{Y_{x,t}}(u) &= y_{x,t} + e_{x,t}(1-u), \end{aligned} \tag{3.1}$$

where $y_{x,t} = \ln m_{x,t}$ are (crisp) log-central death rates and $e_{x,t}$ are fuzziness parameters obtained by means of fuzzification procedure given by Koissi and Shapiro [2].

The EFLC model can then be written as

$$\vec{Y}_{x,t} = \vec{A}_x \oplus (\vec{B}_x \otimes \vec{K}_t), \quad x = 0, 1, \ldots, X, \; t = 1, 2 \ldots, T, \tag{3.2}$$

where $\vec{A}_x$, $\vec{B}_x$, $\vec{K}_t$ are OFN's expressed by means of the following ordered pairs

$$\vec{A}_x = (f_{A_x}, g_{A_x}), \quad \vec{B}_x = (f_{B_x}, g_{B_x}), \quad \vec{K}_t = (f_{K_t}, g_{K_t}), \tag{3.3}$$

with functions f_{A_x}, g_{A_x}, f_{B_x}, g_{B_x} and f_{K_t}, g_{K_t} defined for $u \in [0,1]$ as

$$\begin{aligned} f_{A_x}(u) &= a_x - (1-u)s_{A_x}, \quad & g_{A_x}(u) &= a_x + (1-u)s_{A_x}, \\ f_{B_x}(u) &= b_x - (1-u)s_{B_x}, \quad & g_{B_x}(u) &= b_x + (1-u)s_{B_x}, \\ f_{K_t}(u) &= k_t - (1-u)s_{K_t}, \quad & g_{K_t}(u) &= k_t + (1-u)s_{K_t}. \end{aligned} \tag{3.4}$$

For the EFLC model, it is assumed that the unknown model's parameters are a_x, b_x, k_t and $s_{A_x}, s_{B_x}, s_{K_t}$ incorporated in functions (3.4).

Using the addition and multiplication operators $\oplus, \otimes$ for oriented fuzzy numbers (see [3] for details), the right-hand side of (3.2) takes the form

$$\begin{aligned} \vec{A}_x \oplus (\vec{B}_x \otimes \vec{K}_t) &= (f_{A_x}, g_{A_x}) \oplus (f_{B_x \otimes K_t}, g_{B_x \otimes K_t}) = \\ &= (f_{A_x \oplus B_x \otimes K_t}, \; g_{A_x \oplus B_x \otimes K_t}), \end{aligned} \tag{3.5}$$

where

$$\begin{aligned} f_{A_x \oplus B_x \otimes K_t}(u) &= a_x + b_x k_t - (s_{A_x} + k_t s_{B_x} + b_x s_{K_t})(1-u) + s_{B_x} s_{K_t}(1-u)^2, \\ g_{A_x \oplus B_x \otimes K_t}(u) &= a_x + b_x k_t + (s_{A_x} + k_t s_{B_x} + b_x s_{K_t})(1-u) + s_{B_x} s_{K_t}(1-u)^2. \end{aligned} \tag{3.6}$$

Expressions $s_{B_x} s_{K_t}(1-u)^2$ in (3.6) are close to 0 for small values of s_{B_x}, s_{K_t} and for $u \in [0,1]$. Given this, we consider the following two approximation

$$\begin{aligned} f_{A_x \oplus B_x \otimes K_t}(u) &\approx a_x + b_x k_t - (s_{A_x} + k_t s_{B_x} + b_x s_{K_t})(1-u), \\ g_{A_x \oplus B_x \otimes K_t}(u) &\approx a_x + b_x k_t + (s_{A_x} + k_t s_{B_x} + b_x s_{K_t})(1-u). \end{aligned} \tag{3.7}$$

It follows from (3.7) that right-hand side of (3.2) corresponds to symmetric triangular numbers with central values $a_x + b_x k_t$ and spreads approximated by $s_{A_x} + k_t s_{B_x} + b_x s_{K_t}$.

4 Parameters' estimation

To estimate the parameters of the EFLC model, we applied the Diamond distance $D^2(\cdot,\cdot)$ between the right and left sides of (3.2). Then the estimation problem reduces the minimization of the following criterion

$$F(a_x, b_x, k_t, s_{A_x}, s_{B_x}, s_{K_t}) = \sum_{x=0}^{X}\sum_{t=1}^{T} D^2\left(\vec{A}_x \oplus (\vec{B}_x \otimes \vec{K}_t),\ \vec{Y}_{x,t}\right). \qquad (4.1)$$

To illustrate the theoretical discussions presenting the proposal of a new fuzzy model and estimation results, the LC and EFLC models were estimated based on the age-specific death rates for males and females in Poland from years 1990–2018. Data were sourced from the Human Mortality Database (mortality.org) and the database of the Polish Central Statistical Office (stat.gov.pl). The 2014–2018 death rates were only used to evaluate prediction accuracy.

In the analysis, the ex-post crisp forecasting errors were evaluated and compared with the errors yielded by the LC model. It appeared that the mortality forecasts obtained with the EFLC model produced smaller or comparable prediction errors in relation to the standard LC model. Thus, in terms of prediction accuracy, the EFLC model utilizing oriented fuzzy numbers appeared to be rather similar to the standard Lee-Carter model. What makes it superior to the LC model, is that it allows the areas of fuzziness of the estimated parameters to be determined, and consequently the areas of fuzziness for predicted mortality rates. Another advantage of the EFLC model is that the areas of fuzziness can be identified without assuming any probability distribution of mortality data.

Acknowledgement

The research was supported by a grant from the National Science Centre, Poland, under contract UMO-2015/17/B/HS4/00927.

References

[1] Diamond P., 1988, *Fuzzy least-squares*, Information Sc., 46(3), 141–157.

[2] Koissi M. C., Shapiro A. F., 2006, *Fuzzy formulation of the Lee–Carter model for mortality forecasting*, Insurance: Mathematics and Economics, 39, 287–309.

[3] Kosiński W., Prokopowicz P., 2004, *Algebra liczb rozmytych*, Matematyka Stosowana, 46, 37–63.

[4] Kosiński W., Prokopowicz P., Ślęzak D., 2003, *Ordered fuzzy numbers*, Bulletin of the Polish Academy of Sciences – Mathematics, 51, 327–338.

[5] Lee R. D., Carter L., 1992, *Modeling and forecasting the time series of U.S. mortality*, Journal of the Am. Stat. Assoc., 87, 659–671.