The maximum entropy mortality model: forecasting mortality using statistical moments

Marius D. Pascariu, Adam Lenart & Vladimir Canudas-Romo

To cite this article: Marius D. Pascariu, Adam Lenart & Vladimir Canudas-Romo (2019) The maximum entropy mortality model: forecasting mortality using statistical moments, Scandinavian Actuarial Journal, 2019:8, 661-685, DOI: 10.1080/03461238.2019.1596974

To link to this article: https://doi.org/10.1080/03461238.2019.1596974

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 29 Mar 2019.

Submit your article to this journal

Article views: 687

View related articles

View Crossmark data
The maximum entropy mortality model: forecasting mortality using statistical moments

Marius D. Pascariu a,b, Adam Lenart c and Vladimir Canudas-Romo d

aInterdisciplinary Center on Population Dynamics, Faculty of Business and Social Sciences, University of Southern Denmark, Odense, Denmark; bLongevity & Morbidity R&D Center, SCOR Global Life SE, Paris, France; cInstitute of Public Health, University of Southern Denmark, Odense, Denmark; dSchool of Demography, The Australian National University, Canberra, Australia

ABSTRACT

The age-at-death distribution is a representation of the mortality experience in a population. Although it proves to be highly informative, it is often neglected when it comes to the practice of past or future mortality assessment. We propose an innovative method to mortality modeling and forecasting by making use of the location and shape measures of a density function, i.e. statistical moments. Time series methods for extrapolating a limited number of moments are used and then the reconstruction of the future age-at-death distribution is performed. The predictive power of the method seems to be net superior when compared to the results obtained using classical approaches to extrapolating age-specific-death rates, and the accuracy of the point forecast (MASE) is improved on average by 33% respective to the state-of-the-art, the Lee–Carter model. The method is tested using data from the Human Mortality Database and implemented in a publicly available R package.

1. Introduction

The mortality experience of a population is well described by its hazard rates, and the age-pattern that has been well exploited in numerous methods of mortality modeling (Gompertz 1825; Makeham 1867; Siler 1983; Heligman and Pollard 1980) and forecasting (Lee and Carter 1992; Li and Lee 2005; Haberman et al. 2014). The main reasons why hazard rates have been used predominantly in modeling and forecasting is that they readily represent the change in the risk of death over age and time. In addition, the five components of the pattern of human mortality (infant, child, youth, adult and old age mortality) can be identified clearly in a graphical representation of the hazard curve; and a variety of time series models can be employed to extrapolate the identified trends over time.

Surprisingly, few mortality methods acknowledge that the probability density function of the distribution of deaths can be equally informative when compared to the hazard indices: and more than that it can give an immediate indication on key measure of longevity like how long a population lives on average and the degree of variability of ages at death. By employing statistical knowledge about the shape of the distribution, and how the level of mortality at a certain age is fully dependent on the levels of mortality at all the other ages, brings enormous advantages. When hazard rates are investigated, no conclusion can be made about mean age-at-death, or the inequality experienced by the population when it comes to death, or what is the prevalence of extraordinary long life spans (the outliers in old

CONTACT  Marius D. Pascariu mpascariu@health.sdu.dk  Interdisciplinary Center on Population Dynamics, Faculty of Business and Social Sciences, University of Southern Denmark, J.B. Winslows Vej 9B, 5000 Odense, Denmark

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
age mortality) or even summary statistics related to the mortality experience. For example, Figure 1 illustrates the transition in the age-at-death distribution for men living in England and Wales. We learn that in 1960 the population experienced a pronounced infant mortality level up to the age of 5, and the fact that it takes about 53 years in order for the first 10% of men to die. In 1960, only 10% of the male population had the chance of living beyond the age of 85. Analyzing the same distribution in 2016, we can see that infant mortality dropped to almost insignificant levels (an inspection on the logarithmic scale would still show differences over ages), and that it takes about 63 years to eliminate the first 10% of the male population through death. We also learn that more that 50% of men are surviving to age 85, and more than 10% of the population will surpass the age of 93.

The first attempt to describe the mortality pattern by analyzing the age-at-distribution was made by Pearson (1897), building on Lexis’ work (1879), where different functions were being used to capture the components of the pattern of human mortality. Pearson used a skewed function, arguing that the skewness of old age mortality depends on the incidence of premature mortality. More recently, (Dellaportas et al. 2001) uses death counts to fit the Heligman–Pollard model with Bayesian methods, and Mazzuco et al. (2018) models mortality by fitting a half-normal and a skew-bimodal-normal distribution to the observed empirical age-at-death density function.

Despite being well suited to portray the dynamics of mortality patterns and to study longevity and lifespan variability, age-at-death distributions have generally been neglected in forecasting practice. The life table distribution of deaths, or \( f(x) \), is constrained ensuring that its elements add to the radix of the population, usually \( l_0 = 1 \), thus having \( \sum x f(x) = 1 \). Time series extrapolations of its trends are likely to violate this assumption making burdensome the use of the same extrapolative methods as in the case of hazard rates. Notable attempts at forecasting the age-at-death distribution were made by Oeppen (2008) and Bergeron-Boucher et al. (2017) by adapting the Lee–Carter model to the Compositional Data Analysis framework (Aitchison 1982). While this approach solves the problem of respecting the unit sum constraint, it also requires changing the coordinate system from a Euclidean space to an Aitchison simplex (1986) which might hinder the interpretation of the results. Another novel idea is introduced by Basellini and Camarda (2019) by modeling the shifting

---

1 The commonly accepted life table notation for the age-at-death distribution is \( d(x) \). However, we will use the \( f(x) \) notation in order to maintain consistency with the notation used in statistics for densities.
and compression dynamics of the adult mortality distribution around the modal age at death with a 3-parameter function, parameters that can be extrapolated using standard time series models.

Inspired by the idea of forecasting mortality, given the information offered by the age-at-death distribution, we propose a novel approach to forecasting age-specific mortality levels by making use of statistical moments, i.e. the shape measures of a density. By employing the knowledge about the shape of the distribution mentioned earlier, and how the level of mortality at a certain age is fully dependent on the levels of mortality at all the other ages, brings an enormous advantage to the extrapolation methods based on death frequencies over the methods based on hazard or mortality rates extrapolation. One statistical moment describes one characteristic of the distribution it belongs to. For example, in the case of the distribution of deaths of a population the first moments correspond to: (i) the mean age at death or life expectancy; (ii) the variance which offers information about the inequality of the age at death; (iii) the skewness which states how concentrated the mortality is around young or old ages, and (iv) the kurtosis which indicates the weight of the tails or the presence of outliers (extreme old age) in the distribution. Beyond moments higher than the fourth the interpretation is limited, however these are relevant in the case of more complex distribution (multi-modal densities) which helps in fine-tuning the observed irregularities. Thus, for a distribution, the collection of all the moments uniquely determines its density function. In order to gain a perfect understanding of the underlying density function, one needs to have information about all the moments up to infinity. However a limited number of moments like mean, variance, skewness and kurtosis can already offer a good approximation of the shape of the probability density function of the underlying distribution of deaths, and therefore a good understanding of the levels of mortality experienced by a population at different ages.

