FORECASTING MORTALITY IN SUBPOPULATIONS USING LEE-CARTER TYPE MODELS: A COMPARISON

IVAN LUCIANO DANESI [1]◇, STEVEN HABERMAN [2]†,
AND PIETRO MILLOSSOVICH [3]†

Abstract. The relative performance of multipopulation stochastic mortality models is investigated. When targeting mortality rates, we consider five extensions of the well known Lee-Carter single population extrapolative approach. As an alternative, we consider similar structures when mortality improvement rates are targeted. We use a dataset of deaths and exposures of Italian regions for the years 1974-2008 to conduct a comparison of the models, running a battery of tests to assess the relative goodness of fit and forecasting capability of the different approaches. Results show that the preferable models are those striking a balance between complexity and flexibility.

1. Introduction

Several aspects of modern societies are affected by the level and trend of mortality rates. For example, the private and public retirement systems, as well as other components of the social security system, are planned and modified according to the values taken by current and forecast values of death rates.

Several mortality forecasting models have been proposed in the last few decades. Among the extrapolative methods, that of Lee-Carter (see Lee and Carter [17]) has been the most successful and has since received a great deal of attention. This model has been extensively studied and has been extended in several directions, see Pitacco et al. [22] and De Jong and Tickle [8] for a review.

In many instances, one is interested in forecasting mortality rates for more than one population. Although the separate modelling of each population under scrutiny is possible, it would neglect any existing interaction that motivated the analysis in the first place. Therefore one should focus on a framework where death rates in the populations under study are jointly modelled, in order to allow for correlation between mortality dynamics. As a first example, demographers have long been interested in the study of mortality of males and females in a given population (Lee and Carter [17]). More generally, a population could be split according to some characteristics - smoking habit, occupation, income - in order to analyse the mortality of each subgroup. A similar investigation may involve the populations of related countries or regions of a given country (see for instance Delwarde et al.

This draft: December 4, 2014.

◇ Department of Statistics, University of Padua, Padua, Italy.

† Faculty of Actuarial Science and Insurance, Cass Business School, City University London, London, UK.

[1] danesi@stat.unipd.it

[2] s.haberman@city.ac.uk

[3] pietro.millossovich.1@city.ac.uk

We thank Ermanno Pitacco, Nicola Torelli, Andrés Villegas, Andrew Hunt, Jens Perch Nielsen, Lola Miranda Martines for helpful comments and suggestions. We acknowledge financial support from University of Trieste and Cass Business School. Earlier versions of the paper have been presented at the University of Padua, KU Leuven, the GSP 2013 in Brixen, the IME 2013 in Copenhagen, the IWSM 2013 in Palermo, the ARC 2013 in Philadelphia and the 8th Conference in Actuarial Science & Finance on Samos.

[9], Booth et al. [3]). As a final example, the joint modelling of two populations is the key of any longevity basis risk assessment exercise, see Coughlan et al. [7].

The possibility of extending forecasting methods to related populations has been explored by several researchers, including Lee and Carter [17], Li and Lee [19], Brouhns et al. [4], Russolillo et al. [25], Li and Hardy [18], Dowd et al. [10], Jarner and Kryger [16], Haberman and Villegas [13]. In the context of single population forecasting models, extensive comparisons have been carried out by Cairns et al. [6] and Haberman and Renshaw [11]. In this paper we investigate the relative performance of multipopulation mortality models. More precisely, we considered five parametric structures where death rates of related populations are jointly modelled through a Lee-Carter type formulation. The models include many (but are not limited to) existing contribution in the literature and allow for varying degrees of interaction and complexities between the considered populations. We also examine the performance of five similar parametric structures where the improvement rates, rather than mortality rates themselves, are targeted. The feasibility of such approach has been studied by Haberman and Renshaw [12]. Here, mortality rates are transformed into improvement rates which are then modelled and forecasted. Finally, the inverse transformation is used to calculate the corresponding forecasted mortality rates.

We fit the ten models to a data set of deaths and exposures spanning 18 regions of Italy for the years 1974-2008 and the ages 20-89. We assess the performance of the models against several indicators, including information criteria, in- and out-of-sample goodness of fit of individual death rates and truncated residual lifetimes. The analysis seems to indicate that the most preferred models are those achieving a compromise between complexity of the parametric structure and flexibility in fitting past and future trends.

The structure of the paper is as follows. In Section 2, we present a short overview on Lee-Carter model and the proposed variations. In Section 3, the models are applied to the Italian regional mortality data the indicators used to assess and compare the different models are presented. Section 4 discusses the results.

2. The models

2.1. **The Lee-Carter model.** The original formulation of the model presented in Lee and Carter [17] is

$$\ln m_{x,t} = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t}, \ \sum_x \beta_x = 1, \ \sum_t \kappa_t = 0,$$

where $m_{x,t}$ is the (central) mortality rate relative to age x and calendar year t. The logarithm of $m_{x,t}$ is then specified through an age specific term, α_x, which represents the general mortality shape across age, a bilinear term $\beta_x \kappa_t$ and an error term $\varepsilon_{x,t} \sim N(0, \sigma^2)$. The bilinear term is composed of κ_t, an index representing the change in the level of mortality across time, and β_x, the age specific response to variations in the time index. As the model written in this way is overparametrised, the two additional constraints are introduced in order to identify the model.

In this formulation the random errors are homoskedastic, which is usually a strong and often unrealistic hypothesis. To overcome this problem, Brouhns et al. [4] proposed a modification of the Lee-Carter model where the number of deaths is specified as a Poisson random variable:

$$D_{x,t} \sim \text{Poisson}\left(\text{ETR}_{x,t}\, m_{x,t}\right)$$

where $D_{x,t}$ is the number of deaths aged x last birthday in calendar year t, $\mathrm{ETR}_{x,t}$ is the corresponding central exposed to risk and

$$\log m_{x,t} = \alpha_x + \beta_x \kappa_t, \quad \sum_x \beta_x = 1, \ \sum_t \kappa_t = 0 \tag{2.1}$$

which has the form of the Lee-Carter model, except for the error term.

2.2. **The mortality improvement rate.** An alternative approach proposed in the literature is to model the improvement in mortality rates, rather than the rate itself, see for example Mitchell et al. [21], Haberman and Renshaw [12] and references therein. In Haberman and Renshaw [12], the year-on-year mortality improvement rate is defined by

$$z_{x,t} = 2\frac{1 - m_{x,t}/m_{x,t-1}}{1 + m_{x,t}/m_{x,t-1}}. \tag{2.2}$$

The expression in Equation (2.2) is immediately seen to be the ratio between the incremental mortality improvement $m_{x,t-1} - m_{x,t}$ and the average $(m_{x,t} + m_{x,t-1})/2$ of two adjacent mortality rates. This formulation allows for some smoothing of improvement rates, which are known to be noisy. Alternative definitions of improvement rates can be found in Richards et al. [24] and Baxter [1]. The values of $z_{x,t}$ are then modelled as realizations of independent Gaussian random variables $Z_{x,t}$ assuming constant dispersion, $Z_{x,t} \sim N(\eta_{x,t}, \sigma^2)$. The following first moment predictor structure is then considered:

$$\eta_{x,t} = \beta_x \kappa_t, \quad \sum_x \beta_x = 1. \tag{2.3}$$

Clearly, the time indices κ_t in (2.3) have a different interpretation than those in (2.1). As noted by Haberman and Renshaw [12], if mortality rates are modelled through (2.1), the time indices in (2.3) are then approximately equal to the discrete version of the log derivative of the corresponding time indices in (2.1).

2.3. **The parametric structures.** From now on, the index $i = 1, \ldots, I$ denotes subpopulation i among the I (with $I \geq 2$) populations under study. For each i, we assume that the following data are available: for ages $x = x_1, \ldots, x_k$ and (consecutive) calendar years $t = t_1, \ldots, t_n$,

- $D^i_{x,t}$ — number of deaths aged x last birthday in calendar year t
- $\mathrm{ETR}^i_{x,t}$ — corresponding central exposure.

