From Life Expectancy at Birth to Life Table. A Novel Approach

Andrea Nigri, Susanna Levantesi, José Manuel Aburto

August 26, 2020

Keywords: Life expectancy, Forecasting, Mortality rates, Deep Neural Network.

1 Introduction

The rise in human longevity over the last two centuries has led to a growing interest in modelling and predicting death rates and life expectancy. Reliable estimates of age-specific mortality are essential in the study of health inequalities and well-being between and within countries, however this task comes with several difficulties including lack of reliable data or stochastic variation in death counts. Because of these difficulties in several countries and sub-populations, the regularities observed in past and present trends of life expectancy have made this indicator appealing to model and predict using multiple approaches. A key advantage of modelling life expectancy at birth, or at any age, is that the predictive model deals only with a single indicator over time that summarizes the overall level of mortality, instead of modelling single time series of death rates for each age simultaneously (e.g. Lee-Carter). These approaches consider the continuous increase in life expectancy in several countries (HumanMortalityDatabase (2018), Oeppen and Vaupel (2002)), how improvements have varied over time Vallin and Meslé (2009), and the fact that females tend to reach higher ages than males (**Add Marc Luy's paper here**). However, recent patterns of stalls in mortality improvements and stagnation Ho and Hendi (2018), or temporary reversals (Aburto et al. (2016); García and Aburto (2019)), in life expectancy observed in several countries around the globe, including the USA (**add a citation**), has made more challenging to accurately model and predict life expectancy changes.

Among the mortality forecasting methods, extrapolation is the most common in demographic literature. A collection of mortality models (e.g., the Lee-Carter model (Lee and Carter, 1992) and its extensions, and the Cairns-Blake-Dowd model (Cairns et al., 2006)) work on the matrix of age and period death rates extrapolating the identified time-dependent and cohort-dependent parameters through a time series model.

Another approach to project mortality typically aims at forecasting the age-at-death distribution, as for example Oeppen (2008) and Bergeron-Boucher et al. (2017) whom use the Lee-Carter model to project life table distribution of deaths instead of the logarithm of death rates, Pascariu et al. (2019) whom introduced the forecast of the statistical moments of death distribution and Basellini and Camarda (2019) whom propose a relational model by transforming the age axis with a smooth non-linear warping function. It is worth emphasizing that, basically, none of these models can be generalized on a wide-global dataset over age and time. Some of them adequately work in a subset of countries, some others on a subset of specific age-ranges, and so on. Cairns et al. (2008) proposed a set of criteria that a good mortality model should meet. They refer to good-practice guidelines, such as the consistency with historical data, the long-term dynamics biologically reasonable, but also to statistical and computational features: a

good mortality model should be straightforward to implement and parsimonious. These latter features seem to advantage the univariate time series forecasting, such as the life expectancy at birth, that is easy to compute and interpret.

This last approach directly works on the univariate series of life expectancy (or other summary indicators such as lifespan disparity or entropy). The main contributions are from Lee (2006), who considers a stochastic behavior of changes in the life expectancy trend, assuming that the average changes are functions of the gap with the best-practice life expectancy (BPLE), that is the maximum life expectancy observed among national populations. Similarly, Torri and Vaupel (2012) bring forward a Geometric Brownian motion-based model, overcoming the main limitation of Lee's approach in which future life expectancy can exceed the level of BPLE. Pascariu et al. (2018) take up the forecasting approaches based on the BPLE gap by using the gap between country female life expectancy and BPLE for women, and the gap with female life expectancy in the same country for men. Raftery et al. (2013) introduce the idea of forecasting life expectancy using a two-sex model, developing a Bayesian hierarchical model to obtain joint probabilistic projections of life expectancy for both genders. Finally, Nigri et al. (ress) propose a new approach for forecasting life expectancy and lifespan disparity based on the recurrent neural networks with a long short term memory.

Although the use of life expectancy as a fundamental indicator of health and longevity is widespread, estimating age-specific death rates is needed for analyzing divergence in patterns of mortality at different ages, calculating other mortality indicators such as years of life lost or lifespan inequality, and for insurance pricing and pension liabilities.

The aim of this study is to formulate a model leveraging on deep learning algorithms based on neural networks to derive age-specific mortality profiles from observed or predicted life expectancy levels. Resulting estimates would be useful for guiding public health interventions, informing about age-specific mortality dynamics in contexts with deficient data collection, as well as pension and social security schemes which rely on longevity dynamics. In this article, we focus on developing the "indirect" methodology to derive age- and sex- specific death rates over time from life expectancy levels with a deep neural network framework. However, the model is flexible to be used with different summary measures such as lifespan inequality.