The method proposed here considers the finite moment problem where a positive density, \( f(x) \), is sought from knowledge of a limited number of its power moments. We assess the evolution of several observed moments of the age-at-death distribution in order to forecast them by employing multivariate time series models; and we reconstruct the forecast distribution using the maximum entropy approach (Mead and Papanicolaou 1984) that relies on the average rate at which information is produced by a stochastic source of data or density function, i.e. information entropy. Reconstructing the density function from a set of predicted moments has the advantage of allowing accelerating/decelerating rates of mortality improvement over age and time. It also eliminates the necessity of altering the coordinate system as in Oeppen (2008). We will refer to this method as the maximum entropy mortality model (MEM).

The purpose of this study is to:

- demonstrate that extrapolation methods based on death frequencies can be advantageous to methods based on mortality rates. We are also interested in showing that accurate forecasts of age-specific mortality levels can be obtained using statistical moments and the information provided by the age-at-death distribution;
- compare the MEM against other well established mortality models and determine its newly added value;
- validate the MEM predictions against a benchmark, that is, a simplistic trend extrapolation of the age-specific death rates (naïve model) in order to justify the increase in complexity of the proposed method. This objective is justified and inspired by Bohk-Ewald et al. (2018), a study where 20 major fertility forecasting methods are evaluated. The main findings show that across multiple measures of fertility forecast accuracy only four methods consistently outperform the naïve model.

2. Methods

In this section, we introduce our method of modeling and forecasting mortality and the main concepts required to understand its estimation procedure.
2.1. Statistical moments

The statistical moments are defined as the expected value of the $n$-th power of a random variable $x$. The $n$-th moment, $\mu_n$, for a continuous density function, $f(x)$, about a value $c$ can be defined as:

$$\mu_n = \int_c^\omega (x - c)^n f(x) dx,$$

where $\omega$ is the range of the distribution. If variable $c$ denotes the mean, it is said that $\mu_n$ is the $n$-th moment about the mean or the $n$-th central moment. The moments can be computed about zero as well, in which case the moment is called a raw or crude moment. The normalized moment of a probability distribution is a central moment that is standardized. The normalization is typically a division by an expression of the standard deviation, $\sigma$, which renders the moment scale invariant.

The normalized moment of degree $n$ of a central moment is

$$\tilde{\mu}_n = \mu_n / \sigma^n.$$

Both the raw and the normalized moments are used in this study. While the normalized moments are a good choice to serve as indices in time series extrapolation (as we will see in Section 2.4), the raw moments are more efficient in the estimation of the underlying density functions (Section 2.3). If one type of moments is known, all the others can be derived from these without losing their properties.

2.2. Information entropy

The concept of information entropy was introduced by Shannon (1948) in an effort to mathematically formalize the process of communication in computer science (between two devices). The word information, here, is used in a special sense and must not be confused with meaning. The information is seen as a measure of the numbers of the possible outcomes generated by a probabilistic process and it is defined as the logarithm of this value. Warren (1953) reveals that the quantity which uniquely meets the natural requirements that one sets up for information turns out to be exactly that which is known in thermodynamics as entropy and which is a measure of the degree of randomness.

To explore the numerical values of the entropy measure $H$, let us assume that the individuals in a population can die in one of the following two states: childhood (C) and adulthood (A). The associate probabilities would be $f(C)$ for the first state and $f(A) = 1 - f(C)$ for the second. It follows that:

$$H = - \left[ f(C) \log f(C) + f(A) \log f(A) \right] \quad \text{or} \quad H = - \sum_{i=C}^A f(i) \log f(i).$$

Since the logarithm of a number smaller that 1 is a negative value, the minus sign in the equation is added for convenience only, so that $H$ is always positive. It turns out that the information entropy has its largest value, when the events of dying in childhood and adulthood are equally probable; that is when $f(C) = f(A) = 0.5$. Just as soon as one outcome becomes more probable than the other (e.g. adult mortality greater than childhood mortality) the value of $H$ decreases. When one outcome is very probable ($f(A)$ almost one and $f(C)$ almost zero, say) the value of $H$ is approaching zero, creating a situation in which the stochastic process of dying in different states becomes less random and the outcome almost certain. Therefore, information entropy is also a measure that can show the level of inequality experienced by individuals represented in a distribution such as age-at-death. A small value of the entropy indicates that everyone dies around the same ages, i.e. the population is characterized by a small degree of inequality in terms of age-at-death. If $H$ is large one can say that the inequality is pronounced.
To generalize, the entropy of a random variable $x$ with probability distribution function $f(x)$ is the negative logarithm of the density function for the value, and can be written as:

$$H = -\int f(x) \log_b f(x) \, dx,$$

and

$$H = E[I(x)] = E[-\log_b f(x)],$$

where $E$ is the expected value operator, $I$ is the information content of $x$ and $b$ is the base of the logarithms used.

### 2.3. The finite moment problem

The problem of reconstructing a distribution from a given number of moments is not straightforward. It is known in the mathematical literature as the finite moment problem. The method has been extensively studied from a theoretical perspective and has practical applications in thermodynamics and quantum-physics. It can be regarded as a finite dimensional version of the Hausdorff moment problem (Hausdorff 1921; Shohat and Tamarkin 1943). Various methods for solving this problem have been proposed in the last decades, by making use of orthogonal polynomials (Chihara 1978), splines (John et al. 2007), or other numerical strategies (Frontini et al. 1990). All the procedures aim at constructing specific sequences of functions $f_N(x)$ which eventually converge to the true distribution $f(x)$ as the number of moments $N$, used in estimation, approaches infinity

$$\hat{\mu}_n = \int_{a}^{\omega} x^n f_N(x) \, dx, \quad n = 0, 1, 2 \ldots, N. \tag{6}$$

Equation (6) should be seen as a system of $N+1$ equations, where the moments $\hat{\mu}_0, \ldots, \hat{\mu}_N$ come from the $f_N(x)$ density.

Taking advantage of the regularity of human mortality the reconstruction of a density function can be realized using a small number of moments, usually 3–6. And, a good fit of the true density is achieved by imposing a prior restriction of the class on functions where the solution is sought.