We then compute the corresponding central death rates and year-on-year improvement rates

$$m^i_{x,t} = \frac{D^i_{x,t}}{\mathrm{ETR}^i_{x,t}}, \ z^i_{x,t} = \frac{1 - m^i_{x,t}/m^i_{x,t-1}}{1 + m^i_{x,t}/m^i_{x,t-1}}.$$

Note that, for each age and subpopulation, there are n death rates and $n-1$ improvement rates that can be computed. From now on we assume, as in Brouhns et al. [4], that the force of mortality, denoted $\mu^i_{x,t}$, satisfies $\mu^i_{x+u,t+u} = \mu^i_{x,t}$ for all i, integers x and t and $0 \leq u < 1$. It follows that $\mu^i_{x,t} = m^i_{x,t}$ and both quantities can be used interchangeably.

The Brouhns et al. [4] version of the Lee-Carter model, recalled in (2.1), specifies the numbers of deaths as Poisson random variables,

$$D^i_{x,t} \sim \mathrm{Poisson}(\mathrm{ETR}^i_{x,t}\, m^i_{x,t}),$$

independent across ages, years and subpopulations. The mean of these variables is modelled through a number L of time factors, according to

$$\log m^i_{x,t} = \alpha^i_x + \sum_{j=1}^{L} \beta^i_{x,j}\kappa_{t,j}. \tag{2.4}$$

This expression is in spirit similar to those found in Booth et al. [2] and Hyndman and Ullah [15].

When modelling improvement rates, it is assumed that the $z^i_{x,t}$ are realizations of Gaussian random variables

$$Z^i_{x,t} \sim N(\eta^i_{x,t}, \sigma^2_i),$$

independent across ages, years and subpopulations. Note that the variance is allowed to vary among populations. We express the mean of these variables by generalising (2.3) in a form similar to (2.4), that is

$$\eta^i_{x,t} = \sum_{j=1}^{L} \beta^i_{x,j}\kappa_{t,j}, \tag{2.5}$$

where again the meaning of the time indices is different from those in (2.4).

The aim of (2.4)-(2.5) is to introduce a general framework allowing for different levels of complexities and interactions within and between the subpopulations. The number L of factors will be typically related to the number of populations and the chosen degree of complexity. Some particular cases of (2.4) and (2.5) are considered, in order to make estimation feasible and to ease comparison between models. More precisely, five specifications of (2.4), called P-models, are introduced. Subsequently, the counterparts of these five models in terms of mortality improvement rates (2.5), called M-models, are presented.

(1) P-double:
$$\log m^i_{x,t} = \alpha^i_x + \beta^i_{x,1}\kappa^i_{t,1} + \beta^i_{x,2}\kappa^i_{t,2},$$
with the identifiability constraints $\sum_t \kappa^i_{t,1} = 0$, $\sum_x \beta^i_{x,1} = 1$, $\sum_t \kappa^i_{t,2} = 0$, $\sum_x \beta^i_{x,2} = 1$, $\sum_t \kappa^i_{t,1}\kappa^i_{t,2} = 0$ and $\sum_x \beta^i_{x,1}\beta^i_{x,2} = 0$ for all i.

(2) P-common:
$$\log m^i_{x,t} = \alpha^i_x + \beta^i_{x,1}\kappa_{t,1} + \beta^i_{x,2}\kappa^i_{t,2},$$
with the identifiability constraints $\sum_t \kappa_t = 0$, $\sum_x \beta^i_{x,1} = 1$, $\sum_t \kappa^i_t = 0$, $\sum_x \beta^i_{x,2} = 1$, $\sum_t \kappa_{t,1}\kappa^i_{t,2} = 0$ and $\sum_x \beta^i_{x,1}\beta^i_{x,2} = 0$ for all i.

(3) P-simple:
$$\log m^i_{x,t} = \alpha^i_x + \beta^i_x\kappa^i_t,$$
with the identifiability constraints $\sum_t \kappa^i_t = 0$, $\sum_x \beta^i_x = 1$ for all i.

(4) P-division:
$$\log m^i_{x,t} = \alpha^i_x + \beta^i_x\kappa^i_t,$$
with $\kappa^i_t = \kappa^{(h)}_t$ for $i \in J_h$, where $J_1, \ldots, J_{I'}$ is a partition of $\{1, \ldots, I\}$; the identifiability constraints are $\sum_t \kappa^{(h)}_t = 0$ and $\sum_{i\in J_h, x} \beta^i_x = |J_h|$ for $h = 1, \ldots, I'$. Here $|J|$ is the cardinality of the set J.

(5) P-one:
$$\log m^i_{x,t} = \alpha^i_x + \beta^i_x\kappa_t,$$
with the identifiability constraints $\sum_t \kappa_t = 0$ and $\sum_{i,x} \beta^i_x = I$.

(6) M-double:
$$\eta^i_{x,t} = \beta^i_{x,1}\kappa^i_{t,1} + \beta^i_{x,2}\kappa^i_{t,2},$$
with the identifiability constraints $\sum_x \beta^i_{x,1} = 1$, $\sum_x \beta^i_{x,2} = 1$, $\sum_t \kappa^i_{t,1}\kappa^i_{t,2} = 0$ and $\sum_x \beta^i_{x,1}\beta^i_{x,2} = 0$ for all i.

(7) M-common:
$$\eta^i_{x,t} = \beta^i_{x,1}\kappa_{t,1} + \beta^i_{x,2}\kappa^i_{t,2},$$
with the identifiability constraints $\sum_x \beta^i_{x,1} = 1$, $\sum_x \beta^i_{x,2} = 1$, $\sum_t \kappa_{t,1}\kappa^i_{t,2} = 0$ and $\sum_x \beta^i_{x,1}\beta^i_{x,2} = 0$ for all i.

(8) M-simple:
$$\eta^i_{x,t} = \beta^i_x \kappa^i_t,$$
with the identifiability constraint $\sum_x \beta^i_x = 1$ for all i.

(9) M-division:
$$\eta^i_{x,t} = \beta^i_x \kappa^i_t,$$
with $\kappa^i_t = \kappa^{(h)}_t$ for $i \in J_h$, where $J_1, \ldots, J_{I'}$ is a partition of $\{1, \ldots, I\}$; the identifiability constraints are $\sum_{i \in J_h, x} \beta^i_x = |J_h|$ for $h = 1, \ldots, I'$.

(10) M-one:
$$\eta^i_{x,t} = \beta^i_x \kappa_t,$$
with the identifiability constraint $\sum_{i,x} \beta^i_x = I$.

2.4. **Discussion of the models.** Models (1) and (6) (P-double and M-double) are inspired by Renshaw and Haberman [23] and Booth et al. [2], where a single population Lee-Carter model with two bilinear components is considered. The addition of a bilinear component aims at capturing possible time trends that a single component would fail to reproduce.

Models (2) and (7) (P-common and M-common) exhibit a common and a population specific time index. They can be obtained as a restriction of (1) and (6) respectively, where one of the two time indices is constrained to be the same across all subpopulations. Model P-double, known as augmented common factor model (see Li and Hardy [18]), is inspired by Li and Lee [19], where a common factor is estimated on the considered populations and, in a second stage, a second bilinear component, which can be interpreted as the spread from the common trend, is estimated separately for every population.

Models (3) and (8) (P-simple and M-simple) are obtained from (2) and (7) by removing the common factor. In other words, in each subpopulation mortality is specified through a Lee-Carter model featuring a population specific bilinear component.

The idea behind models (4) and (9) (P-division and M-division) is that, among the I subpopulations under study, some can be grouped to form $I' < I$ clusters. It is implicit that the populations in each cluster share some common characteristics and can be jointly modelled introducing a common time trend. Groups of subpopulations can be identified through standard clustering techniques.

