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

Andrea Nigri, Susanna Levantesi, Josè Manuel Aburto

August 19, 2020

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

1 Introduction

The rise in human longevity during the 20th century leads to a growing interest in modeling and predicting both mortality rates and life expectancy. The frontiers of survival are raising without evidence of near deceleration and the hypothesis of an imminent boundary of human life has been repeatedly disproved by empirical evidence (HMD (2018), Oeppen and Vaupel (2002)). In their influential work, Oeppen and Vaupel coined the "best-practice life expectancy"(BPLE) hypothesis, consisting of the linear increase of the maximum female life expectancy at a constant pace over time since 1840. Afterwards, Vallin and Meslé (2009) proposed a segmented trend line, highlighting that life expectancy cannot keep rising at the same pace. This is also evident in some countries, such as Denmark, England and the US (Aburto et al. (2018); Hiam et al. (2018); Ho and Hendi (2018)), which are experiencing decelerations in life expectancy.

It is a widespread opinion that such constancy is fundamental in analyzing mortality change, thus increasing appeal on methods based on life expectancy extrapolation. Nevertheless the growing availability of reliable data, in lockstep with the improving of statistical-mathematical methods, has allowed the creation of ever-finer mortality forecasting models.

The mortality literature suggests three families of predictive models. The first one 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), Torri and Vaupel (2012), Raftery et al. (2013), Pascariu et al. (2018) and Nigri et al. (2019). The second one refers to extrapolative models that use the matrix of age and period mortality rates. The Lee-Carter model (Lee and Carter, 1992) and its extensions including additional age, period and cohort parameters and the Cairns-Blake-Dowd model (Cairns et al., 2006) belong to this category. The third one 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.

Although the use of life expectancy as a fundamental indicator of health and longevity is widespread, estimating mortality rates is, however, needed for calculating insurance pricing and pension liabilities. Indeed, 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. In light of that, the estimation of future mortality rates still plays a central role in the insurance industry and national public health systems, thus offering cues for further discussion among both demographers and actuaries. In fact, from a global perspective, the forecasting of summary measures (such as life expectancy) is a common and well-established practice (Miller (1986);Lee (1993)), while the reconstruction of mortality rates from those measures is still an open debate.
The literature offers very few research contributions, such as the model introduced by Ševčíková et al. (2016) which, adopting a method based on the Lee-Carter model, exploit an inverse approach for converting life expectancy forecasting into age-specific death rates using the UN's 2014 probabilistic population projections.

Our paper aims at contributing to fill the gap between forecasting life expectancy and mortality rates by proposing a novel approach, based on Neural Networks, to obtain age-specific mortality rates from a demographic summary measure. In particular, we develop a Deep Neural Network model for exploring (and learning) the relationship between life expectancy and age specific mortality rates. Such a model is then used to estimate the entire life table over time, starting from a given value of life expectancy at birth, exploiting the advantages of a univariate time series forecasting.

In the numerical application, we use data from Human Mortality Data Base (HMD (2018)). The model has been performed for Italy, USA, Australia, Russia 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. 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.