Demography has a long standing tradition of developing formal demographic methods and using statistical approaches to indirectly estimating indicators. According to the UN manual, the term "indirect" qualifies the demographic estimation technique that origins in the fact that such technique produces estimates of certain parameters on the basis of information that is only indirectly related to its value[1]. Indirect methods were developed by different systems of model life tables (United Nations (1955);United Nations (1967); Sullivan (1972)); Brass' relational model (Brass et al. (1968);Brass (1971)); Murray et al. (2003); Wilmoth et al. (2012); Mayhew and Smith (2013). Our work is also related to the recent improvements of forecasting techniques developed by Ševčíková et al. (2016) and Pascariu et al. (2017) whom, adopting a method based on the Lee-Carter model, exploit an inverse approach for converting life expectancy forecasting into age-specific death rates.

In the field of age-specific mortality forecasting from summary measures, the need for accurate and reliable estimates is yet to be met. The most recent and prominent contributions, Ševčíková et al. (2016) and Pascariu et al. (2017), suffer from some structural drawbacks. Both approaches are based on the log-linear assumption, specifically, the estimation of mortality rates from life expectancy at birth can be accomplished by exploiting the regularities of age patterns of

---

[1]The classic example is the use of the proportion of children dead among those ever borne by women aged 20-24 years to estimate the probability of dying before age 2. The observed proportion of children dead is clearly related to mortality conditions but it is not a pure mortality measure because it is affected by other non-mortality parameters.

mortality. This kind of regularity might not hold for particular ages. The Pascariu et al. (2017) method based on SVD estimation provides systemic divergences, underestimating from age 70 onwards and overestimating infant mortality, where the relationships between mortality rates in logarithmic scale and life expectancy at birth provide a flat correlation. Pascariu et al. (2017) found that the optimal number of years to be used in the fitting of the model is between 30 and 35 years. On the other hand, a short period may provide hight jumps in parameter estimation, that force authors to compensate using another estimation step including the parameter noise smoothing. In this regard, a measure of the improvements provides by the smoothing is neglected. Furthermore, some parameters have not a demographic interpretation, such as the k in Pascariu et al. (2017) that is just an additional step to optimize estimation. Finally, the so-called rotation of the β parameter in the spirit of Li et al. (2013) implies additional modeling with arbitrary thresholds into the Ševčíková et al. (2016) and Pascariu et al. (2017) frameworks. All these aspects might be overcome by using a fully non-parametric approach such as a deep neural network model. It provides a completely data-driven approach, allowing at investigate the hidden relationship among input-output data, by avoiding an a-priori probabilistic assumption.

The scope of our contribution is to derive mortality surface from life expectancy at birth using a deep neural network model. In the numerical application, we use data from Human Mortality Data Base (HumanMortalityDatabase (2018)). The model has been performed for Italy, USA, Australia, France and Japan, from 1950 to 2014 and for both sexes. Our results are then compared with the projected mortality rates obtained from the Lee-Carter model (Lee and Carter, 1992) and the Linear Link Pascariu et al. (2017) model in both versions (SVD and Poisson). The paper is structured as follows. Section 2 describes the model based on the deep neural networks. Section 3 provides the results, while Section 4 concludes.

2   Model

The term neural network (NN) originated as a mathematical model that replicates the biological neural networks of the human brain. NN architecture includes neurons, synaptic connections that link the neurons, and learning algorithms. Typically, NN is formed by three types of layers called input, hidden and output layer and each one has several neurons. Each unit in a network gets "weighted" information through synaptic links from the other connected ones and returns an output by using an activation function transforming the weighted sum of input signals. Aiming at creating a bridge between Deep Neural Network (DNN) and demography, we will describe the steps to obtain the target output. Equation 1 describes the specific NN structure providing the logarithm of mortality surface $\mathbf{M_{(a,t)}}$ with $\mathbf{a} \in \{0, 1, ..., 100\}$ vector of ages and $\mathbf{t} \in \{t_1, t_2, ..., t_n\}$ vector of years:

$$log[\mathbf{M_{(a,t)}}] = f^k \left(\begin{bmatrix} w_{1,1}^{(k)} & w_{1,2}^{(k)} & w_{1,3}^{(k)} & \cdots & w_{1,n}^{(k)} \\ w_{2,1}^{(k)} & w_{2,2}^{(k)} & w_{2,3}^{(k)} & \cdots & w_{2,n}^{(k)} \\ w_{3,1}^{(k)} & w_{3,2}^{(k)} & w_{3,3}^{(k)} & \cdots & w_{3,n}^{(k)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{n,1}^{(k)} & w_{n,2}^{(k)} & w_{n,3}^{(k)} & \cdots & w_{n,n}^{(k)} \end{bmatrix} \begin{bmatrix} H_1^{(k-1)} \\ H_2^{(k-1)} \\ H_3^{(k-1)} \\ \vdots \\ H_n^{(k-1)} \end{bmatrix} + \begin{bmatrix} b_1^k \\ b_2^k \\ b_3^k \\ \vdots \\ b_n^k \end{bmatrix} \right) \tag{1}$$