Here, we follow the maximum entropy reconstruction (MaxEnt) and the algorithm developed by Mead and Papanicolaou (1984) as a definite procedure for the construction of a sequence of approximations to the true density. This method is based on the information entropy given by the density function. As a strategy for finding the local maxima of the entropy functional $\mathcal{L} = \mathcal{L}(f)$, we employ the method of Lagrange multipliers, $\lambda_n$ for the $n$-th moment:

$$\mathcal{L} = H + \sum_{n=0}^{N} \lambda_n [\hat{\mu}_n - \mu_n]. \tag{7}$$

The entropy, as defined in Equation 4, is maximized under the condition that the first $N+1$ moments, $\hat{\mu}_n$, are equal to the true moments $\mu_n$, where $n$ takes values between 0 and $N$. Functional variation of $\mathcal{L}$ with respect to the unknown density $f(x)$ yields

$$\frac{\delta \mathcal{L}}{\delta f(x)} = 0 \implies f = f_N(x) = \exp \left( -\lambda_0 - \sum_{n=1}^{N} \lambda_n x^n \right), \tag{8}$$

and the $n-$th raw moment

$$\mu_n = \int_{a}^{\omega} x^n \exp \left( -\lambda_0 - \sum_{n=1}^{N} \lambda_n x^n \right) \, dx. \tag{9}$$
Considering the availability of the first $N+1$ moments, the Equations (8) and (9) (that are closely related to Equation 6) should be viewed as a non-linear system of $N+1$ equations for the unknown Lagrange multipliers $\lambda_0, \lambda_1, \ldots, \lambda_N$. If we assume that the density $f(x)$ is normalized such that the first moment is always equal to 1 ($\mu_0 = 1$, i.e. respecting the unit sum constraint), the first equation in (9) then reads

$$\mu_0 = \int_a^\omega x^0 f_N(x) dx = \int_a^\omega \exp \left( -\lambda_0 - \sum_{n=1}^N \lambda_n x^n \right) dx = 1$$

and results in the first Lagrange multiplier, $\lambda_0$, being expressed in terms of the remaining Lagrange multipliers:

$$\int_a^\omega \exp \left( -\sum_{n=1}^N \lambda_n x^n \right) = e^{\lambda_0}.$$ 

The system of equations then reduces to

$$\mu_n = \frac{\int_a^\omega x^n \exp \left( -\sum_{n=1}^N \lambda_n x^n \right)}{\int_a^\omega \exp \left( -\sum_{n=1}^N \lambda_n x^n \right)}, \quad n = 0, 1, 2, \ldots, N.$$  

For a numerical solution, one introduces $\Gamma = \Gamma(\lambda_1, \lambda_2, \ldots, \lambda_N)$ through the Legendre transformation

$$\Gamma = \ln(e^{\lambda_0}) + \sum_{n=1}^N \mu_n \lambda_n,$$

where the $\mu_n$’s are the true statistical moments of the underlying density. The stationary points of the potential $\Gamma$ are solutions to the equations

$$\frac{\delta \Gamma}{\delta \lambda_n} = 0 \implies \mu_n, \quad n = 0, 1, 2, \ldots, N,$$

which is the solution to the finite moment problem. The convexity of $\Gamma$ guarantees that if a stationary point is found for some finite values of $\lambda_1, \lambda_2, \ldots, \lambda_N$ it must be a unique absolute minimum. A more detailed description and analytic demonstration of the method and also alternative algorithms for solving the finite moment problem can be found in Mead and Papanicolaou (1984).

Figure 2 shows the observed and reconstructed distribution of deaths using different numbers of observed statistical moments for the male population living in the USA in 1990. This distribution of deaths is a good study case because it is characterized by a pronounced level of mortality in the first year of life, an accident hump around age 20 and an exponential increase in adult and old age mortality. More recent distributions exhibit less pronounced local maxima or modes, therefore making them easier to estimate. Knowing only the first two moments, mean and variance, is not sufficient to obtain a good reconstruction of the underlying distribution. The obtained coverage, i.e. the common surface of the observed and estimated distributions, would be around 80%. However, the more moments we employ in the estimation procedure the bigger the coverage becomes. The results indicate that six moments are enough to obtain a coverage above 96% where the infant and adult mortality is captured adequately. We note here that above a large enough coverage level, the measure does not necessarily indicate a more accurate approximation of the true distribution, but better identification of the main body of the distribution.
2.4. The maximum entropy mortality model

The idea behind our forecasting method is simple. The future age-specific levels of mortality for a population are determined by extrapolating a limited number of statistical moments given by the available life table age-at-death distributions. The extrapolation is realized with multivariate time series models. The age-at-death distribution is estimated at any point in time from the predicted moments using the MaxEnt algorithm (introduced in Section 2.3).

Prior to generating future realizations, the moments of ordinal 3 and higher are normalized, and the logarithmic transformation is applied to the absolute values of all observed moments. This ensures that the relevant shape measures remain positive on any forecasting horizon (e.g. the mean and the variance of the distribution). The period index of interest to be used in forecasting, with first order differences, can be defined from the empirical central moments, $\tilde{\mu}_{n,t}$ in Equation (2), as follows:

$$y_{n,t} = \log |\tilde{\mu}_{n,t}| - \log |\tilde{\mu}_{n,t-1}|.$$ (15)
We are using a multivariate random-walk, with a vector of drift parameters \( \theta_n \), to drive the dynamics of the multiple period indices, so that

\[
y_{n,t} = \theta_n + y_{n,t-1} + \varepsilon_{n,t} \quad \text{with} \quad t = 1, 2, \ldots, \tau \quad \text{and} \quad n = 1, 2, \ldots, N,
\]

where \( \varepsilon_{n,t} \sim N(0, \Omega) \), with \( \Omega = CC' \). \( C \) represents the Cholesky factorization matrix of the variance-covariance matrix \( \Omega \). The parameters \( \theta_n \) and \( \Omega \) are estimated by ordinary least squares (OLS). A similar model is used by Haberman and Renshaw (2011) to generate trajectories of the multiple period indices in various mortality models.

Once the \( y_{n,t} \) forecasts are obtained, one can compute the statistical moments, estimate the distribution of deaths at time \( t \) using MaxEnt, and derive any other life table indicator by applying standard life table calculations (Preston et al. 2001).

### 2.5. Prediction intervals

We simulate prediction intervals for the indices of interest using an algorithm, which makes full allowance for the forecast error generated by the multivariate random-walk model.

**Algorithm:**

1. \( M \) simulations are performed on a forecasting horizon \( J \).
2. For simulation \( m = 1, 2, \ldots, M \):
   1. randomly sample a variable \( z_m^* \) from the multivariate normal distribution \( N(0, I) \);
   2. compute \( y_{n,t+j}^* = y_{n,t+j} + \hat{\theta}_n + \sqrt{j} \hat{C} z_m^* \);
   3. compute statistical moments \( \mu_{t+j,n,m}^* \);
   4. estimate the density using the MaxEnt algorithm and determine \( d_{x,t+j,m}^* \).

### 3. Case study: England and Wales 1960–2016 male mortality experience, ages 0–95

#### 3.1. The data

The data source used in this article is the Human Mortality Database (2018), which contains historical mortality data for 43 different countries and territories. HMD constitutes a reliable data source because it includes high-quality historical mortality data that was subject to a uniform set of procedures, guaranteeing the cross-national comparability of the information.

In order to test and illustrate the performance of the method, we fit the model using life table death counts for the male population in England and Wales between 1960 and 2016. Additional results for various countries are presented in Appendix 2.

#### 3.2. Model comparison

In addition to the MEM model, the following mortality models are evaluated and used to forecast mortality:

- The multivariate random-walk with drift model:

\[
\log(m_{x,t}) = \theta_x + \log(m_{x,t-1}) + \varepsilon_{x,t}.
\]

This model represents a simple linear extrapolation of the logarithm of the age-specific death rates, \( m_x \), based on the first and the last observed values in the multivariate time series.
- The Lee and Carter (1992) mortality model:

\[ \log(m_{x,t}) = \alpha_x + \beta_x k_t + \epsilon_{x,t}, \] (18)

which is a numerical algorithm to estimate the age-specific effects \( \alpha_x \) and \( \beta_x \) and employs the singular value decomposition (SVD) to derive a univariate time series vector \( k_t \), that becomes the main leading indicator of future mortality.

