MILLIMAN WHITE PAPER

Mortality trend prediction using machine learning

Alexandre Boumezoued
Amal Elfassihi

Milliman

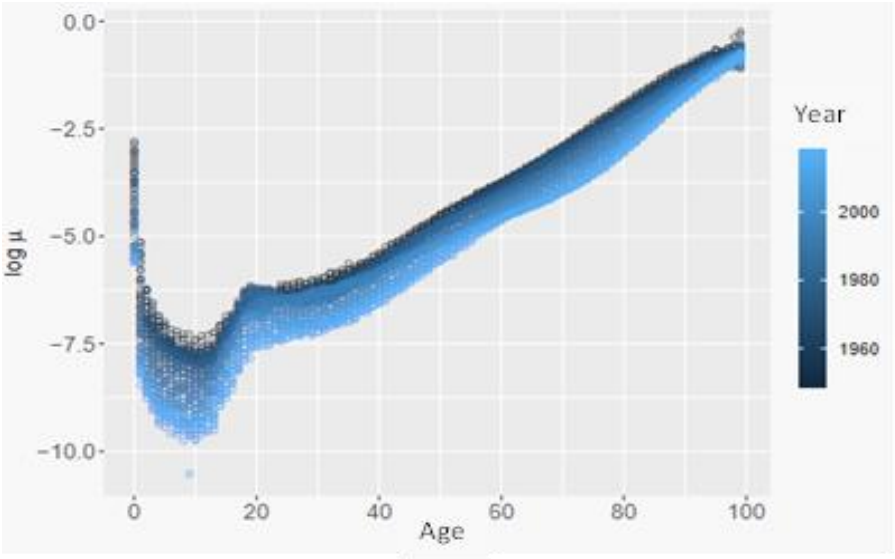
Context

During the 20th century, human mortality was globally reduced: in most industrialised countries, mortality among adults and the elderly shows a decreasing annual probability of death. This is well illustrated by Figure 1, which represents the logarithm of the mortality rate for men in the French population as a function of age, over the time period 1950 to 2016.

In the literature on mortality modelling, machine learning approaches have recently emerged. The work of Deprez et al. (2017) has shown that machine learning algorithms are useful for assessing the quality of fit of mortality estimates provided by standard stochastic mortality models [1]. The authors applied a gradient boosting model to analyse how modelling should be improved based on an individual's characteristics, such as age or birth cohort. This regression approach (nonparametric) then makes it possible to detect the weaknesses of different stochastic mortality models. Hainaut (2018) used neural networks to find the latent mortality factors of a Lee-Carter model and predict them according to a random walk with drift [2]. Richman and Wüthrich (2019) extended the Lee-Carter model to multiple populations using deep neural networks [3].

Thus, we propose in this article to predict future mortality from past data and to compare the output predictions of classic stochastic mortality models (baseline models) with those of machine learning model approaches (neural network and random forest).

FIGURE 1: LOGARITHM OF THE MORTALITY RATE FOR MEN IN THE FRENCH POPULATION AS A FUNCTION OF AGE, OVER THE TIME PERIOD 1950-2016



Note that the companion Milliman paper to the present one by Postema & van Es explores using a Temporal Fusion Transformer (TFT) model on multi-population, age-specific mortality data that has been enriched with socioeconomic data collected by the World Bank [4].

It should be noted that the approaches described in this paper do not incorporate assumptions about future advances in medical science or specific environmental changes: no information other than historical data is taken into account by our models. This means that the models described are unable to predict sudden improvements in mortality due to the discovery of new medical treatments. Likewise, future deterioration caused by epidemics or pandemics, the appearance of new diseases, climate change or worsening pollution cannot enter into these models. The actuary should keep this in mind when designing a reinsurance program.

Data

THE HUMAN MORTALITY DATABASE

The Human Mortality Database (HMD) is a joint initiative of the Department of Demography at the University of California at Berkeley in the United States and the Max Planck Institute for Demographic Research in Rostock, Germany [7]. It was created in 2000 with the objective of bringing together detailed data on mortality and population, and to serve as a reference for anyone interested in human longevity. It contains data in the form of periodic tables by year, age and sex, and also in the form of cohort tables.

CONSTRUCTION OF THE DATA SET

Our data set is defined in several steps.

1. Retained countries

All of the HMD countries were retained, with the exception of countries which do not have at least 10 years of data before year 2000. Thus, 38 countries are included in the data set.

2. Discussion of the high ages retained in the data set

For ages greater than 94, there are data points corresponding to estimated forces of mortality greater than 1.

Thus, for example, we have for the point $x_0 = (g = Male, k = Slovenia, a = 99, t = 1995)$:

$$\hat{\mu}_{x_0} = \frac{d_{x_0}}{E_{x_0}} = \frac{6}{3.5} \approx 1.71$$

This (realized) mortality force data $\hat{\mu}_{x_0}$ being an estimate of the true mortality force μ_{x_0} , we can calculate the 95% confidence interval (CI) for μ_{x_0}.