where, for a generic layer k, $f^{(k)}$ is the activation function, $W^{(k)}$ the weights matrix, $H^{(k)}$ the hidden layers, and $b^{(k)}$ the bias, used to control the triggering value of the activation function. A graphical representation of the DNN model related to eq. 1 is given in Fig. 1.
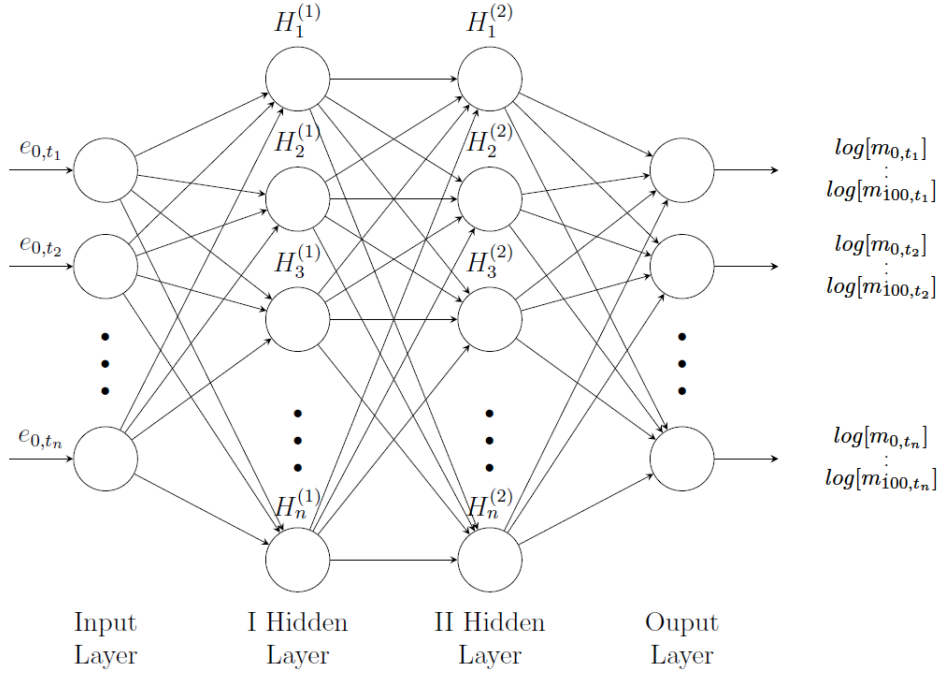
Figure 1: Graphical representation of our DNN model. Circles represent neurons, and lines synapses. Synapses take the input and multiply it by a weight (the "strength" of the input in determining the output). Neurons add the outputs from all synapses and apply an activation function.

In the following, we will describe the concepts from DNN in a demographic view, by defining the output as a function of the DNN structure and the input data which is the time series of life expectancy at birth. The theoretical relationship defining the matrix of mortality surface $log[\mathbf{M_{(a,t)}}]$ given the vector of life expectancy at birth at year t, and the DNN neurons processing is represented by:

$$log[\mathbf{M_{(a,t)}}] = f^{(k)}\big(\mathbf{W^{(k)}}\, f^{(k-1)}\big(\ldots f^{(1)}\big(\mathbf{W^1}\, \mathbf{e_{0,t}} + \mathbf{b^{(1)}}\big)\ldots\big) \tag{2}$$

Where $f^{(1)}\big(\mathbf{W^1}\, \mathbf{e_{0,t}} + \mathbf{b^{(1)}}\big) = \mathbf{H^{(1)}}$ is the first hidden layer that accepts the vector of life expectancy at birth, $\mathbf{e_{0,t}} = e_{0,t_1}, e_{0,t_2}, ..., e_{0,t_n}$, as input. After the estimation procedure, which implies to estimate the weights, we obtain the final output, $log[\mathbf{\hat{M}_{(a,t)}}]$, that is the mortality surface resulting from the application of the DNN parameters (weights matrix $\mathbf{\hat{W}}$ and bias $\mathbf{\hat{b}}$) obtained by the optimization procedure described in the following.

The estimation of the network parameters is obtained by the minimization of the overall loss function $\mathcal{L}$. We select the Mean Square Error (MSE) as loss function:

$$\mathcal{L}(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Where y is $log[\mathbf{M_{(a,t)}}]$, $\hat{y}$ is $log[\mathbf{\hat{M}_{(a,t)}}]$ and n is the number of observations. It is worth noting that the loss function implicitly involves the life expectancy function as shown in eq. 2. To minimize the loss function, we use the Gradient Descent optimization algorithm, which iteratively moves in the direction of steepest descent as defined by the negative of the gradient. This algorithm proceeds by minimizing $\mathcal{L}$ at each step t, therefore differentiating the loss function

with respect to the weights ($\mathbf{W}$):

$$\nabla\mathcal{L}(log[\mathbf{M_{(a,t)}}], log[\hat{\mathbf{M}}_{(\mathbf{a,t})}]) = \begin{bmatrix} \frac{\partial\mathcal{L}(log[\mathbf{M_{(a,t)}}], log[\hat{\mathbf{M}}_{(\mathbf{a,t})}])}{\partial w_{1,1}^{(1)}} \\ \frac{\partial\mathcal{L}log[\mathbf{M_{(a,t)}}], log[\hat{\mathbf{M}}_{(\mathbf{a,t})}])}{\partial w_{1,2}^{(1)}} \\ \frac{\partial\mathcal{L}log[\mathbf{M_{(a,t)}}], log[\hat{\mathbf{M}}_{(\mathbf{a,t})}])}{\partial w_{1,3}^{(1)}} \\ \vdots \\ \frac{\partial\mathcal{L}(log[\mathbf{M_{(a,t)}}], log[\hat{\mathbf{M}}_{(\mathbf{a,t})}])}{\partial w_{n,n}^{(k)}} \end{bmatrix} \quad (3)$$