Finally, models (5) and (10) (P-one and M-one), sometimes known as common factor or joint κ model, assumes that there is a unique cluster of subpopulations. This is equivalent to assuming that mortality improvement in all subpopulations are driven by a single time index, while keeping different age varying coefficients. Alternatively, (5) and (10) can be obtained from (2) and (7) by eliminating the specific time factor. This model, or a version of it, was considered, among others, by Li and Hardy [18], Russolillo et al. [25], Delwarde et al. [9].

Both set of models (1)-(5) and (6)-(10) are presented in decreasing order of complexity and number of parameters, see Table 1. More precisely, each model can be obtained from the preceding one by restricting appropriately some of the parameters. In other words (1)-(5) and (6)-(10) are complete sequences of nested models.

Note that, in models (1), (2) and (6), (7), where more than one period index appear for each subpopulation, the the orthogonality constraints suggested by Hunt and Blake [14] are used.

The models considered allow for different degrees of interactions between the subpopulations. Trivial correlation between improvement rates in different subpopulations holds in models such as P-one and M-one where a single time index drive changes in all groups. Models P-division and M-division imply perfectly correlated improvement rates within each cluster, while allowing for non perfect correlation between different clusters. Models P/M-simple, -common and -double all feature at least a subpopulation specific time index so that improvement rates correlation is never trivial and its range is allowed to vary with the complexity of the model.

model	number of time factors	number of parameters	number of constraints
P-double	$2I$	$(3k+2n)I$	$6I$
P-common	$I+1$	$(2k+n)(I+1)$	$4I+2$
P-simple	I	$(2k+n)I$	$2I$
P-division	I'	$2kI+nI'$	$2I'$
P-one	1	$2kI+n$	2
M-double	$2I$	$2(k+n')I$	$4I$
M-common	$I+1$	$(k+n')(I+1)$	$3I+1$
M-simple	I	$(k+n')I$	I
M-division	I'	$kI+n'I'$	I'
M-one	1	$kI+n'$	1

TABLE 1. Number of time factors, parameters and constraints for the ten models ($n' = n-1$).

3. The application

This section introduces the dataset used and the statistics and indices employed to assess the models. The comments on the corresponding outputs are deferred to Section 4.

3.1. **The dataset.** The models presented in Section 2 are applied to a multiple population mortality dataset containing the mortality rates of Italian regions. Italy is divided into 18 regions out of the official 20, since two small sized regions (Val d'Aosta and Molise) are merged with one of their neighbours.

In general, regions of a country are related as they share some common characteristics. However, it is also true that Italian regions are inhomogeneous, either economically as well as along other dimensions, and this is reflected in the mortality experience. In this application $I = 18$, and the index $i = 1, \ldots, 18$ is used for indicating the regions, in this order: Piemonte-Valle d'Aosta, Lombardia, Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Liguria, Emilia-Romagna, Toscana, Umbria, Marche, Lazio, Abruzzo-Molise, Campania, Puglia, Basilicata, Calabria, Sicilia, Sardegna. The geographical areas can be seen in Figure 1.

The data cover a span of 35 years, from 1974 to 2008[1]. In the analysis the focus is on male mortality data for the age range 20-89.

Figure 2, contain four plots representing $\log m^i_{x,t}$ for four fixed ages and three out of the eighteen populations (Lombardia, Lazio and Sicilia). These plots confirm that the evolution of mortality follows similar patterns for the different populations.

[1]The data were provided by Istat (`www.istat.it`).

FIGURE 1. Italy divided in the considered 18 areas.

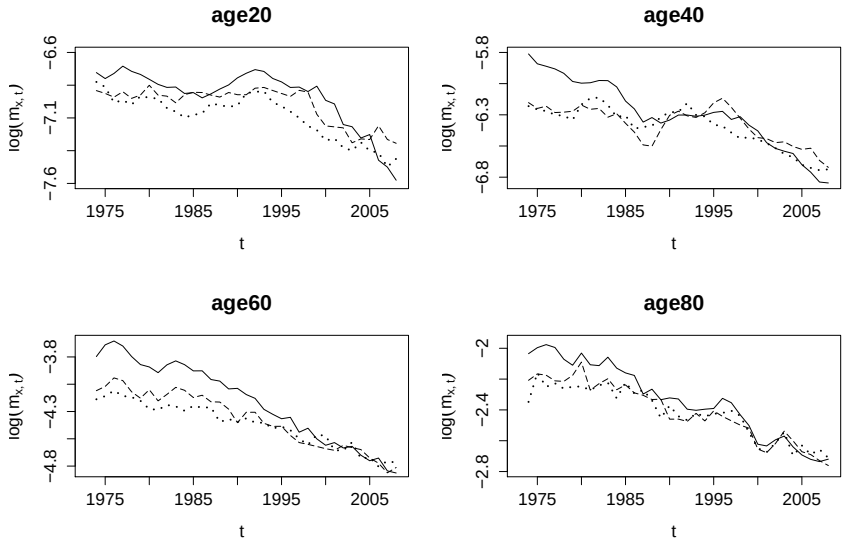

FIGURE 2. Evolution of $\log m^i_{x,t}$ for fixed ages and 3 regions: Lombardia (solid line), Lazio (dashed line) and Sicilia (dotted line).

Before estimating models P-division and M-division, we need to decompose the Italian regions into cluster of similar regions. We use a standard clustering algorithm where the dissimilarity index is given by the life expectancy at birth, computed in [20]. The final $I' = 5$ clusters are the following:

$$I_1 = \{1, 2, 3, 4, 5, 6\},\ I_2 = \{7, 8, 9\}$$

$$I_3 = \{10, 11, 12\},\ I_4 = \{13, 18\},\ I_5 = \{14, 15, 16, 17\}$$

3.2. **Estimation procedure.** The models are estimated using the first 25 years of observed data, from 1974 to 1998, while the remaining 10 years, from 1999 to 2008, are used in order to assess the quality of the out-of-sample forecast.

The parameters of the ten models are estimated by maximum likelihood using gradient based optimization techniques. The initial values of the parameters in the optimisation algorithm were set using the following criteria:

- for P-models, the starting values are obtained by assuming the normal distribution for the errors and using the procedure described in Lee and Carter [17];
- for M-models, the starting values are obtained using the first (or the first two) components of the singular value decomposition of the improvement rates matrix;
- in all the models containing common factors (that is (2), (4), (5) and (7), (9), (10)), the chosen starting values for the common parameters are those computed on the larger subpopulation.

Note that models where no common time factor is present, namely (1), (3) and (6), (8), the model likelihood factorizes into the subpopulation likelihoods and estimation can then be done separately for each subpopulation. In models (4) and (9), estimation can be done separately for each cluster of subpopulations.

We denote by $\hat{m}^i_{x,t}$, $\hat{z}^i_{x,t}$ and $\hat{\eta}^i_{x,t}$ the estimated values of $m^i_{x,t}$, $z^i_{x,t}$ and $\eta^i_{x,t}$.

3.3. **Forecast procedure.** Forecasting requires to model the time varying coefficients as a time series. According to the model considered, a different number of time series is involved in the application (see Table 1). When more than one period index appears in each subpopulation — as in (1), (2) and (6), (7) —, then the two (groups of) indices are modelled as independent time series, consistently with the orthogonality constrained imposed in that case, see Hunt and Blake [14].

As a general rule, if a VARIMA$(p, 1, q)$ time series is used to specify the time indices in P-models, then, following Haberman and Renshaw [12], a VARMA(p, q) should be used for M-models targeting improvement rates. Let then

$$y_t = \begin{vmatrix} y^1_t \\ \vdots \\ y^m_t \end{vmatrix},$$

be the vector of time indices that is to be specified, where the dimension m vary according to the model:

- P-double and M-double — two groups of eighteen time varying coefficients;
- P-common and M-common — one group of eighteen specific plus one common time varying coefficient;
- P-simple and M-simple — one group of eighteen time varying coefficients;
- P-division and M-division — one group of five cluster specific time varying coefficients;
- P-one and M-one — one time varying coefficient.