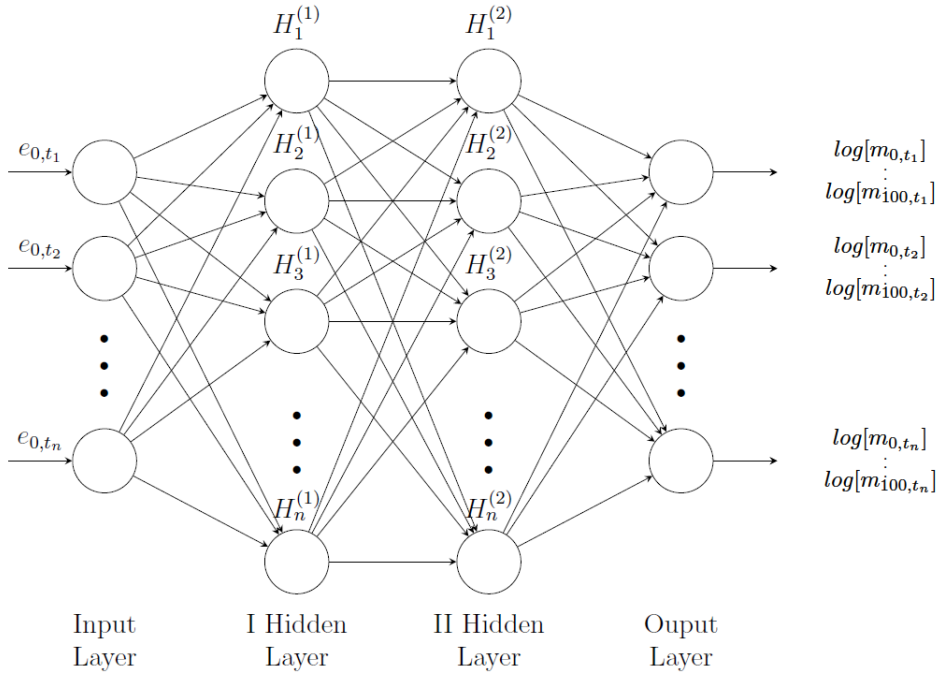


Figure 1: Graphical representation of our DNN model. Circles represent neurons, and lines synapses. Synapses take the input and multiply it by a weight (the "strength" of the input in determining the output). Neurons add the outputs from all synapses and apply an activation function.

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:

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

Where $f^{(1)}(\mathbf{W^1}\, \mathbf{e_{0,t}} + \mathbf{b^{(1)}}) = \mathbf{H^{(1)}}$ 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, $log[\hat{\mathbf{M}}_{\mathbf{(a,t)}}]$, that is the mortality surface resulting from the application of the DNN parameters (weights matrix $\hat{\mathbf{W}}$ and bias $\hat{\mathbf{b}}$) 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 $log[\hat{\mathbf{M}}_{\mathbf{(a,t)}}]$ 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 ($\mathbf{W}$):

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

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:

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

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, $\nabla\mathcal{L}_t(log[\mathbf{M}_{(\mathbf{a},\mathbf{t})}], log[\hat{\mathbf{M}}_{(\mathbf{a},\mathbf{t})}])$, is multiplied by a scalar, η, often called learning rate, according to the following scheme:

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

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[1] 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.

[1]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. (2019b), 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):

$$\mathbf{log\,(M_{a,t})} = \boldsymbol{\alpha}_{\mathrm{a}} + \boldsymbol{\beta}_{\mathrm{a}}\boldsymbol{\kappa}_{\mathrm{t}}, \tag{6}$$

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

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

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, $m_{\mathrm{a,t}}$, can be expressed as a linear function of the logarithm of life expectancy at birth, $e_{0,\mathrm{t}}$. Formally:

$$\begin{aligned} &\mathbf{log\,(M_{a,t})} = \boldsymbol{\beta}_{\mathrm{a}}\,\mathbf{log}(\boldsymbol{e}_{0,\mathrm{t}}) + \boldsymbol{\nu}_{\mathrm{a}}\boldsymbol{k} + \boldsymbol{\varepsilon}_{\mathrm{a,t}} \quad \text{for} \quad \mathrm{a} \geq \mathbf{0} \\ &\textstyle\sum_{\mathrm{a}=0}^{\omega} \nu_{\mathrm{a}} = 1, \quad \text{and} \quad \nu_{\mathrm{a}} \geq 0 \end{aligned} \tag{8}$$

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 $\varepsilon_{\mathrm{a,t}}$ denotes a set of normally distributed errors with mean zero and variance σ^2:

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

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 $\boldsymbol{M^*_{a,t}} = \mathbf{exp}\left\{\boldsymbol{\beta}_a \mathbf{log}(\boldsymbol{e^*_{0,t}}) + \boldsymbol{\nu}_a \boldsymbol{k}\right\}$, 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 HMD (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* $\times$ *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.