For the generic weights $w_{n,n}^{(k)}$ and the k-th layer, the algorithm proceeds using the chain derivation rule described in the following equation:

$$\frac{\partial\mathcal{L}(log[\mathbf{M_{(a,t)}}], log[\hat{\mathbf{M}}_{(\mathbf{a,t})}])}{\partial w_{n,n}^{(k)}} = \frac{\partial\mathcal{L}(log[\mathbf{M_{(a,t)}}], log[\hat{\mathbf{M}}_{(\mathbf{a,t})}])}{\partial H_n^{(k)}} \frac{\partial H_n^{(k)}}{\partial z_n^{(k)}} \frac{\partial z_n^{(k)}}{\partial w_{n,n}^{(k)}} \quad (4)$$

where $z_n^{(k)} = w_n^{(k)} H_n^{(k-1)} + b_n^{(k)}$.

To update the weights ($\tilde{\mathbf{W}}$), the gradient of the loss function, $\nabla\mathcal{L}_t(log[\mathbf{M_{(a,t)}}], log[\hat{\mathbf{M}}_{(\mathbf{a,t})}])$, is multiplied by a scalar, η, often called learning rate, according to the following scheme:

$$\tilde{\mathbf{W}} = \mathbf{W} - \eta\nabla\mathcal{L}_t(log[\mathbf{M_{(a,t)}}], log[\hat{\mathbf{M}}_{(\mathbf{a,t})}]) \quad (5)$$

In a figurative way, the idea behind the gradient descent is similar to "climbing down a hill" until a global or local minimum is reached. At each update, the search moves in the opposite direction of the gradient and the learning rate η determines the amplitude of this movement, controlling the adjustment in the weights, thus determining how fast or slow we will move towards the optimal weights. A very large learning rate leads to a sub-optimal solution. A very small learning rate involves too many iterations to find the optimal solution. Then, the learning rate can be considered the most important hyper-parameter for tuning NN.

The search for the optimal parameters is then carried out through an optimization process. The NN initial weights are selected in an arbitrary (random) way so they are not optimal parameters. The iterations of the algorithm lead to the optimization of the weights and minimization of the error. Many algorithms have been developed to solve the problem concerning the magnitude of the gradients. We use the **RMSProp** (Root Mean Square Propagation) algorithm, proposed by Hinton et al. (2012), which uses a moving average of squared gradients to normalize the gradient. Differently from other Gradient Descent algorithms, **RMSProp** uses an adaptive learning rate instead of treating the learning rate as a hyperparameter. This means that the learning rate changes over time.

Thus, we will implement a preliminary fine-tuning to identify the optimal hyperparameters: epochs[2] and neurons number for each hidden layer.

The choices concerning the type of architecture (e.g., the number of hidden layers, units for each layer) and the hyperparameter (e.g., learning rate, activation functions, and epochs) value remains a heuristic problem for NN users: the choice often depends on the type of data and it is not an easy step. A preliminary round of hyperparameters tuning, before the testing, is highly desired.

---

[2]The number of times algorithm explores the entire training dataset.

3 Other models

In this paper, the performance of the proposed model is compared to the Lee-Carter, a well-established approach considered as the reference model for the forecast of mortality. Furthermore, nowadays in the demographic scenario the indirect estimation techniques increase their appeal, offering a remarkable forecasting accuracy in lockstep with a higher level of parsimony. As a benchmark, we introduce a further comparison with the Linear-Link (LL) model proposed by Pascariu et al. (2017), which is one of the recent and most prominent approaches to reconstruct vital rates from the demographic summary measures.

3.1 Lee-Carter model

We present the Lee-Carter (LC) model in the Poisson framework as proposed by Brouhns et al. (2002):

$$\log\left(\mathbf{M_{a,t}}\right) = \boldsymbol{\alpha}_a + \boldsymbol{\beta}_a \boldsymbol{\kappa}_t, \tag{6}$$

Where $\boldsymbol{\alpha}_a$ is the age-specific parameter providing the average age profile of mortality; $\boldsymbol{\beta}_a \cdot \boldsymbol{\kappa_t}$ is the age-period term describing the mortality trends ($\boldsymbol{\kappa}_t$ is the time index and $\boldsymbol{\beta}_a$ modifies the effect of $\boldsymbol{\kappa}_t$ across ages). The following constraints on $\boldsymbol{\kappa}_t$ and $\boldsymbol{\beta}_a$ avoid identifiability problems with the parameters:

$$\sum_{t \in \mathcal{T}} \kappa_t = 0 \quad \sum_{a \in \mathcal{A}} \beta_a = 1$$

Mortality forecasting is obtained by modeling the time index κ_t by an autoregressive integrated moving average (ARIMA) process. In general, a random walk with drift properly fits the data:

$$\kappa_t = \kappa_{t-1} + \delta + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_k^2) \tag{7}$$

where δ is the drift parameter and ϵ_t are the error terms, normally distributed with null mean and variance σ_k^2.