In the present application, for P-models a (multidimensional) random walk with drift, defined by

$$y_t = \Phi_0 + y_{t-1} + \xi_t,$$

where Φ_0 is the drift vector and ξ_t is a white noise process with $\xi_t \sim N(0, \Sigma^\xi)$, is used. For M-models, we use a simple regression to a constant

$$y_t = \Phi_0 + \xi_t,$$

where again ξ_t is a white noise process.

Once the time varying coefficients have been forecast, the values $\hat{m}^i_{x,t}$ for $t > t_n$ can be computed for P-models. As for the M-models, forecast central death rates

are computed using forecast improvement rates $\hat{z}^i_{x,t}$ by applying iteratively the formula

$$\hat{m}^i_{x,t_n+j} = \hat{m}^i_{x,t_n+j-1}\frac{2-\hat{z}^i_{x,t_n+j}}{2+\hat{z}^i_{x,t_n+j}}, \quad j=1,2,3,\dots. \tag{3.1}$$

The procedure starts for $j=1$ with $\hat{m}^i_{x,t_n}$ replaced by $\hat{m}^{i*}_{x,t_n}$ computed according to

$$\hat{m}^{i*}_{x,t_n} = \overline{m}^i_{x,t_n-1}\frac{2-\overline{z}^i_{x,t_n}}{2+\overline{z}^i_{x,t_n}},$$

where $\overline{m}^i_{x,t_n-1}$ and $\overline{z}^i_{x,t_n}$ are the simple averages of m^i_{x,t_n-2}, m^i_{x,t_n-1}, m^i_{x,t_n} and of z^i_{x,t_n-1}, z^i_{x,t_n}, respectively. This adjustment is adopted in order to lessen the dependence on the last observed value.

3.4. **Goodness of fit indices based on information criteria.** The two penalised log-likelihood indices most commonly used to assess the goodness of fit are (see Burnham and Anderson [5]) the Akaike information criterion (AIC) and the Bayes information criterion (BIC), defined respectively by

$$\mathrm{AIC} = 2d - 2\ell, \quad \mathrm{BIC} = d\log(g) - 2\ell,$$

with d the dimension of the parametrised prediction structure (see the second column in Table 1) and g is the sample size ($g = I\cdot k\cdot n$ for P-models and $g = I\cdot k\cdot(n-1)$ for M-models). By construction, the BIC puts a higher penalty on the number of parameters compared to the AIC. The maximized log-likelihood ℓ is given for P-models by

$$\ell = \sum_{i,\,x,\,t}\left(D^i_{x,t}\log\hat{m}^i_{x,t} - \mathrm{ETR}^i_{x,t}\hat{m}^i_{x,t}\right)$$

up to an additive constant independent of the chosen model, and by

$$\ell = -\frac{1}{2}\sum_{i,\,x,\,t}\left(\log(2\pi\hat{\sigma}^2_i) + \frac{(z^i_{x,t}-\hat{\eta}^i_{x,t})^2}{\hat{\sigma}^2_i}\right)$$

for M-models, where

$$\hat{\sigma_i}^2 = \sum_{x,t}\frac{(z^i_{x,t}-\hat{\eta}^i_{x,t})^2}{(n-1)k}$$

is the maximum likelihood estimate of σ^2_i.

When d is large relative to g, typically if $g/d < 40$, an adjusted version of the Akaike information criterion defined by

$$\mathrm{AIC}_c = \mathrm{AIC} + \frac{2d(d+1)}{g-d-1}$$

can be considered, see [5]. Clearly AIC_c converges to AIC as g gets large relative to d.

The best models are those corresponding to smaller values of the index. Note that the indices cannot be compared across the two main model structures, but only among the P-models and the M-models separately. For this reason, the results are presented separately in Table 2. To ease the comparison, differences between an index and the minimum value, denoted by Δ, are displayed.

	P-double	P-common	P-simple	P-division	P-one
d	4572	2955	2934	2622	2526
ℓ	-28332454	-28332976	-28334463	-28334716	-28335099
g	31500	31500	31500	31500	31500
AIC	56674052	56671861	56674794	56674676	56675251
Δ-AIC	2191	0	2933	2815	3389
rank-AIC	2	1	4	3	5
g/d	7	11	11	12	12
AIC_c	56675605	56672473	56675397	56675183	56675731
Δ-AIC_c	3132	0	2924	2679	3218
rank-AIC_c	4	1	3	2	5
BIC	56712264	56696558	56699316	56696725	56696538
Δ-BIC	15901	196	2953	228	0
rank-BIC	5	2	4	3	1

	M-double	M-common	M-simple	M-division	M-one
d	3312	1731	1674	1375	1283
ℓ	42056	39903	38760	37058	36617
g	30240	30240	30240	30240	30240
AIC	-77489	-76345	-74173	-71366	-70668
Δ-AIC	0	1144	3316	6122	6821
rank-AIC	1	2	3	4	5
g/d	9	17	18	22	24
AIC_c	-76674	-76134	-73977	-71235	-70554
Δ-AIC_c	0	539	2697	5438	6120
rank-AIC_c	1	2	3	4	5
BIC	-49943	-61948	-60250	-59931	-59997
Δ-BIC	12005	0	1698	2018	1951
rank-BIC	5	1	2	4	3

TABLE 2. Penalised log-likelihood indices for the ten models (when applicable, values are rounded to the nearest integer).

3.5. **Mean absolute percentage and residual analysis.** The in-sample or out-of-sample goodness of fit of a model can be measured with the Mean Absolute Percentage Error (MAPE), defined by

$$\text{MAPE}^i = \frac{1}{n_1 \cdot k} \sum_{x,t} \left| \frac{m^i_{x,t} - \hat{m}^i_{x,t}}{m^i_{x,t}} \right| \tag{3.2}$$

where n_1 is either the number of in-sample or out-of-sample years. In the current case, $k = 70$ and $n_1 = 25$ for P-models and $n_1 = 24$ for M-models (in-sample) or $n_1 = 10$ (out of sample). The MAPE for the in-sample fitting are summarized in Table 3 and in Table 4 for the out-of-sample forecast.

A graphical analysis of the residual plots, constructed by plotting the scaled residuals with respect to age, year and cohort, can be useful for investigating if the models are able to describe the general shape of the data and to capture any systematic patterns. The scaled residuals are defined by

$$r^i_{x,t} = \frac{D^i_{x,t} - \text{ETR}^i_{x,t}\hat{m}^i_{x,t}}{\sqrt{\text{ETR}^i_{x,t}\hat{m}^i_{x,t}}}$$

for P-models, and by

$$r^i_{x,t} = \frac{z^i_{x,t} - \hat{\eta}^i_{x,t}}{\sqrt{\hat{\sigma}^2_i}}$$

for M-models. It should be noted that the scaled residuals are computed with respect to the target quantity in the fitting procedure: the number of deaths $D^i_{x,t}$ for P-models and the mortality improvement rates for M-models. The residual plots for the ten models for the region Lombardia are presented in Figures 3 and 4.