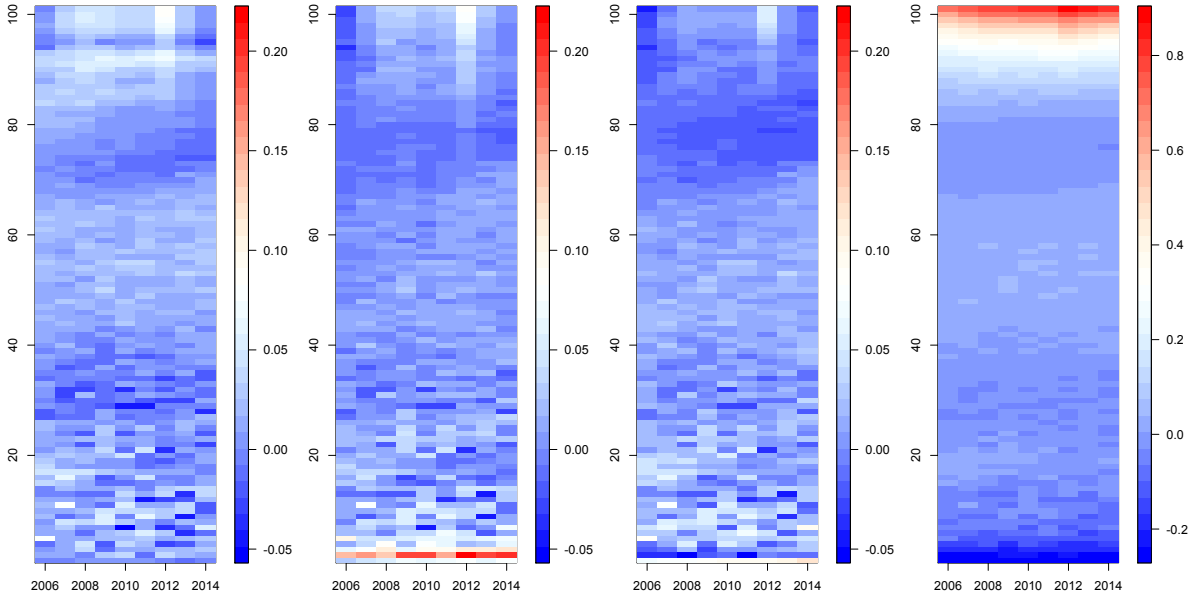
4.1 Results

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:

$$\Delta^i_{a,t} = \frac{log(\hat{m}^i_{a,t}) - log(m_{a,t})}{log(m_{a,t})}$$

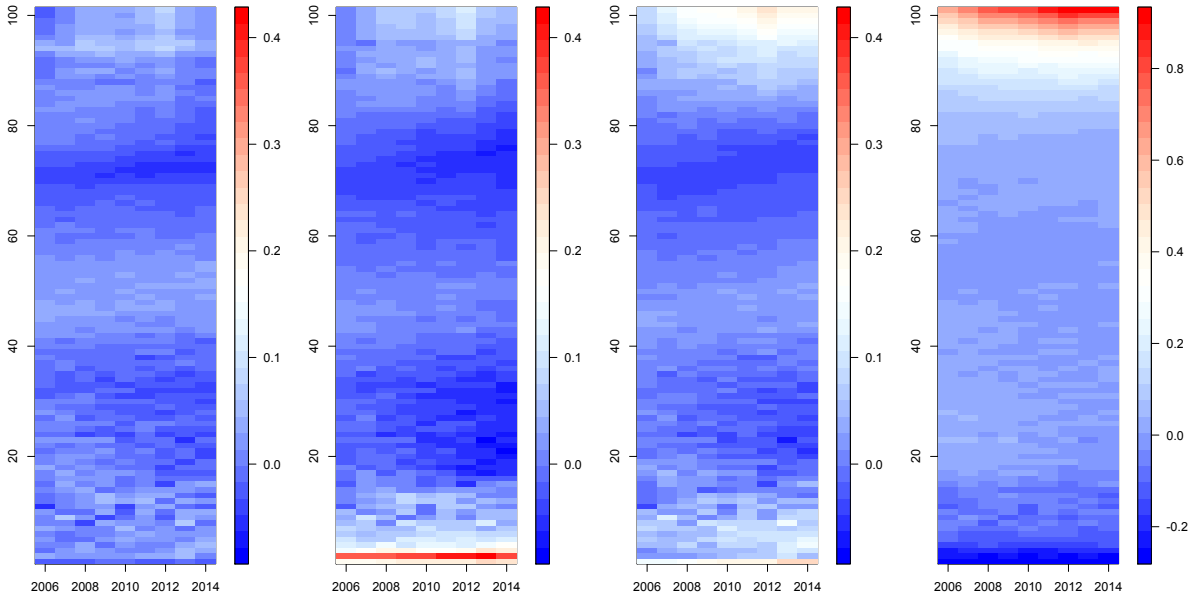
In Fig. 2 and 3 we show the heat maps (for the Italian case) which are useful for appreciating the models' differences. The DNN approach shows very high accuracy performances thanks