- The Hyndman and Ullah (2007) functional mortality model:

\[ \log(m_{x,t}) = \alpha_x + \sum_{k=1}^{K} \beta_{x,k} \phi_{t,k} + \epsilon_{x,t} + \sigma_{t} \epsilon_{x,t} \] (19)

This model is an extension of the Lee–Carter model where the sum term allows for smooth functions of age and \( \sigma_t \) allows the amount of noise to become age-specific.

- The Renshaw and Haberman (2006) model – a cohort-based extension of the Lee–Carter model:

\[ \log(m_{x,t}) = \alpha_x + \beta(1)_x k_t + \beta(0)_x \gamma_{t-x} + \epsilon_{x,t} \] (20)

The model incorporates a cohort effect \( \gamma_{t-x} \) to the Lee–Carter predictor to better capture discontinuities in mortality trends. We are using the simplified and more stable version of the model where \( \beta(0)_x = 1 \) at all ages, as suggested in Haberman and Renshaw (2011).

- The Oeppen (2008) model – the compositional-data mortality model:

\[ \text{clr}(f_{x,t}) = \alpha_x + \beta_x k_t + \epsilon_{x,t} \] (21)

Again, a variant of the Lee–Carter model where the index of interest to be modeled is the life table age-at-death distribution \( f(x) \), subject to a clr transformation (instead of a logarithmic one).

For all models, \( \epsilon_{x,t} \) are independent and identically distributed random variables (i.i.d) normally distributed with mean zero.

Thus, six models are evaluated. Four of them are targeting the log-transformed death rates, \( \log m_{x,t} \), and the other two (Oeppen and MEM) are focusing on modeling the age-specific frequencies of the age-at-death distribution or mortality data in compositional format. We mention that the random-walk model is chosen because of its simplicity, the Lee–Carter and Hyndman–Ullah methods are included in comparison because of their popularity and acceptance in the demographic and actuarial literature. The Renshaw–Haberman is a good example of how the projections can be improved by controlling for cohort effects and finally the Oeppen model is selected because of its similarity with the MEM model (mainly, the \( f(x) \) focus). Discussing the advantages and the features of the models used in comparison is beyond the scope of this article. For more details about the models, the readers are referred to the original articles.

The analysis is performed using the R programming language (R Core Team 2018). The Lee–Carter and the Hyndman–Ullah models are fitted and forecasted using the demography R package (2017) and the Renshaw–Haberman implementation is adopted from the StMoMo software (2015). The source code of the other three models can be downloaded and installed in the form of an R software package from the authors’ GitHub repository.

### 3.3. Evaluation and predictive power measurements

We assess the performance of the proposed MEM method and the other five mortality models based on forecasts of life expectancy at all ages\(^2\) as compared to the observed life expectancies. All the models

\(^2\) Because the five mortality extrapolation methods are modeling different life table indicators (\( m_x \) vs. \( f_x \)), in order to perform a fair comparison one needs to make sure that the life table computation guarantees the transitivity between the indicators. That is,
are fitted and evaluated over the 0–95 age range. Since a perfect fit of the data can always be obtained by using a model with enough parameters, and due to the fact that a good fit does not guarantee good forecasting performance (Hyndman and Athanasopoulos 2018), we will not evaluate the models based on their ability to fit the historical data. The models are evaluated based on the out-of-sample forecasting performance over the observed data, where we refer to sample as the data-set used in fitting. The predictive power is our ultimate goal, translated into a high degree of accuracy of forecast trajectories.

For each scenario and model, we are computing a matrix of accuracy errors. The overall accuracy of the forecast for a specific model-scenario is the average of all the values in the corresponding matrix. For a given population multiple scenarios of equal dimension (see Section 3.4) are explored in order to account for the robustness of the models. The aggregation is done by averaging the results over all scenarios for each model/accuracy-measure/country.

Many accuracy measures have been published. See Hyndman and Koehler (2006) for a comprehensive review of the most common accuracy measures used in forecasting literature. Only six of them are considered here:

- ME – Mean Error;
- MAE – Mean Absolute Error;
- MAPE – Mean Absolute Percentage Error;
- sMAPE – Symmetric Mean Absolute Percentage Error;
- sMRAE – Symmetric Mean Relative Absolute Error;
- MASE – Mean Absolute Scaled Error.

The sMRAE and MASE are accuracy measures computed relative to a benchmark model. In the case of sMRAE, the reference model in our study is the multivariate random-walk with drift, because we consider this model to be the simplest reliable method to extrapolate age-specific death rates. The MASE measure assesses the accuracy of a forecast with reference to a simple 1-step random-walk model (without drift). For all the presented accuracy measures, except ME, a smaller value is preferred over a larger one. A model performs better in terms of ME, compared to another model, if the obtained value is closer to zero, i.e. the smallest value in absolute terms. Because the six measures are evaluating the accuracy by analyzing different aspects of the realized forecasts, it is possible but not mandatory to obtain a different classification of the model performance, depending on the considered measure. We computed a general classification (GC) of the resulted accuracy performance of the analyzed models by considering the median classification over the six measures for each model. The best performing model is marked with: rank (1). A detailed description of the accuracy assessment process is provided in Appendix 1.

### 3.4. Out-of-sample forecasting strategy

The period between 1960 and 2016 is long enough and relevant at the same time for assessing the predictive power of the estimated models. A typical testing scenario would use sub-periods of 20 years of data to fit/train the models, and subsequent periods of 20 years of data to forecast and validate the results. Multiple testing scenarios are defined by rolling forward the training/validation windows in steps of 1 year. This strategy will be called: 20-20-1. Therefore, considering our timeline, in the first scenario we will use data from 1960 to 1979 for fitting the models. The resulting models will be used to predict mortality for the period between 1980 and 1999, and validate the forecasts against the observed values in the same period. We will refer to this as the 1960–1979–1999 scenario. By moving the evaluation windows 1 year forward, the second scenario would be 1961–1980–2000. And so on.

---

given a certain mortality level, the same values of life expectancy (indicator evaluated in this article) are obtained regardless of whether the life table construction starts from $m_x$, $q_x$, $l_x$ or $f_x$. In this article, the goal is achieved by using the life table methods implemented in the MortalityLaws R package (2018).
until the last scenario: 1977–1996–2016. In total, 18 scenarios have been defined, containing equal fitting and forecasting period lengths, making possible in this way the aggregation (by averaging) and comparison of the accuracy results over all scenarios in addition to the specific scenario results.

### 3.5. Results

Across multiple measures of forecast accuracy computed based on the mortality experience of the male population living in England and Wales, we find that the predictive power of the MEM method stands out when compared to the other models. The back-testing of the male mortality in England and Wales indicates that only the Renshaw–Haberman model returns slightly more accurate results. Table 1 displays a summary of the aggregated measures over 18 defined scenarios. We also learn that the differences between the multivariate random-walk with drift, Lee–Carter and Hyndman–Ullah models are insignificant in the case of this population; that the Oeppen method consistently offers better results among the Lee–Carter type models, obtaining the third position among the best performing models in this study.