Since the models are used to evaluate the general trend of mortality, an index which spans several years of forecast would be a more appropriate way for comparing

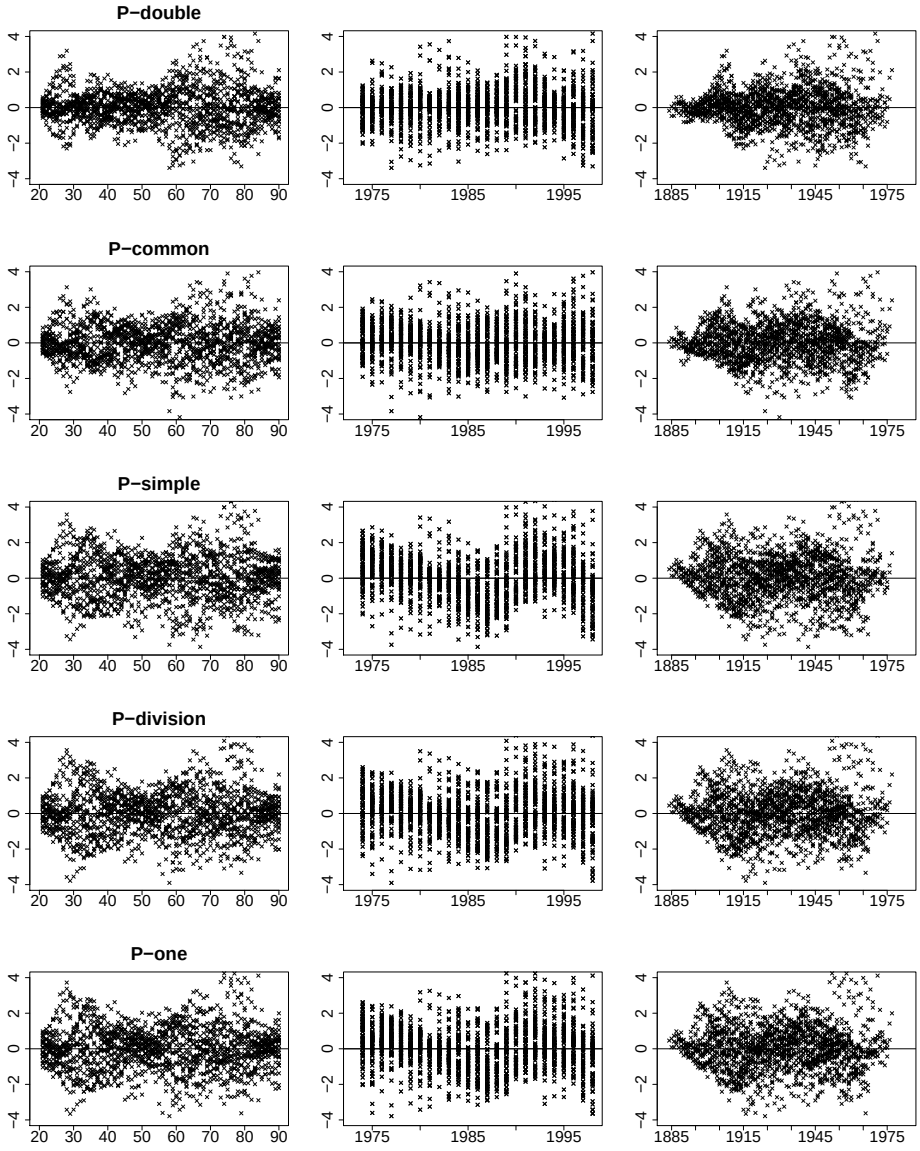

FIGURE 3. Age, year and cohort residual plots for P-models. Region: Lombardia.

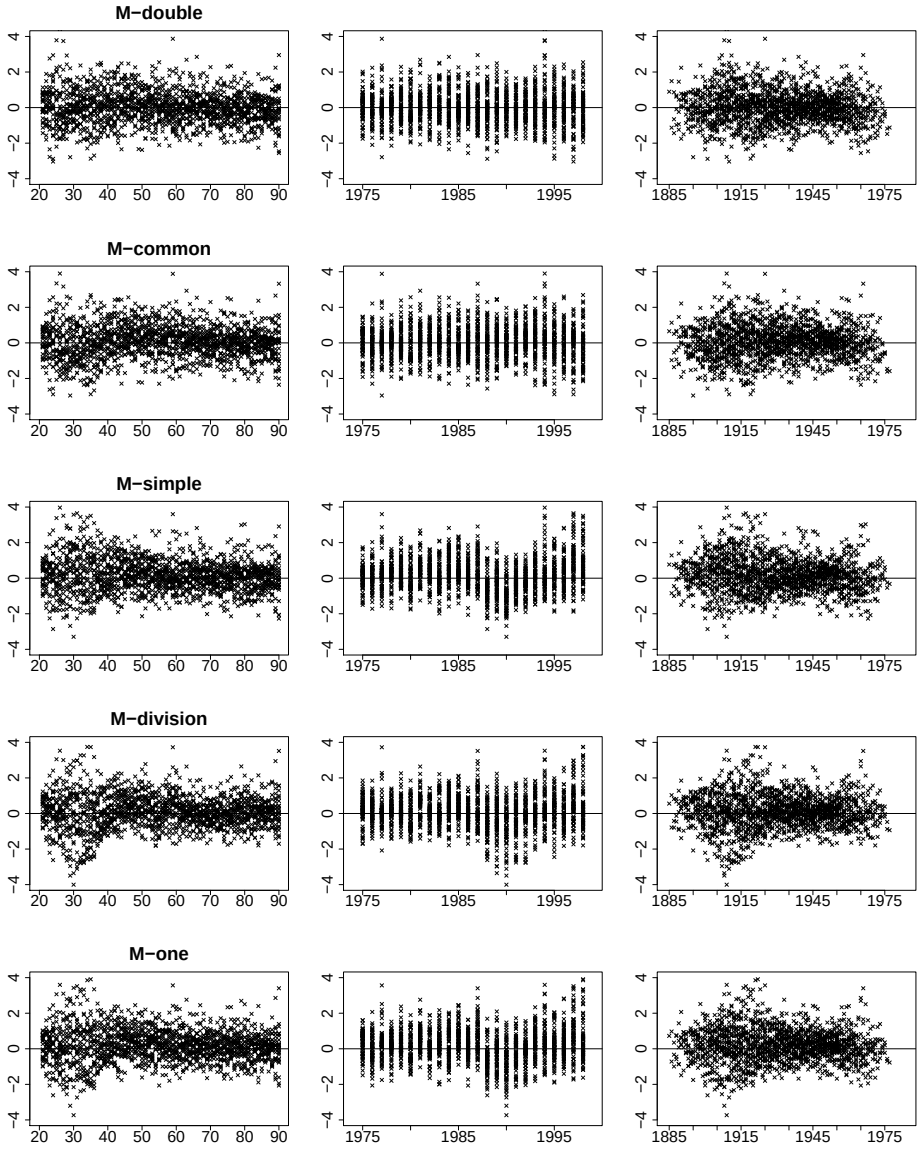

FIGURE 4. Age, year and cohort residual plots for M-models. Region: Lombardia.

REGION	P-double	P-common	P-simple	P-division	P-one	M-double	M-common	M-simple	M-division	M-one
1	4.04	4.51	5.21	5.3	5.3	9	9.15	10.41	11.38	13.54
2	4.18	4.84	5.94	5.9	5.9	9.49	12.66	13.11	17.35	14.04
3	5.84	6.56	6.74	6.94	6.99	16.59	18.25	18.18	14.22	20.4
4	3.78	4.63	5.2	5.36	5.37	10.15	10.33	12.07	13.02	15.21
5	5.32	6.16	6.05	6.5	6.58	12.19	15.62	13.36	18.19	16.45
6	4.86	5.59	6.4	6.83	6.75	14.18	10.51	15.24	28.11	14.24
7	4.65	5.12	6.32	6.29	6.31	11.29	11.2	19.17	19.82	13.91
8	4.4	4.71	5.21	5.3	5.3	12.48	9.79	15.64	17.32	11.96
9	6.94	7.5	7.69	7.76	8.2	16.47	11.62	17.43	17.05	15.63
10	4.98	5.67	6.12	6.19	6.28	11.93	9.33	13.13	10.41	12.28
11	4.02	4.31	4.99	5.22	5.3	7.39	8.04	9.67	12.83	10.72
12	5.38	5.95	5.8	5.97	5.96	15.61	12.57	16	11.78	13.26
13	3.45	4.05	4.08	4.1	4.41	9.67	8.74	10.26	10.33	10.81
14	4	4.75	5.42	5.46	5.51	9.35	8.15	13.07	12.61	10.56
15	7.42	8.51	8.01	8.28	8.92	13.9	12.38	14.49	15.68	17.79
16	4.43	5.57	5.51	5.66	5.75	13.19	12.1	13.66	12.84	14.26
17	3.76	4.45	4.93	5.02	5.09	8.06	11.31	10.28	10.77	12.73
18	4.88	5.47	6.35	7.05	6.94	10.81	9.85	12.71	14.2	14.33
mean	4.8	5.46	5.89	6.06	6.16	11.76	11.2	13.77	14.88	14.01
st. dev.	1.05	1.13	0.94	1.01	1.09	2.72	2.51	2.7	4.25	2.45