3.2 Linear-Link model

Linear-Link model combines strong linear relations comparing life expectancies and age-specific death rates on a log-log scale. This method is implemented in a publicly available R package. Given a life expectancy at birth level, the Linear-Link model derives mortality pattern using a linear relation. Thus, the logarithm age-specific mortality rate at time t and age a, $m_{a,t}$, can be expressed as a linear function of the logarithm of life expectancy at birth, $e_{0,t}$. Formally:

$$\log\left(\mathbf{M_{a,t}}\right) = \beta_a \log(e_{0,t}) + \nu_a k + \varepsilon_{a,t} \quad \text{for} \quad a \geq 0$$

$$\sum_{a=0}^{\omega} \nu_a = 1, \quad \text{and} \quad \nu_a \geq 0 \tag{8}$$

Where ω is the highest attainable age. The canonical form of the Linear-Link model is based on the least squares estimation of the slope β_a of the linear relation between life expectancy at birth and the mortality rates over the observation time t. Thereby, β_a can be regarded as an age-specific parameter and $\varepsilon_{a,t}$ denotes a set of normally distributed errors with mean zero and variance σ^2:

$$\sum_a \left[\log(m_{a,t}) - \beta_a \log(e_{0,t})\right]^2 = \sum_a \left[\varepsilon_{a,t}\right]^2 \tag{9}$$

The model specification involves a second step estimation, by computing the singular value decomposition (SVD) of the matrix of regression residuals, obtaining the ν_a parameter estimation.

In order to avoid projecting age-specific noise in the jump-off life table a smoothing of the β_a and ν_a parameters using splines is performed. Finally, the parameter k optimizes the mortality curve given in the previous step, where the difference between target life expectancy e_0^* and an estimated life expectancy e_0 is below a given level, 0.001. Where e_0 represents the level of life expectancy at birth based on the initial mortality rates $\boldsymbol{M_{a,t}^* = \exp\left\{\beta_a \log(e_{0,t}^*) + \nu_a k\right\}}$, where $\boldsymbol{k} = 0$. Alternatively, the parameters of the model can be estimated by assuming that deaths follow a Poisson distribution. In the present analysis, we consider both canonical Linear-Link (LL) and Poisson Linear Link model (LL Pois).

4 Numerical Application

In the numerical application, we consider historical mortality data collected by the HumanMortalityDatabase (2018) for Australia, Italy, Japan, Russia and USA, by gender. We will carry out an out-of-sample test considering the same period for all countries: 1950-2014. The time frame 1950-2005 is used as train-test chunk, according to a common splitting rule 80%-20%. While, the remaining time frame 2006-2014 is used for the model validation.

In order to assess the models' accuracy, we calculate the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) on the validation period 2006-2014.

$$\text{MAE}: \quad \sum^{n} \frac{\mid m_{a,t} - \hat{m}_{a,t} \mid}{n}, \tag{10}$$

$$\text{RMSE}: \quad \sqrt{\frac{\sum^n (m_{a,t} - \hat{m}_{a,t})^2}{n}}. \tag{11}$$

To be applied, our model requires that data be arranged in a *year* × *age* matrix Y and vector X. The X vector is composed of input data, that in our case study is the time series of life expectancy at age 0. The Y matrix contains the target output i.e. the mortality rates by age and year. The model learns the hidden pattern in the input data in a given year and gives back the output in the same year.

To obtain the train-test split and avoiding the time dependence, years are randomly sampled. The train-test we perform is slightly different from a canonical one because the test vector is selected without using a time-dependent selection. Thus, in the train set, we may obtain for instance the years 1980, 1995, 1970, 2000. Conversely, in the test vector, we may observe for instance the years 2010, 1985, 1999, 1997, 1975, 2001. Clearly, the train and test set do not show a time-dependent pattern. After the tuning procedure, for each country separately, we use the architectures with six hidden layers with five hundred neurons per layer. We choose the Rectified Linear Unit (ReLU) activation functionthat outperformed the other functions tested for all countries.