When the models are tested over the mortality experience of multiple populations using the same strategy we discovered a similar pattern, namely, in the majority of the cases the methods based of age-at-death distribution (MEM and Oeppen) would be ranked as the best performing models. The performance of the Renshaw–Haberman model is interesting to observe, because in some countries like the Netherlands, Sweden and England and Wales, it is able to capture the cohort effects and successfully extrapolate them into the future predicting adequately the true mortality experience. However, the same strategy applied in France, Sweden or the USA completely places the model on the last position in the general classification and the reliance on observed cohort effects can create more damage to an improvement of the forecasts. The results are shown in Table 2 and Appendix 2.

The projected trajectories given by these methods can be inspected across different life table indicators, which are equivalent in the sense that they represent the same level of mortality. In Figure 4, we show the resulting mean trends in life expectancy at age 0, 25, 45, 65, 75 and 85; and in Figure 5 the trends in central death rate at the same ages are represented. It is noticeable that the MEM can cope with different levels of mortality improvement over age and time. This is the main reason why the model is able to return significantly better forecasts.

### Table 1. Forecast accuracy measures aggregated over 18 scenarios in the 1960–2016 period.

| Model                     | ME   | MAE  | MAPE  | sMAPE | sMRAE | MASE  | GC |
|---------------------------|------|------|-------|-------|-------|-------|----|
| M.Random-Walk w Drift     | 0.76 (6) | 0.77 (6) | 3.07 (5) | 3.16 (5) | 100.00 (5) | 4.20 (5) | (5) |
| Lee–Carter                | 0.74 (4) | 0.74 (4) | 3.02 (4) | 3.11 (4) | 99.00 (4) | 4.10 (4) | (4) |
| Hyndman–Ullah             | 0.76 (5) | 0.77 (5) | 3.10 (6) | 3.19 (6) | 100.67 (6) | 4.22 (6) | (6) |
| Renshaw–Haberman          | 0.40 (1) | 0.44 (1) | 1.92 (1) | 1.92 (1) | 76.07 (1) | 2.64 (1) | (1) |
| Oeppen                    | 0.60 (3) | 0.61 (3) | 2.59 (3) | 2.66 (3) | 89.09 (3) | 3.44 (3) | (3) |
| MEM-6                     | 0.48 (2) | 0.49 (2) | 2.07 (2) | 2.11 (2) | 77.85 (2) | 2.73 (2) | (2) |

### Table 2. General classification of the model predictive power for male mortality experience in England and Wales and nine developed countries. Period analyzed: 1960–2016.

|          | M.R.WD | Lee–Carter | Hyndman–Ullah | Renshaw–Haberman | Oeppen | MEM |
|----------|--------|------------|---------------|------------------|--------|-----|
| Australia| (6)    | (4)        | (3)           | (4)              | (2)    | (1) |
| Canada   | (5)    | (6)        | (3)           | (2)              | (4)    | (1) |
| France   | (5)    | (3)        | (3)           | (6)              | (2)    | (1) |
| England and Wales | (5) | (4) | (6) | (1) | (3) | (2) |
| Italy    | (4)    | (5)        | (6)           | (2)              | (3)    | (1) |
| The Netherlands | (3) | (6) | (5) | (1) | (4) | (2) |
| Spain    | (3)    | (1)        | (2)           | (6)              | (4)    | (5) |
| Switzerland | (3) | (3) | (6) | (5) | (2) | (1) |
| Sweden   | (4)    | (5)        | (6)           | (1)              | (3)    | (1) |
| USA      | (4)    | (3)        | (5)           | (6)              | (1)    | (2) |
3.6. How many moments to use in MEM forecasting?

In general, the number of statistical moments to be considered in the MEM model depends on the regularity of the age-at-death distribution in the population of interest. The more moments employed, the more accurate the estimation of the underlying distribution becomes. However, the cost of using a larger number of moments is paid in processing speed and the likelihood of convergence of the MaxEnt algorithm. For seven or more moments, a more complex time series model for moment extrapolation might be required.

Table 3. Forecast accuracy measures aggregated over 18 scenarios in the 1960–2016 period.

| Model                          | ME    | MAE   | MAPE  | sMAPE | sMRAE | MASE  | GC |
|--------------------------------|-------|-------|-------|-------|-------|-------|----|
| M.Random-Walk w Drift          | 0.76 (6) | 0.77 (6) | 3.07 (5) | 3.16 (5) | 100.00 (5) | 4.20 (5) | (5) |
| MEM-2                          | 0.75 (5) | 0.76 (5) | 4.02 (6) | 4.20 (6) | 101.98 (6) | 4.65 (6) | (6) |
| MEM-3                          | 0.54 (4) | 0.55 (4) | 2.70 (4) | 2.78 (4) | 86.25 (4) | 3.29 (4) | (4) |
| MEM-4                          | 0.49 (2) | 0.50 (2) | 2.36 (3) | 2.42 (3) | 82.00 (3) | 2.95 (3) | (3) |
| MEM-5                          | 0.49 (3) | 0.50 (3) | 2.15 (2) | 2.20 (2) | 79.52 (2) | 2.82 (2) | (2) |
| MEM-6                          | 0.48 (1) | 0.49 (1) | 2.07 (1) | 2.11 (1) | 77.85 (1) | 2.73 (1) | (1) |
We tested the MEM models of order 2–6, that is models that are estimated based on the first 2, 3 until 6 moments (plus $\mu_0$). The testing was carried out in the same manner and over the same scenarios as in Section 3.5. From the results in Table 3, we learn that only the MEM-2 is disqualified by the benchmark model, all the other variants of the MEM return significantly better results.

4. Conclusion and discussion

The maximum entropy mortality model represents a new approach to modeling age-specific mortality levels. Nevertheless, its novelty refers only to the authors' idea of defining an algorithm to forecast mortality using well established methods and concepts like statistical moments, information entropy, the finite moment problem and the multivariate random-walk with drift introduced decades or centuries ago. All these individually have an important application in multiple scientific fields.

The main advantage of the MEM is that the possible forecast age-specific trends are no longer based on the assumption of constant changes in mortality as in the Lee–Carter model. A different speed of improvement can be predicted across ages by taking into account the observed dynamics of the distribution of deaths and the changes in its shape and location. This makes possible the identification of the location of the longevity risk across the age-axis. The model has the required features to predict the different rates of change in life expectancy at different ages, e.g. by maintaining a linear increase in life expectancy at birth and at the same time inducing accelerating rates for ages above 65. This result is consistent with the observed trends in the past and across many developed countries. The method was not designed to capture cohort effects, instead our approach is purely statistical. However, the results presented here show that MEM can capture cohort effects more accurately than most of the other models. This is partially due to the approach of modeling mortality in a compositional framework and focusing on the death distributions (i.e. if a life is saved at age $x$ it must be placed back in the death-distribution at later age, say $x+\tau$).

Even if it is not shown here, the model introduced in this article is flexible enough in the sense that different covariates, influencing the mortality dynamics, can be considered in order to further improve the predictions. Examples of such covariates are information on the prevalence of smoking or obesity but also the trends in life expectancy at birth and modal age at death. This can be done by extending the time series model used in extrapolation (the multivariate random-walk with drift) by attaching extra cause-specific parameters. Similarly, a wide range of multivariate autoregressive time series models can be used to capture the coherent trends given by the observed empirical statistical moments, however this is subject to the requirements imposed by the available data.