TABLE 3. MAPE of fitted with respect to observed data.

the predictive capacity of the models. We use the truncated expected residual lifetime computed along cohort trajectories. We truncate the expectation at 10 years in order not to introduce a mortality extrapolation at higher ages. The probabilities of death $q^i_{x,t}$ can be calculated by

$$q^i_{x,t} \approx 1 - \exp(-m^i_{x,t}).$$

The expected residual lifetime truncated after 10 years for population i, denoted by $e^i_{x:\overline{10|}}$, is computed by

$$e^i_{x:\overline{10|}} = \frac{\sum_{j=1}^{10} l^i_{x+j}(t_n+j)\left[1-\frac{1}{2}q^i_{x+j,t_n+j}\right]}{l^i_x(t_n)},$$

where $l^i_{x+j}(t_n+j)$, $j \geq 1$ is the number of survivors after j years in a fictitious population aged x in year t_n, with initial arbitrary size $l^i_x(t_n)$ (the last available year), see Haberman and Renshaw [12]. The population size $l^i_{x+j}(t_n+j)$ is recursively computed through

$$l^i_{x+j}(t_n+j) = \left[1-q^i_{x+j-1,t+j-1}\right] l^i_x(t_n+j-1),\ j = 1,2,3,\ldots.$$

This index is computed for all the regions for ages 60, 70 and 80. The results are summarised in Tables 5, 6 and 7, for ages 60, 70 and 80, respectively.

4. Discussion

In this section we compare the models employing the diagnostic tools, and indices introduced in Section 3. The values of these statistics are presented in Tables 2-7 and Figures 3-4. The analysis indicates that, when targeting death rates, the P-common model gives the best compromise between parsimony and goodness of fit (in and out of sample), followed by the P-division model. When modeling improvement rates, the corresponding structure M-common again provides a good balance between parsimony and in-sample goodness of fit. Alternatively, M-division and M-simple can be considered, in particular when out-of-sample performance is important.

Parsimony vs. goodness of fit. The penalised log-likelihood indices (see Table 2) give mixed indication of the relative model performance among the two families considered. First, the AIC ranks P- and M-models according to their complexity, so that models with more parameters perform better. The BIC reverts this ranking by awarding models with a simpler structure, with the exception of the common structures (models (2) and (6), which turn out to be the most robust with respect to both criteria.

The goodness of fit of the P- and M-common structures is confirmed by Figures 3-4, where plots of residuals vs. calendar year, age and cohort are depicted for one region (similar plots for other regions, available upon request, present essentially the same behaviour). These plots show that simpler structures, such as P/M-simple, P/M-division and P/M-one seems to evidence a pattern in residuals when plotted against calendar years.

In- and out-of-sample performance. Tables 3 and 4 contain respectively in- and out-of-sample performance for all the regions considered. The values of MAPE seem high, but it should be recalled that individual death rates are rather volatile. Note that models targeting improvement rates in-sample performance is uniformly worse than those targeting death rates. However, the difference between the two approaches is less striking when considering out-of-sample results. It should be mentioned that, in the M-approach, central death rates are obtained from improvement rates through iterative multiplication, see (3.1). In general, more complex

REGION	P-double	P-common	P-simple	P-division	P-one	M-double	M-common	M-simple	M-division	M-one
1	16.09	14.85	18.47	19.55	27.38	18.9	33.37	19.19	19.55	21.85
2	25.9	28.27	37.75	35.98	41.55	20.24	35.27	30.3	25.25	29.62
3	14.86	13.68	16.3	18.31	32.01	17.61	35.12	20.53	19.32	24.21
4	16.06	13.64	15.91	16.58	29.16	17.28	32.32	19.43	17.3	22.52
5	16.44	15.94	16.38	14.08	26.85	17.94	34.35	18.89	25.77	23.42
6	31.84	35.93	46.42	45.14	43.39	24.71	42.45	33.79	55.36	35.56
7	23.22	27.78	34.83	34.73	37.03	23.44	37.65	30.28	29.73	29.09
8	15.94	14.49	19.28	19.63	25.62	16.84	32.85	20.59	21.52	18.41
9	17.16	15.46	19.14	16.69	21.26	22.43	37.66	24.96	24.34	22.04
10	14.2	14.72	18.42	17.67	22.81	17.1	33.9	16.84	18.06	17.22
11	19.56	19.02	24.98	25.19	27.49	16.36	34.04	17.69	21.54	22.1
12	8.81	9.78	8.84	10.21	15.12	13.91	29.7	14.06	10.9	12.32
13	11.24	9.6	12.36	12.82	16.87	11.87	29.73	12.21	12.39	12.19
14	14.33	15.65	20.53	19.35	23.29	16.94	33.37	21.04	17.78	17.89
15	18.47	16.93	18.58	18.65	21.88	19.91	34.99	19.89	20.1	19.9
16	15.6	14.17	18.44	21.47	25.57	17.08	32.46	17.07	17.08	16.52
17	12.89	16.58	21.83	22.53	25.41	12.87	31.44	14.76	14.97	13.59
18	20.59	19.92	25.43	24.87	26.07	20.21	34.04	19.82	18.35	19.69
mean	17.4	17.58	21.88	21.86	27.15	18.09	34.15	20.63	21.63	21.01
st. dev.	5.26	6.53	9.02	8.56	7.27	3.3	2.91	5.63	9.36	5.94

TABLE 4. MAPE of forecast data with respect to observed data

REGION	Observed	P-double		P-common		P-simple		P-division		P-one		M-double		M-common		M-simple		M-division		M-one	
1	9.32	9.31	0.15	9.29	0.31	9.31	0.15	9.3	0.28	9.24	0.93	9.33	0.07	9.35	0.28	9.33	0.08	9.33	0.06	9.31	0.14
2	9.26	9.24	0.21	9.27	0.09	9.23	0.35	9.27	0.02	9.19	0.81	9.3	0.44	9.33	0.68	9.27	0.08	9.31	0.45	9.28	0.2
3	9.32	9.31	0.16	9.31	0.16	9.32	0.04	9.31	0.21	9.24	0.95	9.34	0.21	9.35	0.23	9.33	0.09	9.33	0.05	9.31	0.18
4	9.29	9.27	0.21	9.27	0.28	9.27	0.29	9.27	0.28	9.2	0.98	9.33	0.38	9.36	0.68	9.32	0.33	9.35	0.55	9.31	0.16
5	9.25	9.23	0.17	9.2	0.54	9.2	0.57	9.22	0.28	9.15	1.06	9.29	0.48	9.33	0.83	9.31	0.61	9.27	0.22	9.28	0.37
6	9.32	9.31	0.14	9.32	0	9.32	0.04	9.3	0.23	9.24	0.84	9.3	0.21	9.33	0.04	9.29	0.31	9.31	0.13	9.29	0.36
7	9.36	9.35	0.14	9.34	0.19	9.34	0.18	9.34	0.17	9.3	0.65	9.4	0.45	9.41	0.5	9.36	0	9.36	0.01	9.4	0.38
8	9.37	9.35	0.16	9.36	0.08	9.35	0.15	9.35	0.19	9.31	0.6	9.39	0.26	9.42	0.62	9.37	0.07	9.37	0.08	9.4	0.39
9	9.37	9.36	0.17	9.35	0.2	9.35	0.24	9.37	0.01	9.34	0.4	9.38	0.1	9.4	0.33	9.37	0	9.37	0.01	9.41	0.42
10	9.42	9.41	0.13	9.4	0.23	9.4	0.22	9.42	0.09	9.38	0.42	9.41	0.19	9.43	0.12	9.41	0.13	9.42	0.02	9.42	0
11	9.32	9.3	0.16	9.31	0.13	9.3	0.18	9.29	0.28	9.26	0.61	9.35	0.4	9.36	0.49	9.35	0.39	9.33	0.17	9.33	0.16
12	9.35	9.34	0.08	9.34	0.07	9.34	0.11	9.32	0.28	9.3	0.54	9.32	0.35	9.36	0.06	9.32	0.33	9.34	0.09	9.34	0.14
13	9.2	9.18	0.15	9.18	0.22	9.18	0.18	9.18	0.22	9.15	0.54	9.19	0.02	9.23	0.34	9.19	0.03	9.2	0.02	9.21	0.17
14	9.37	9.35	0.15	9.34	0.26	9.34	0.29	9.35	0.17	9.32	0.5	9.38	0.09	9.4	0.33	9.34	0.26	9.36	0.14	9.38	0.08
15	9.33	9.32	0.1	9.33	0.05	9.33	0.04	9.33	0.1	9.31	0.14	9.29	0.42	9.32	0.05	9.29	0.35	9.29	0.36	9.31	0.2
16	9.35	9.33	0.16	9.34	0.14	9.34	0.09	9.32	0.3	9.3	0.56	9.33	0.15	9.38	0.31	9.33	0.16	9.35	0.05	9.35	0
17	9.34	9.33	0.13	9.31	0.34	9.3	0.38	9.3	0.43	9.27	0.69	9.35	0.1	9.37	0.32	9.34	0.05	9.33	0.05	9.36	0.28
18	9.29	9.27	0.22	9.27	0.21	9.27	0.26	9.28	0.15	9.26	0.33	9.29	0.01	9.32	0.3	9.29	0.02	9.29	0.02	9.31	0.16
mean			0.15		0.19		0.21		0.21		0.64		0.24		0.36		0.18		0.14		0.21
st.dev.			0.04		0.13		0.14		0.11		0.25		0.16		0.23		0.17		0.16		0.13