4.1 Results

The analysis includes numerical and graphical representation of the goodness of fit. The following figures show the relative difference of the logarithm of central death rates between our model (DNN) and the observed values. For comparison, we also shows the relative difference for the LC model. For a given model i, the relative difference is defined as:

$$\Delta_{a,t}^i = \frac{log(\hat{m}_{a,t}^i) - log(m_{a,t})}{log(m_{a,t})}$$

In Fig. 2 and 3 we show the heat maps (for the Italian case) which are useful for appreciating the models' differences. The DNN approach shows very high accuracy performances thanks

to the deep architecture, which allows for investigating the hidden relationship between life expectancy at birth and mortality surface over time.
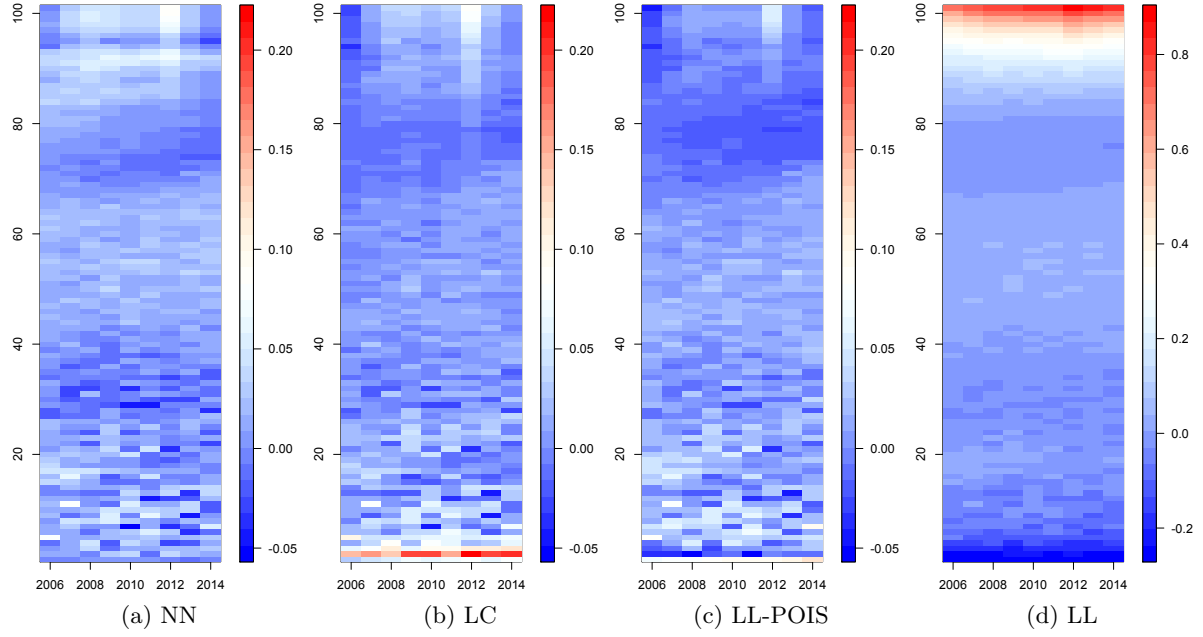

(a) NN   (b) LC   (c) LL-POIS   (d) LL

Figure 2: $\Delta_{a,t}^{i}$ - Italy, females

(a) NN   (b) LC   (c) LL-POIS   (d) LL

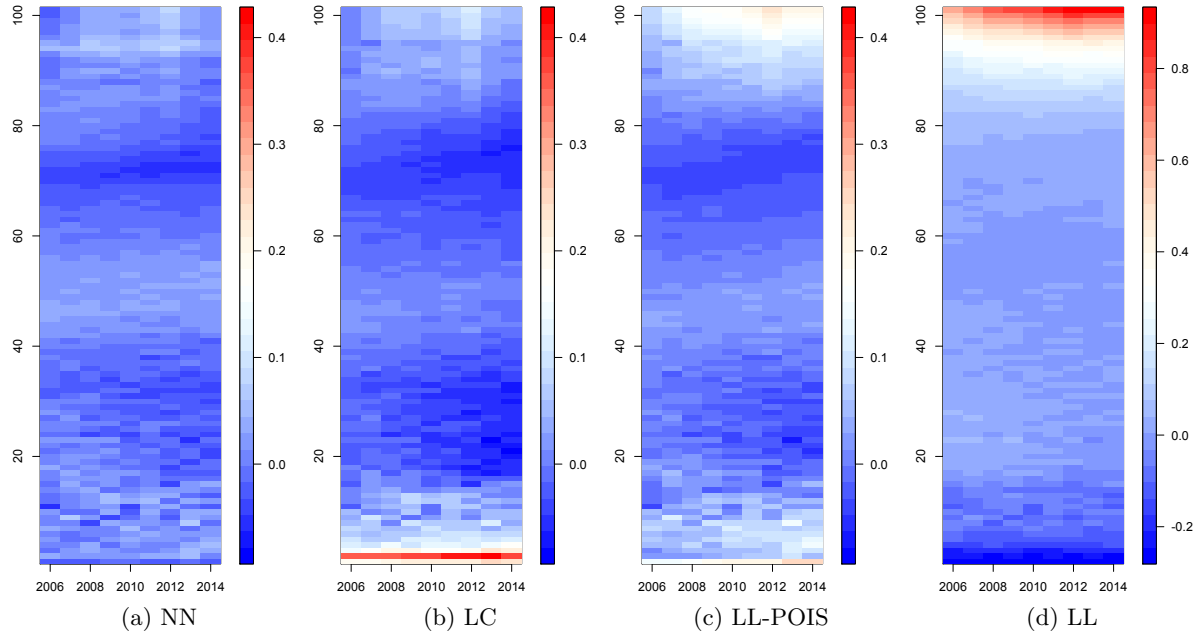Figure 3: $\Delta_{a,t}^{i}$ - Italy, males