Including higher order moments in the prediction can sometimes increase the importance of relatively small effects seen in the age-at-death distribution such as the accident hump for males or infant mortality. In countries with a low level of mortality, four moments can return pertinent results but in countries that still exhibit a pronounced multi-modal distribution of deaths a larger number of moments might be required. The coverage proportion, introduced in the article, is a very good measure for studying the models ability to estimate the shape of the true distribution, but it has an important drawback: two errors of the same magnitude at different ends of the distribution will be assigned the same weight in the measure. This is a drawback because underestimating or overestimating the force of mortality at younger ages has a higher impact on the general level of mortality of a population than underestimating or overestimating the force of mortality at an older age. If we investigate the evolution of longevity by looking at life expectancy, we can see that relying solely on information given by the mean age at death and life span disparity in order to generate forecasts is an endeavor doomed to failure (as showed in Table 3). The model would mostly return pronounced pessimistic or overoptimistic results across ages without a correspondence to reality. A real improvement can be noticed if 4–6 moments are used.

Other possible extensions of the MEM method worth exploring in the future are the use of an optimal weighting scheme which decreases the importance of higher order moments or the use of
various smoothing methods that can be applied to the data prior to computing the observed moments and fitting the model.

An important finding revealed in this study is the superiority of the age-at-death distribution based extrapolation methods, like the MEM and Oeppen (2008) over the death rate-based extrapolation methods.

Reproducible research

The presented model and algorithm is implemented using the \texttt{R} programming language (R Core Team 2018) and can be downloaded and installed in form of an \texttt{R} software package from authors GitHub repository (https://github.com/mpascariu/MortalityForecast). The results and figures for the countries presented in the article can be reproduced using the code and data saved in the \texttt{MortalityForecast} package – version 0.8.0.

Acknowledgments

The authors thank researchers at the Interdisciplinary Center on Population Dynamics, University of Southern Denmark, for their helpful comments and for many useful discussions which have helped improve this work. The first author wrote parts of this paper while visiting the School of Demography at the Australian National University; he thanks them for their kind disposition.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work is conducted within the ”Modelling and Forecasting Age-Specific Death at Older Age”, Project [No. 95-103-31186], under the management of the University of Southern Denmark, Institute of Public Health with the generous financial support of the SCOR Corporate Foundation for Science. The authors thank the funding institution. Researchers operated independently from the funders on this work, and the funding organizations had no role in the study design, data collection, analysis, interpretation, writing of the report or the decision to submit it for publication.

ORCID

Marius D. Pascariu http://orcid.org/0000-0002-2568-6489
Adam Lenart http://orcid.org/0000-0001-9288-9510
Vladimir Canudas-Romo http://orcid.org/0000-0001-6532-0089

References

Aitchison, J. (1982). The statistical analysis of compositional data. 	extit{Journal of the Royal Statistical Society. Series B (Methodological)} 44(2), 139–177.
Aitchison, J. (1986). 	extit{A concise guide to compositional data analysis}. London: Chapman and Hall. 2015.
Baselli, U. & Camarda, C. G. (2019). Modelling and forecasting adult age-at-death distributions. 	extit{Population Studies} 73(1), 119–138.
Bergeron-Boucher, M.-P., Canudas-Romo, V., Oeppen, J. & Vaupel, J. W. (2017). Coherent forecasts of mortality with compositional data analysis. 	extit{Demographic Research} 37, 527–566.
Bøh-Ewald, C., Li, P. & Myrskylä, M. (2018). Forecast accuracy hardly improves with method complexity when completing cohort fertility. 	extit{Proceedings of the National Academy of Sciences} 115(37), 9187–9192.
Chihara, T. S. (1978). 	extit{An introduction to orthogonal polynomials}. New York: Gordon and Breach.
Dellaportas, P., Smith, A. F. & Stavropoulos, P. (2001). Bayesian analysis of mortality data. 	extit{Journal of the Royal Statistical Society: Series A (Statistics in Society)} 164(2), 275–291.
Frontini, M., Rodriguez, G. & Seatzu, S. (1990). An algorithm for computing minimum norm solutions of finite moment problem. In Mason, J. C. and Cox, M. G., editors, 	extit{Algorithms for Approximation II}, p. 361–368. London: Chapman and Hall.
Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. 	extit{Philosophical transactions of the Royal Society of London} 115, 513–583.
Appendix 1. Accuracy assessment

The performance of the six statistical models is assessed based on forecasts of life expectancy. Depending on the model output, the complete remaining life expectancy at age $x$, $e(x)$, can be defined in terms of central death rate $m(x)$,

$$e(x) = \int_x^\infty \exp[-m(s)]ds,$$

and with reference to the death frequency $f(x)$ as follows,

$$e(x) = \int_x^\infty sf(s)ds / \int_x^\infty f(s)ds - x.$$
The discrete approximations follow the method presented in Preston et al. (2001) and implemented in the MortalityLaws software package. Let \( e(x, t) \) be the expected remaining lifetime of an \( x \)-year-old calculated using the observed mortality rates at time \( t \), and \( \hat{e}(x, t) \) the forecast value, respectively. Then, for each forecasting-scenario, model and country, a matrix of accuracy errors is obtained,

\[
E = \begin{bmatrix}
\delta(0, 1) & \delta(0, 2) & \ldots & \delta(0, \tau) \\
\delta(1, 1) & \delta(1, 2) & \ldots & \delta(1, \tau) \\
\vdots & \vdots & \ddots & \vdots \\
\delta(\omega, 1) & \delta(\omega, 2) & \ldots & \delta(\omega, \tau)
\end{bmatrix}
\]  

(A3)

where \( \delta(x, t) \) is the distance between the observed and predicted life expectancy calculated as,

\[
\delta(x, t) = e(x, t) - \hat{e}(x, t), \quad x = 0, \ldots, \omega, \quad t = 1, \ldots, \tau.
\]  

(A4)

The aggregate accuracy measures are computed over a forecasting window of 20 years and over the 0-95 age range, i.e. \( \tau = 20 \) and \( \omega = 95 \). The aggregation process is done by averaging the accuracy results over age and time. To further simplify the notations, we define algebraically the mean operation of a random variable \( z(x, t) \) as:

\[
\text{mean}\left[ z(x, t) \right] = \frac{\sum_{x=0}^{\omega} \sum_{t=1}^{\tau} z(x, t)}{(\omega + 1)\tau}
\]  

(A5)

Then the six accuracy measures can be summarized as:

- **ME** – Mean Error
  \[
  ME = \text{mean}\left[ \delta(x, t) \right]
  \]  

(A6)

- **MAE** – Mean Absolute Error
  \[
  MAE = \text{mean}\left[ |\delta(x, t)| \right]
  \]  

(A7)

- **MAPE** – Mean Absolute Percentage Error
  \[
  MAPE = \text{mean}\left[ 100 \times \frac{|\delta(x, t)|}{e(x, t)} \right]
  \]  