TABLE 5. Expected residual lifetime at age 60 (10 years truncated). Each couple of columns refers to a model: the left column contains the estimated index, the right one the percentage error with respect to the observed value.

REGION	Observed	P-double		P-common		P-simple		P-division		P-one		M-double		M-common		M-simple		M-division		M-one	
1	8.36	8.35	0.11	8.34	0.26	8.35	0.15	8.32	0.46	8.21	1.84	8.29	0.84	8.3	0.77	8.3	0.77	8.3	0.74	8.27	1.1
2	8.23	8.21	0.21	8.24	0.09	8.19	0.51	8.25	0.19	8.11	1.48	8.28	0.62	8.31	0.99	8.21	0.27	8.26	0.31	8.23	0.07
3	8.42	8.41	0.03	8.42	0.09	8.43	0.21	8.41	0.04	8.27	1.68	8.38	0.46	8.41	0.11	8.39	0.27	8.42	0.01	8.36	0.61
4	8.34	8.33	0.11	8.3	0.54	8.32	0.21	8.33	0.18	8.2	1.72	8.31	0.33	8.35	0.12	8.32	0.26	8.32	0.27	8.28	0.78
5	8.26	8.25	0.17	8.18	0.99	8.18	0.93	8.23	0.38	8.1	1.88	8.24	0.19	8.31	0.65	8.27	0.17	8.23	0.35	8.23	0.38
6	8.39	8.38	0.11	8.39	0.05	8.38	0.07	8.35	0.47	8.23	1.9	8.28	1.33	8.34	0.59	8.28	1.31	8.34	0.64	8.3	1.06
7	8.48	8.48	0.07	8.46	0.32	8.47	0.19	8.46	0.24	8.36	1.44	8.48	0.1	8.43	0.59	8.35	1.62	8.34	1.63	8.44	0.57
8	8.42	8.41	0.17	8.4	0.2	8.42	0.01	8.42	0.06	8.33	1.09	8.31	1.31	8.4	0.21	8.31	1.33	8.31	1.35	8.39	0.37
9	8.46	8.45	0.09	8.43	0.38	8.43	0.32	8.49	0.29	8.39	0.79	8.32	1.65	8.39	0.82	8.29	2	8.3	1.95	8.38	1
10	8.55	8.53	0.15	8.51	0.37	8.51	0.41	8.53	0.17	8.46	1	8.45	1.18	8.51	0.45	8.45	1.11	8.49	0.62	8.49	0.6
11	8.28	8.26	0.23	8.25	0.38	8.25	0.36	8.24	0.49	8.18	1.27	8.22	0.78	8.24	0.58	8.22	0.76	8.16	1.49	8.21	0.95
12	8.51	8.52	0.03	8.51	0.04	8.52	0.03	8.46	0.62	8.4	1.37	8.43	0.97	8.5	0.21	8.43	1	8.49	0.28	8.46	0.62
13	8.1	8.08	0.23	8.07	0.4	8.08	0.17	8.08	0.28	8.02	1.01	8.09	0.1	8.15	0.63	8.09	0.07	8.1	0.06	8.13	0.42
14	8.4	8.39	0.2	8.39	0.21	8.38	0.3	8.4	0.05	8.33	0.91	8.35	0.59	8.42	0.17	8.32	1	8.38	0.24	8.38	0.32
15	8.38	8.36	0.2	8.44	0.71	8.45	0.83	8.46	0.91	8.4	0.24	8.42	0.49	8.48	1.16	8.42	0.44	8.42	0.46	8.49	1.32
16	8.41	8.39	0.23	8.39	0.15	8.4	0.09	8.36	0.5	8.31	1.15	8.33	0.88	8.43	0.24	8.33	0.91	8.38	0.35	8.39	0.23
17	8.32	8.3	0.21	8.29	0.31	8.29	0.31	8.29	0.4	8.22	1.16	8.26	0.68	8.32	0.08	8.27	0.55	8.27	0.57	8.32	0.04
18	8.31	8.29	0.25	8.29	0.25	8.28	0.3	8.31	0.05	8.27	0.5	8.36	0.59	8.41	1.16	8.36	0.62	8.36	0.64	8.4	1.06
mean			0.16		0.32		0.3		0.32		1.25		0.73		0.53		0.8		0.66		0.64
st.dev.			0.07		0.24		0.25		0.23		0.47		0.44		0.36		0.53		0.56		0.38

TABLE 6. Expected residual lifetime at age 70 (10 years truncated). Each couple of columns refers to a model: the left column contains the estimated index, the right one the percentage error with respect to the observed value.