The logarithm of mortality surface is shown in Fig. 4 for the Italian males, year 2010.

```
Italy Male, year:2010
Red: NN, Black: L-C, Green: LL Pois, Blue: LL
```
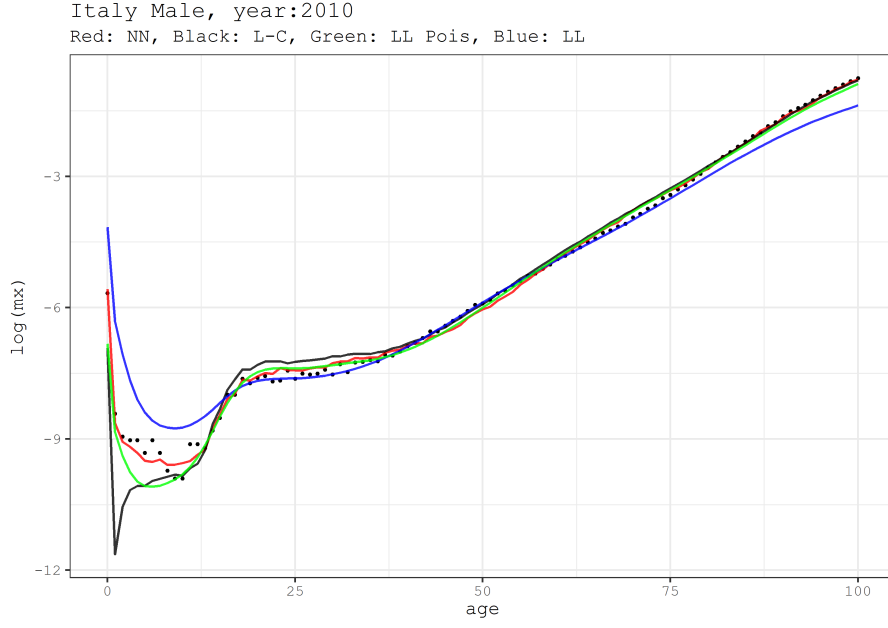
Figure 4: $log(m_a)$ - Italy, males, year 2010.

The values of RMSE and MAE on the validation period 2006-2014 are reported in Table 1.

Table 1: Performance of DNN the validation set for each country.

Country	Female		Male	
Australia	*MAE*	*RMSE*	*MAE*	*RMSE*
Lee-Carter	0.1381	0.2080	0.1734	0.2345
LL	0.3101	0.4688	0.4679	0.5820
LL Pois	0.1517	0.2225	0.1633	0.2164
DNN	**0.1295**	**0.1984**	**0.1385**	**0.1742**
France	*MAE*	*RMSE*	*MAE*	*RMSE*
Lee-Carter	0.1310	0.1905	0.1644	0.2504
LL	0.2765	0.3911	0.2707	0.3967
LL Pois	0.1220	0.1593	0.1047	0.1579
DNN	**0.1041**	**0.1379**	**0.0914**	**0.1208**
Italy	*MAE*	*RMSE*	*MAE*	*RMSE*
Lee-Carter	0.1128	0.2308	0.2286	0.4685
LL	0.2939	0.4659	0.2760	0.4760
LL Pois	0.1040	0.1586	0.1657	0.2667
DNN	**0.0986**	**0.1368**	**0.1063**	**0.1403**
Japan	*MAE*	*RMSE*	*MAE*	*RMSE*
Lee-Carter	0.5164	0.7009	0.1828	0.2886
LL	0.2937	0.4338	0.2649	0.4490
LL Pois	0.4271	0.5481	**0.1532**	**0.2200**
DNN	**0.1302**	**0.1771**	0.1752	0.2338
USA	*MAE*	*RMSE*	*MAE*	*RMSE*
Lee-Carter	**0.0987**	0.1330	0.10363	0.1375
LL	0.2062	0.3223	0.2909	0.3847
LL Pois	0.1067	0.1535	0.0969	0.1240
DNN	0.1000	**0.1276**	**0.0964**	**0.1150**

5 Conclusions

Refined tools to forecast future mortality surface are becoming common among life insurers, pension plans, and social security schemes that have to manage future cash flows depending on longevity dynamics. The estimation of future mortality rates still plays a central role in the insurance industry and national public health systems. Our model is applied to estimating mortality rates using life expectancy or future target life expectancy. It can be also involved in projected life tables for countries with deficient data and for historical life table construction. Therefore, our model has a high degree of generalization which makes it easily applicable to different situations.