(A8)

- **sMAPE** – Symmetric Mean Absolute Percentage Error
  \[
  sMAPE = \text{mean}\left[ 200 \times \frac{|\delta(x, t)|}{e(x, t) + \hat{e}(x, t)} \right]
  \]  

(A9)

- **sMRAE** – Symmetric Mean Relative Absolute Error
  Let \( \delta_B(x, t) \) be the benchmark error. Specifically \( \delta_B(x, t) \) is represented by the errors generated by a naïve model, here the multivariate random-walk with drift model. Then,

  \[
  sMRAE = \text{mean}\left[ 200 \times \frac{|\delta(x, t)|}{|\delta(x, t)| + |\delta_B(x, t)|} \right]
  \]  

(A10)

- **MASE** – Mean Absolute Scaled Error
  \[
  MASE = \text{mean}\left[ \frac{\delta(x, t)}{\frac{1}{\tau - 1} \sum_{t=2}^{\tau} [e(x, t) - e(x, t - 1)]} \right]
  \]  

(A11)

For a specific country and model, multiple scenarios of equal dimension (fitted/forecasted years) are explored in order to account for the robustness of the models. Usually, 18 scenarios per country. The aggregation is done by averaging the results over all scenarios for each model/accuracy-measure/country.

### Appendix 2. Out-of-sample forecasts

In this section, we present the results of our study performed in nine additional developed countries.
A.1 Out-of-sample forecasts: Australia, Male population, 1960–2014

Table A1. Forecast accuracy measures aggregated over 16 scenarios in the 1960–2014 period.

| Model                          | ME    | MAE   | MAPE  | sMAPE | sMRAE | MASE  | GC   |
|--------------------------------|-------|-------|-------|-------|-------|-------|------|
| M.Random-Walk w Drift          | 0.62(5) | 0.64(5) | 2.51(5) | 2.56(5) | 100.00(6) | 3.21(5) | (6)  |
| Lee–Carter                     | 0.55(4) | 0.58(4) | 2.26(4) | 2.30(4) | 96.03(5) | 2.91(4) | (4)  |
| Hyndman–Ullah                  | 0.48(3) | 0.51(3) | 2.11(3) | 2.14(3) | 92.80(4) | 2.64(3) | (3)  |
| Renshaw–Haberman               | 0.83(6) | 1.11(6) | 4.40(6) | 6.63(6) | 84.62(3) | 5.52(6) | (4)  |
| Oeppen                         | 0.36(2) | 0.42(2) | 1.77(2) | 1.80(2) | 80.89(2) | 2.18(2) | (2)  |
| MEM-6                          | 0.12(1) | 0.39(1) | 1.67(1) | 1.67(1) | 78.92(1) | 2.08(1) | (1)  |

Figure A1. Out-of-sample forecast of the remaining life expectancy and central death rates at various ages using the five mortality models (Australia, Male population, Scenario 16: 1975–1994–2014).
A.2 Out-of-sample forecasts: Canada, Male population, 1960–2011

Table A2. Forecast accuracy measures aggregated over 13 scenarios in the 1960–2011 period.

| Model                        | ME   | MAE  | MAPE | sMAPE | sMRAE | MASE  | GC |
|------------------------------|------|------|------|-------|-------|-------|----|
| M.Random-Walk w Drift        | 0.56 | 0.57 | 2.07 | 2.10  | 100.00| 3.58  | 5  |
| Lee–Carter                   | 0.60 | 0.62 | 2.25 | 2.29  | 104.39| 3.88  | 6  |
| Hyndman–Ullah                | 0.44 | 0.50 | 1.98 | 2.01  | 92.82 | 3.27  | 3  |
| Renshaw–Haberman             | 0.12 | 0.35 | 1.55 | 1.54  | 83.31 | 2.51  | 2  |
| Oeppen                       | 0.53 | 0.55 | 2.07 | 2.10  | 99.10 | 3.52  | 4  |
| MEM-6                        | 0.17 | 0.26 | 1.11 | 1.12  | 71.32 | 1.81  | 1  |

Figure A2. Out-of-sample forecast of the remaining life expectancy and central death rates at various ages using the five mortality models (Canada, Male population, Scenario 13: 1972–1991–2011).
### A.3 Out-of-sample forecasts: France, Male population, 1960–2016

Table A3. Forecast accuracy measures aggregated over 18 scenarios in the 1960–2016 period.

| Model                        | ME    | MAE   | MAPE  | sMAPE | sMRAE | MASE | GC |
|------------------------------|-------|-------|-------|-------|-------|------|----|
| M.Random-Walk w Drift        | 0.51 (6) | 0.53 (5) | 2.10 (5) | 2.14 (5) | 100.00 (5) | 3.47 (5) | (5) |
| Lee–Carter                   | 0.49 (5) | 0.52 (4) | 1.90 (4) | 1.93 (3) | 96.00 (4) | 3.23 (4) | (3) |
| Hyndman–Ullah                | 0.47 (4) | 0.51 (3) | 1.90 (3) | 1.93 (4) | 93.75 (3) | 3.21 (3) | (3) |
| Renshaw–Haberman             | 0.44 (3) | 0.53 (6) | 2.37 (6) | 2.36 (6) | 101.03 (6) | 3.88 (6) | (6) |
| Oeppen                       | 0.27 (1) | 0.38 (2) | 1.56 (1) | 1.58 (1) | 84.97 (1) | 2.51 (1) | (2) |
| MEM-6                        | 0.27 (2) | 0.38 (1) | 1.68 (2) | 1.69 (2) | 86.80 (2) | 2.61 (2) | (1) |

Figure A3. Out-of-sample forecast of the remaining life expectancy and central death rates at various ages using the five mortality models (France, Male population, Scenario 18: 1977–1996–2016).
### A.4 Out-of-sample forecasts: Italy, Male population, 1960–2014

Table A4. Forecast accuracy measures aggregated over 16 scenarios in the 1960–2014 period.

| Model                          | ME  | MAE  | MAPE | sMAPE | sMRAE | MASE | GC |
|-------------------------------|-----|------|------|-------|-------|------|----|
| M.Random-Walk w Drift         | 0.82(4) | 0.83(4) | 3.21(4) | 3.30(4) | 100.00(4) | 4.64(4) | (4) |
| Lee–Carter                    | 0.92(5) | 0.92(5) | 3.40(5) | 3.50(5) | 103.97(6) | 5.01(5) | (5) |
| Hyndman–Ullah                 | 0.92(6) | 0.93(6) | 3.41(6) | 3.51(6) | 103.79(5) | 5.03(6) | (6) |
| Renshaw–Haberman              | 0.54(2) | 0.58(2) | 2.53(2) | 2.57(2) | 85.64(2) | 3.55(2) | (2) |
| Oeppen                        | 0.80(3) | 0.80(3) | 3.02(3) | 3.10(3) | 95.54(3) | 4.40(3) | (3) |
| MEM-5                         | 0.37(1) | 0.41(1) | 2.01(1) | 2.06(1) | 71.72(1) | 2.58(1) | (1) |