We have:

$$\mu_{x_0} \approx \hat{\mu}_{x_0} \pm 1.96 \sqrt{\frac{\hat{\mu}_{x_0}}{E_{x_0}}} = \hat{\mu}_{x_0} \left(1 \pm \frac{1.96}{\sqrt{d_{x_0}}}\right) \approx 1.71 \pm 1.96 \times 0.70.$$

The 95% CI is therefore approximately [0.34, 3.09].

Consequently it appears that the coefficient of variation is extremely high, which implies a significant uncertainty on μ_{x_0}. In the remainder of the comparative analysis, we have chosen to exclude ages greater than 94 in our data set.

3. The retained data set

The data set consists of n observations and is written:

$$D = \{(x_i, \log \mu_i) \mid \mu_i \text{ is available}, a \in [0,94], t \in [1950, 2016]\}$$

where x denotes the vector of input variables and the logarithm of the (estimated) mortality force $\log \mu$ is the output variable.

Note that when a mortality rate is recorded as zero or is not available in the data, the mean imputation method is applied, i.e., the data point is replaced by the average mortality rate over all the other countries for the same year, age and gender.

4. Separation of the data set into a training set and a test set

The total data set D includes HMD data for the years 1950 to 2016. The data split in two different ways into a training set D_{train} and a test set D_{test}:

- **Split 1:** D_{train} includes data from 1950 to 1990 and D_{test} includes data from 1991 to 1999, i.e., about 20% of D_1, which includes data from 1950 to 1999.
- **Split 2:** D_{train} includes data from 1950 to 1999 and D_{test} includes data from 2000 to 2016, which is about 25% of the total dataset D.

Comparison of mortality prediction approaches

BASELINE MODELS

1. The Lee-Carter model

The Lee-Carter model, also known as M1 model, can be written as follows:

$$log\big(\mu(x,t)\big) = \alpha(x) + \beta(x)\kappa(t),$$

where:

- $\mu(x,t)$ is the force of mortality
- $\alpha(x)$ captures the age pattern of mortality
- $\beta(x)$ captures the sensitivity of each age to the trend
- $\kappa(t)$ models the evolution of mortality over time and its stochastic variations

The force of mortality is then projected over the years corresponding to the test sample D_{test}, by modelling the series of $\kappa(t)$ by an autoregressive integrated moving average (ARIMA) process whose orders are chosen by the Box-Jenkins method [8].

2. The APC model

The Age-Period-Cohort (APC) model, also known as M3 model, can be written as follows:

$$log\big(\mu(x,t)\big) = \alpha(x) + \kappa(t) + \gamma(t-x),$$

where:

- $\mu(x,t)$ is the force of mortality
- $\alpha(x)$ captures the age pattern of mortality
- $\kappa(t)$ models the evolution of mortality over time and its stochastic variations
- $\gamma(t-x)$ is the cohort effect parameter

The force of mortality is then projected over the years corresponding to the test sample D_{test}, by modelling the series of $\kappa(t)$ by an ARIMA process whose orders are chosen by the Box-Jenkins method. The $\gamma(t-x)$ series is modelled by an ARIMA(1,1,0) process.

MACHINE LEARNING APPROACHES

1. Neural network

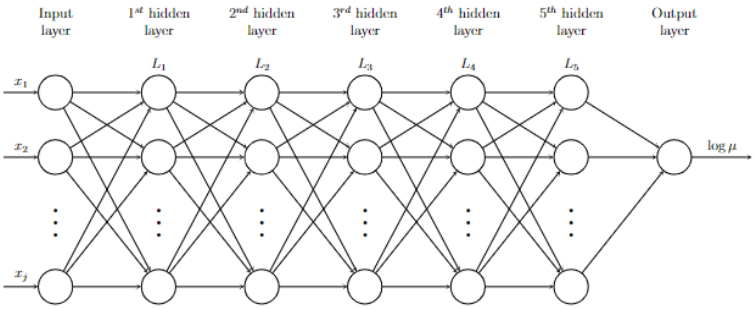
Neural networks refer to a system whose design is originally schematically inspired by the functioning of neurons in the human brain. Neural networks aim to approximate nonlinear functions defined on finite-dimensional space, relying on the composition of layers of simple functions. The relevance of neural networks comes from the universal approximation theorem and the Kolmogorov-Arnold representation theorem. Here, a feedforward neural network is considered, also called multilayered perceptron.

The following input variables are considered: year, age, cohort, sex and country.