REGION	Observed	P-double		P-common		P-simple		P-division		P-one		M-double		M-common		M-simple		M-division		M-one	
1	6.36	6.35	0.16	6.31	0.81	6.34	0.27	6.3	0.93	6.14	3.45	6.18	2.77	6.12	3.7	6.21	2.31	6.23	2.05	6.19	2.63
2	6.29	6.28	0.11	6.29	0.12	6.2	1.31	6.3	0.22	6.11	2.88	6.17	1.9	6.21	1.2	6.06	3.64	6.06	3.54	6.1	3.02
3	6.51	6.49	0.33	6.47	0.55	6.51	0.03	6.48	0.48	6.28	3.51	6.21	4.64	6.38	1.92	6.27	3.73	6.4	1.66	6.27	3.67
4	6.45	6.45	0.04	6.44	0.05	6.45	0.01	6.44	0.19	6.25	3.13	6.24	3.29	6.28	2.63	6.31	2.21	6.3	2.34	6.25	3.12
5	6.27	6.24	0.45	6.19	1.29	6.22	0.79	6.26	0.12	6.09	2.89	6.19	1.25	6.26	0.1	6.23	0.7	6.18	1.49	6.12	2.44
6	6.4	6.38	0.16	6.4	0.13	6.39	0.04	6.35	0.78	6.19	3.25	6.33	0.95	6.33	1.07	6.32	1.11	6.44	0.74	6.35	0.65
7	6.51	6.5	0.24	6.52	0.11	6.55	0.56	6.54	0.47	6.39	1.84	6.35	2.44	6.31	3.09	6.19	4.88	6.19	4.93	6.35	2.47
8	6.41	6.38	0.57	6.39	0.3	6.4	0.28	6.39	0.38	6.27	2.18	6.24	2.73	6.37	0.66	6.2	3.27	6.2	3.26	6.3	1.77
9	6.35	6.3	0.79	6.26	1.51	6.26	1.55	6.32	0.46	6.21	2.23	6.25	1.62	6.33	0.4	6.21	2.23	6.21	2.23	6.32	0.53
10	6.47	6.44	0.5	6.45	0.26	6.46	0.13	6.49	0.36	6.39	1.16	6.28	2.88	6.35	1.75	6.3	2.63	6.35	1.8	6.38	1.35
11	6.26	6.23	0.37	6.25	0.14	6.25	0.16	6.23	0.47	6.13	1.93	6.21	0.8	6.23	0.37	6.21	0.66	6.19	1.07	6.2	0.83
12	6.54	6.51	0.38	6.47	1.03	6.49	0.71	6.42	1.88	6.34	3.1	6.33	3.22	6.46	1.23	6.35	2.96	6.46	1.26	6.46	1.14
13	6.14	6.11	0.35	6.07	1.07	6.12	0.34	6.1	0.6	6.01	2.11	5.99	2.45	6.04	1.49	5.99	2.37	6.01	2.02	6.06	1.23
14	6.31	6.28	0.51	6.29	0.3	6.3	0.27	6.33	0.29	6.23	1.28	6.17	2.23	6.23	1.26	6.11	3.24	6.09	3.49	6.24	1.16
15	6.49	6.47	0.32	6.29	3.01	6.31	2.78	6.32	2.57	6.26	3.53	6.32	2.54	6.42	1.05	6.32	2.62	6.37	1.79	6.28	3.23
16	6.36	6.31	0.68	6.32	0.53	6.34	0.24	6.29	1.04	6.23	2.06	6.22	2.13	6.36	0.02	6.22	2.13	6.29	1	6.32	0.63
17	6.12	6.07	0.74	6.14	0.36	6.13	0.26	6.11	0.04	6.04	1.33	6.04	1.25	6.12	0.03	6.07	0.74	6.04	1.18	6.14	0.46
18	6.47	6.46	0.2	6.42	0.79	6.42	0.85	6.47	0.01	6.41	0.98	6.43	0.66	6.52	0.66	6.43	0.68	6.49	0.24	6.52	0.79
mean			0.38		0.69		0.59		0.63		2.38		2.21		1.26		2.34		2.01		1.73
st.dev.			0.22		0.73		0.7		0.66		0.86		1.02		1.05		1.21		1.17		1.08

TABLE 7. Expected residual lifetime at age 80 (10 years truncated). Each couple of columns refers to a model: the left column contains the estimated index, the right one the percentage error with respect to the observed value.

models tend to have a better in-sample performance while the relative difference is again diminished for out-of-sample. An exception is M-common, whose out-of-sample behaviour is notably worse than the other models of the same family.

The predictive capacity of the models can be analysed also through the 10 years truncated expected residual lifetime at ages 60, 70 and 80 and the corresponding MAPE, contained in Tables 5-7. As these expectations depend on a set of death probabilities, a smoothing effect of the noise embedded in death rates is apparent as can be seen in the performance which is now satisfying in absolute terms. The decreasing performance pattern observed in the three tables is due to the increasing volatility of death rates with age. Among P-models, the most complex structure, P-double, clearly dominates the other models, with P-one predictive capacity being seriously undermined. The intermediate models, P-common, -simple and -division, present a similar performance. As for M-models, results are mixed, with predictive capacity not always related to model complexity. With the exception of age 60, M-common seems to be the most performing model.

References

[1] S. D. Baxter. Should projections of mortality improvements be subject to a minimum value? *British Actuarial Journal*, 13:375–464, 2007.

[2] H. Booth, J. Maindonald, and L. Smith. Applying Lee-Carter under conditions of variable mortality decline. *Population Studies*, 56(3):325–336, 2002.

[3] H. Booth, R. J. Hyndman, L. Tickle, and P. de Jong. Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions models. *Demographic Research*, 15(9):289–310, 2006.

[4] N. Brouhns, M. Denuit, and J.K. Vermunt. A Poisson log-bilinear approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31(3):373–393, 2002.

[5] P. Burnham and R. Anderson. Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2):261–304, 2004.

[6] A. Cairns, D. Blake, K. Dowd, D. Coughlan, D. Epstein, A. Ong, and I. Balevich. A quantitative comparison of stochastic mortality models using data from England & Wales and the United States. *North American Actuarial Journal*, 13(1):1–35, 2009.

[7] G. Coughlan, M. Khalaf-Allah, Y. Ye, S. Kumar, A. Cairns, D. Blake, and K. Dowd. Longevity hedging 101: A framework for longevity basis risk analysis and hedge effectiveness. *North American Actuarial Journal*, 15(2):150–176, 2011.

[8] P. De Jong and L. Tickle. Extending Lee–Carter mortality forecasting. *Mathematical Population Studies*, 13(1):1–18, 2006.

[9] A. Delwarde, M. Denuit, M. Guillen, and A. Vidiella. Application of the Poisson log-bilinear projection model to the G5 mortality experience. *Belgian Actuarial Bulletin*, 6(1):54–68, 2006.

[10] K. Dowd, A. Cairns, D. Blake, G. Coughlan, and M. Khalaf-Allah. A gravity model of mortality rates for two related populations. *North American Actuarial Journal*, 15(2):334–356, 2011.

[11] S. Haberman and A. Renshaw. A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics*, 48(1):35–55, 2011.

[12] S. Haberman and A. Renshaw. Parametric mortality improvement rate modelling and projecting. *Insurance: Mathematics and Economics*, 50(3):309–333, 2012.

[13] S. Haberman and A. Villegas. On the modelling and forecasting of socio-economic mortality differentials: an application to deprivation and mortality in England. *North American Actuarial Journal*, 18(1):168–193, 2014.
[14] A. Hunt and D. Blake. Identifiability in age/period mortality models. 2014.
[15] R. J. Hyndman and Md. S. Ullah. Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, 51(10):4942–4956, 2007.
[16] S. F. Jarner and E. M. Kryger. Modelling adult mortality in small populations: The SAINT model. *ASTIN Bulletin*, 41(2):377–418, 2011.
[17] R. D. Lee and L.R. Carter. Modelling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87(14):659–675, 1992.
[18] J.S. Li and M.R. Hardy. Measuring basis risk in longevity hedges. *North American Actuarial Journal*, 15(2):177–200, 2011.
[19] N. Li and R. D. Lee. Coherent mortality forecasts for a group of populations: an extension of the Le–Carter method. *Demography*, 42(3):575–594, 2005.
[20] G. Minelli, V. Manno, S. M. D'Ottavi, M. Masocco, G. Rago, M. Vichi, L. Frova, S. Marchetti, M. Demaria, and S. Conti. La mortalità in Italia nell'anno 2009. *Rapporti ISTISAN*, (12/15), 2012.
[21] D. Mitchell, P. Brockett, R. Mendoza-Arriaga, and K. Muthuraman. Modeling and forecasting mortality rates. *Insurance: Mathematics and Economics*, 52 (2):275–285, 2013.
[22] E. Pitacco, M. Denuit, S. Haberman, and A. Olivieri. *Modelling Longevity Dynamics for Pensions and Annnuity Business*. Oxford University Press, 2009.
[23] A. Renshaw and S. Haberman. Lee-Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, 33(2):255–272, 2003.
[24] S. J. Richards, J. G. Kirkby, and I. D. Currie. The importance of year of birth in two-dimensional mortality data. *British Actuarial Journal*, 12(1):5–61, 2005.
[25] M. Russolillo, G. Giordano, and S. Haberman. Extending the Lee–Carter model: a three-way decomposition. *Scandinavian Actuarial Journal*, 2011(2): 96–117, 2011.