Figure A4. Out-of-sample forecast of the remaining life expectancy and central death rates at various ages using the five mortality models (Italy, Male population, Scenario 16: 1975–1994–2014).
### Table A5. Forecast accuracy measures aggregated over 18 scenarios in the 1960–2016 period.

| Model                      | ME    | MAE   | MAPE  | sMAPE | sMRAE | MASE | GC |
|----------------------------|-------|-------|-------|-------|-------|------|----|
| M.Random-Walk w Drift      | 0.75 (3) | 0.77 (3) | 2.99 (3) | 3.07 (3) | 100.00 (3) | 4.55 (3) | (3) |
| Lee–Carter                 | 0.79 (6) | 0.81 (6) | 3.23 (5) | 3.33 (5) | 104.27 (5) | 4.88 (6) | (6) |
| Hyndman–Ullah              | 0.79 (5) | 0.81 (5) | 3.24 (6) | 3.33 (6) | 104.37 (6) | 4.88 (5) | (5) |
| Renshaw–Haberman           | 0.30 (1) | 0.36 (1) | 1.48 (1) | 1.49 (1) | 70.61 (1) | 2.22 (1) | (1) |
| Oeppen                     | 0.77 (4) | 0.80 (4) | 3.19 (4) | 3.29 (4) | 103.62 (4) | 4.89 (4) | (4) |
| MEM-5                      | 0.59 (2) | 0.62 (2) | 2.53 (2) | 2.59 (2) | 90.04 (2) | 3.81 (2) | (2) |

**Figure A5.** Out-of-sample forecast of the remaining life expectancy and central death rates at various ages using the five mortality models (The Netherlands, Male population, Scenario 18: 1977–1996–2016).
A.6 Out-of-sample forecasts: Spain, Male population, 1960–2014

Table A6. Forecast accuracy measures aggregated over 16 scenarios in the 1960–2014 period.

| Model                        | ME   | MAE  | MAPE  | sMAPE | sMRAE | MASE  | GC  |
|------------------------------|------|------|-------|-------|-------|-------|-----|
| M.Random-Walk w Drift        | 0.12 | 0.31 | 1.35  | 1.34  | 100.00| 2.28  | (3) |
| Lee–Carter                   | 0.09 | 0.34 | 1.17  | 1.17  | 97.62 | 2.19  | (1) |
| Hyndman–Ullah                | 0.13 | 0.34 | 1.18  | 1.18  | 97.72 | 2.21  | (2) |
| Renshaw–Haberman             | 0.45 | 0.84 | 3.44  | 4.65  | 114.59| 6.14  | (6) |
| Oeppen                       | -0.14| 0.37 | 1.38  | 1.37  | 105.86| 2.54  | (4) |
| MEM-5                        | -0.30| 0.50 | 1.65  | 1.63  | 111.11| 3.42  | (5) |

Figure A6. Out-of-sample forecast of the remaining life expectancy and central death rates at various ages using the five mortality models (Spain, Male population, Scenario 16: 1975–1994–2014).
A.7 Out-of-sample forecasts: Switzerland, Male population, 1960–2016

Table A7. Forecast accuracy measures aggregated over 18 scenarios in the 1960–2016 period.

| Model                      | ME    | MAE   | MAPE  | sMAPE | sMRAE | MASE  | GC  |
|----------------------------|-------|-------|-------|-------|-------|-------|-----|
| M.Random-Walk w Drift      | 0.58(3) | 0.61(3) | 2.19(4) | 2.22(4) | 100.00(4) | 3.34(3) | (3) |
| Lee–Carter                 | 0.62(4) | 0.64(4) | 2.17(3) | 2.20(3) | 100.91(5) | 3.43(4) | (3) |
| Hyndman–Ullah              | 0.66(5) | 0.68(5) | 2.30(5) | 2.34(5) | 106.15(6) | 3.65(5) | (6) |
| Renshaw–Haberman           | 1.13(6) | 1.23(6) | 4.24(6) | 6.45(6) | 93.44(3) | 6.69(6) | (5) |
| Oeppen                     | 0.45(2) | 0.50(2) | 1.74(2) | 1.76(2) | 89.54(2) | 2.69(2) | (2) |
| MEM-5                      | 0.35(1) | 0.40(1) | 1.41(1) | 1.42(1) | 81.59(1) | 2.17(1) | (1) |

Figure A7. Out-of-sample forecast of the remaining life expectancy and central death rates at various ages using the five mortality models (Switzerland, Male population, Scenario 18: 1977–1996–2016).
A.8 Out-of-sample forecasts: Sweden, Male population, 1960–2016

Table A8. Forecast accuracy measures aggregated over 18 scenarios in the 1960–2016 period.

| Model                          | ME    | MAE   | MAPE  | sMAPE | sMRAE | MASE  | GC |
|-------------------------------|-------|-------|-------|-------|-------|-------|----|
| M.Random-Walk w Drift         | 0.85  | 0.86  | 3.02  | 3.09  | 100.00| 5.24  | (4)|
| Lee–Carter                    | 0.87  | 0.88  | 3.06  | 3.13  | 98.85 | 5.33  | (5)|
| Hyndman–Ullah                 | 0.92  | 0.92  | 3.19  | 3.28  | 104.01| 5.60  | (6)|
| Renshaw–Haberman              | 0.68  | 0.72  | 2.70  | 2.75  | 86.45 | 4.49  | (1)|
| Oeppen                        | 0.81  | 0.82  | 2.87  | 2.94  | 93.12 | 4.98  | (3)|
| MEM-5                         | 0.68  | 0.70  | 2.54  | 2.60  | 86.90 | 4.26  | (1)|

Figure A8. Out-of-sample forecast of the remaining life expectancy and central death rates at various ages using the five mortality models (Sweden, Male population, Scenario 18: 1977–1996–2016).
A.9 Out-of-sample forecasts: USA, Male population, 1960–2016

Table A9. Forecast accuracy measures aggregated over 18 scenarios in the 1960–2016 period.

| Model                                      | ME    | MAE    | MAPE   | sMAPE  | sMRAE  | MASE   | GC  |
|--------------------------------------------|-------|--------|--------|--------|--------|--------|-----|
| M.Random-Walk w Drift                      | 0.24(5)| 0.30(4)| 1.32(4)| 1.33(4)| 100.00(4)| 2.57(4)| (4) |
| Lee–Carter                                 | 0.23(4)| 0.30(3)| 1.29(3)| 1.31(3)| 98.69(3)| 2.51(3)| (3) |
| Hyndman–Ullah                              | 0.15(3)| 0.42(5)| 1.76(5)| 1.78(5)| 111.15(5)| 3.49(5)| (5) |
| Renshaw–Haberman                           | 2.84(6)| 3.35(6)| 12.37(6)| 21.27(6)| 126.88(6)| 27.97(6)| (6) |
| Oeppen                                     | 0.09(1)| 0.22(1)| 1.07(2)| 1.08(2)| 87.29(1)| 2.04(2)| (1) |
| MEM-6                                      | -0.10(2)| 0.24(2)| 1.03(1)| 1.02(1)| 92.31(2)| 2.03(1)| (2) |

Figure A9. Out-of-sample forecast of the remaining life expectancy and central death rates at various ages using the five mortality models (USA, Male population, Scenario 18: 1977–1996–2016).