Among all the architectures that have been tested by [3], the DEEP6 architecture is the one that gives the best results. This architecture has in particular the following characteristics:

- Five hidden layers
- A tanh activation function
- A sigmoid output function
- A combination of Dropout and BatchNorm Regularization layers in order to avoid the phenomenon of overfitting
- Two skip connections which connect the output layer to the last intermediate hidden layer, as well as to the first hidden layer.

FIGURE 2: ARCHITECTURE OF THE NEURAL NETWORK



Note: For clarity, only some neurons of the hidden layers have been shown; Dropout and BatchNorm layers are not shown.

2. Random forest

Random forests come from a bagging method applied to trees [5]. This starts from a simple observation: the trees are likely to be very similar to each other, because it is the most explanatory variables that will stand out to the detriment of the others. The idea of the model is thus beyond the random selection of the samples on which the trees are estimated, as it also randomly selects the variables involved in the model at each iteration. In particular, at each partition, the algorithm considers only m explanatory variables randomly among the p explanatory variables of the problem. Often, the order of magnitude of m is given by the relation $m \approx \sqrt{p}$. That means each basic model is less efficient, but the aggregation of very diverse models produces relevant results.

The following input variables are considered: year, age, cohort and sex.

For each of the countries, a random forest with two populations (Men and Women) is calibrated on the training sample. As this procedure has a high computational cost, the number of trees is limited to 300. This choice makes it possible to guarantee both an adequate percentage of variance explained by the model and a low average of the squared residuals.

COMPARISON OF MODEL PERFORMANCES

1. Error metric choice

In [3], the authors choose to use the mean of squared errors (MSE) on the estimated mortality force μ as the error metric. As the output variable in the baseline mortality models considered is $\log \mu$ and not μ, it seems more natural to us to consider as a metric the MSE error on $\log \mu$, rather than on μ. Besides, by choosing this metric, the errors of the points with high mortality forces in absolute value (which correspond to high ages) do not have a predominant impact on the calculated error.

2. Results

The table in Figure 3 depicts the results obtained and shows that, for Split 1 and Split 2, the errors obtained with the machine learning approaches are significantly lower than those obtained with the baseline models. In particular, the random forest approach provides a significant error gain compared to the two baseline mortality models.

Thus, although these algorithms are more difficult to implement and interpret, they manage to learn from data and project future mortality rates for HMD countries with a significantly higher degree of accuracy than classic mortality models.

FIGURE 3: COMPARISON OF ERROR RESULTS OBTAINED BY THE MODELS; ERROR VALUES ARE MULTIPLIED BY 10^3

MODEL	SPLIT 1	SPLIT 2
Age-Period-Cohort	112	146
Lee-Carter	84	152
Neural Network	77	127
Random Forest	64	108

Conclusion

Using machine learning algorithms can significantly improve the forecasts of future mortality rates when compared to traditional mortality models such as the Lee-Carter model or the Age-Period-Cohort model. Using such techniques can further help to improve mortality risk modelling and refine best estimate mortality tables.

However, concerns remain regarding the lack of interpretability of machine learning algorithms in general. The companion Milliman paper by Postema & van Es proposes a method that relieves some of those concerns by using a Temporal Fusion Transformer (TFT) model on multi-population, age-specific mortality data that has been enriched with socioeconomic data collected by the World Bank.

REFERENCES

[1] Deprez, P., Shevchenko, P. V., and Wwüthrich, M. V. (2017). Machine learning techniques for mortality modeling. European Actuarial Journal, 7(2):337–352.

[2] Hainaut, D. (2018). A neural-network analyzer for mortality forecast. ASTIN Bulletin: The Journal of the IAA, 48(2):481–508.

[3] Richman, R., & Wüthrich, M. V. (2021). A neural network extension of the Lee–Carter model to multiple populations. Annals of Actuarial Science, 15(2), 346-366.

[4] Jan Thiemen Postema & Raymond van Es (2022). Improved mortality rate forecasting using machine learning and open data. https://www.milliman.com/en/insight/Improved-mortality-rate-using-ML-and-open-data.

[5] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

[6] Kolmogorov, Andreĭ Nikolaevich. The representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables. Doklady Akademii Nauk SSSR 108.2 (1956): 179-182.

[7] Pankratz, A. (2009). Forecasting with univariate Box-Jenkins models: Concepts and cases, volume 224. John Wiley & Sons.

[8] HMD (2018). The Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). See https://www.mortality.org.

Milliman

Milliman is among the world's largest providers of actuarial and related products and services. The firm has consulting practices in life insurance and financial services, property & casualty insurance, healthcare, and employee benefits. Founded in 1947, Milliman is an independent firm with offices in major cities around the globe.

milliman.com

CONTACT

Alexandre Boumezoued

Alexandre.Boumezoued@milliman.com

Amal Elfassihi

Amal.Elfassihi@milliman.com