References

Aburto, J. M., Beltrán-Sánchez, H., García-Guerrero, V. M., and Canudas-Romo, V. (2016). Homicides in mexico reversed life expectancy gains for men and slowed them for women, 2000–10. *Health Affairs*, 35(1):88–95.

Basellini, U. and Camarda, C. G. (2019). Modelling and forecasting adult age-at-death distributions. *Population studies*, 73(1):119–138.

Bergeron-Boucher, M.-P., Canudas-Romo, V., Oeppen, J., and Vaupel, J. W. (2017). Coherent forecasts of mortality with compositional data analysis. *Demographic Research*, 37:527–566.

Brass, W. (1971). *Biological aspects of demography*. Society for the Study of Human Biology Symposium Series. Taylor & Francis.

Brass, W., Coale, A., Demeny, P., Heisel, D., Lorimer, A., Romaniuk, A., and Van de Walle, E. (1968). *The Demography of Tropical Africa*. Princeton University Press.

Brouhns, N., Denuit, M., and Vermunt, J. (2002). A poisson log-bilinear approach to the construction of projected life tables. *Insurance: Mathematics and Economics*, 31:373–393.

Cairns, A., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, 73:687–718.

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

García, J. and Aburto, J. M. (2019). The impact of violence on Venezuelan life expectancy and lifespan inequality. *International Journal of Epidemiology*, 48(5):1593–1601.

Hinton, G., Srivastava, N., and Swersky, K. (2012). Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. *Lecture Notes Distributed in CSC321 of University of Toronto*.

Ho, J. and Hendi, A. (2018). Recent trends in life expectancy across high income countries: retrospective observational study. *BMJ*, 362(k2562).

HumanMortalityDatabase (2018). University of california, berkeley (usa), and max planck institute for demographic research (germany). Data downloaded on 01/01/2019.

Lee, R. (2006). Mortality forecasts and linear life expectancy trends. perspectives on mortality forecasting. the linear rise in life expectancy: History and prospects. *Social Insurance Studies*, 3.

Lee, R. and Carter, L. (1992). Modeling and forecasting us mortality. *Journal of the American Statistical Association*, 87:659–671.

Li, N., Lee, R., and Gerland, P. (2013). Extending the lee-carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography*, 50(6):2037–2051.

Mayhew, L. and Smith, D. (2013). A new method of projecting populations based on trends in life expectancy and survival. *Population Studies*, 67(2):157–170.

Murray, C. J., Ferguson, B. D., Lopez, A. D., Guillot, M., Salomon, J. A., and Ahmad, O. (2003). Modified logit life table system: principles, empirical validation, and application. *Population Studies*, 57(2):165–182.

Nigri, A., Levantesi, S., and Marino, M. (in press). Life expectancy and lifespan inequality forecasting. a deep learning approach. *Scandinavian Actuarial Journal*.

Oeppen, J. (2008). Coherent forecasting of multiple-decrement life tables: a test using japanese cause of death data. *Compositional Data Analysis Conference*.

Oeppen, J. and Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296(5570):1029–1031.

Pascariu, M. D., Basellini, U., Aburto, J. M., and Canudas-Romo, V. (2017). The linear link: Deriving age-specific death rates from life expectancy.

Pascariu, M. D., Canudas-Romo, V., and Vaupel, J. W. (2018). The double-gap life expectancy forecasting model. *Insurance: Mathematics and Economics*, 78:339–350.

Pascariu, M. D., Lenart, A., and Canudas-Romo, V. (2019). The maximum entropy mortality model: forecasting mortality using statistical moments. *Scandinavian Actuarial Journal*, 8:661–685.

Raftery, A. E., Chunn, J. L., Gerland, P., and Ševčíková, H. (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography*, 50(3):777–801.

Sullivan, J. M. (1972). Models for the estimation of the probability of dying between birth and exact ages of early childhood. *Population Studies*, 26(1):79–97. PMID: 22070174.

Torri, T. and Vaupel, J. W. (2012). Forecasting life expectancy in an international context. *International Journal of Forecasting*, 28(2):519–531.

United Nations (1955). *Age and Sex Patterns of Mortality: Model Life Tables for Under-Developed Countries.* Population Studies, No. 22. Department of Social Affairs.

United Nations (1967). *Methods for Estimating Basic Demographic Measures from Incomplete Data: Manuals and Methods of Estimating Populations.* Manual IV. Population Studies, No. 42. Department of Social Affairs.

Vallin, J. and Meslé, F. (2009). The segmented trend line of highest life expectancies. *Population and Development Review*, 35(1):159–187.

Ševčíková, H., Li, N., Kantorová, V., Gerland, P., and Raftery, A. E. (2016). *Age-specific mortality and fertility rates for probabilistic population projections*, volume 39. Springer Series on Demographic Methods and Population Analysis.

Wilmoth, J., Zureick, S., Canudas-Romo, V., Inoue, M., and Sawyer, C. (2012). A flexible two-dimensional mortality model for use in indirect estimation. *Population Studies*, 66(1):1–28.