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: $\Delta^i_{a,t}$ - Italy, females



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

Figure 3: $\Delta^i_{a,t}$ - Italy, males

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

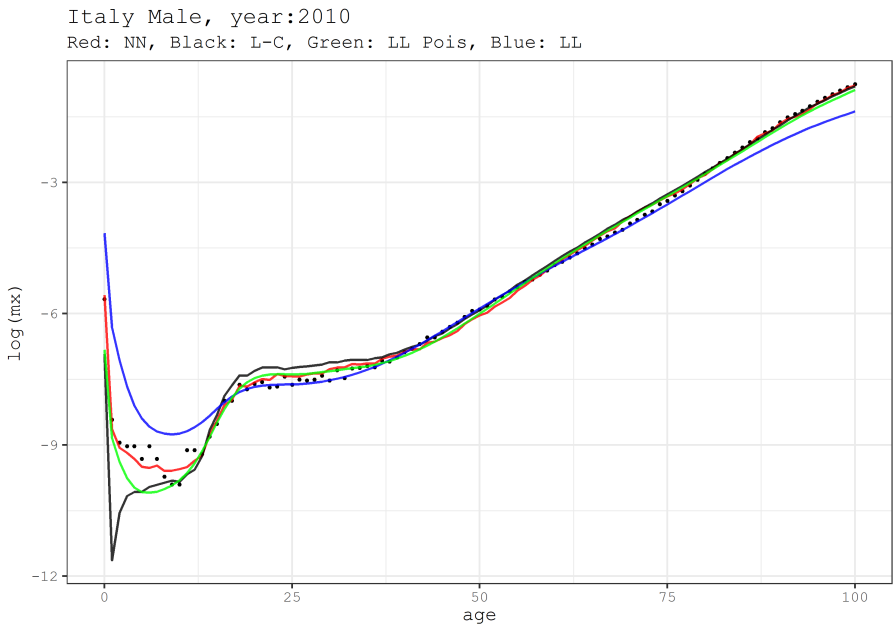


Figure 4: $log(m_a)$ - Italy, males, year 2010.

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**

4.2 Discussion

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., Wensink, M., van Raalte, A., Lindahl-Jacobsen, R. (2018). Potential gains in life expectancy by reducing inequality of lifespans in Denmark: an international comparison and cause-of-death analysis. BMC Public Health, 18(1), 831.

Basellini, U., Camarda, C.G. (2019). Modelling and forecasting adult age-at-death distributions. Popul. Stud. 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. Demogr. Res. 37, 527–566.

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

Cairns, A.J.G., Blake, D., 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.J.G., Blake, D., Dowd, K. (2008). Modelling and management of mortality risk: a review. Scand Actuar J 2–3:79–113 Pensions Institute Discussion Paper No. PI-0814.

Hiam, L., Harrison, D., McKee, M., Dorling, D. (2018). Why is life expectancy in England and Wales 'stalling'?. Journal of Epidemiology and Community Health, 72(5): 404-408.

Hinton, G., Srivastava, N., 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. 2014. Available online: `http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf`

Human Mortality Database (2018). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Data downloaded on 01/01/2019. `https://www.mortality.org`.

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

Lee, R. D. (1993). Modeling and forecasting the time series of us fertility: Age distribution, range, and ultimate level. International Journal of Forecasting 9(2), 187–202.

Lee, R.D. (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. D., and Carter, L. R. (1992). Modeling and forecasting US mortality. Journal of the American Statistical Association, 87: 659–71.

Miller, R. B. (1986). A bivariate model for total fertility rate and mean age of childbearing. Insurance: Mathematics and Economics 5(2), 133-140

Nigri, A., Levantesi, S., Marino, M., 2019. Life Expectancy and Lifespan Inequality forecasting. A Deep Learning approach. (Working paper).

Oeppen, J., Vaupel, J.W. (2002). Broken Limits to Life Expectancy. Science, 296(5570): 1029–1031.

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

Pascariu, M. D., Basellini, U., Aburto, J.M.,Canudas-Romo, V. The Linear Link: Deriving Age-Specific Death Rates from Life Expectancy. (`https://www.scor.com/sites/default/files/pascariu_-_2018_-_modelling_and_forecasting_mortality.pdf`)

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

Pascariu, M.D., Lenart, A., Canudas-Romo, V. (2019). The maximum entropy mortality model: forecasting mortality using statistical moments. Scand. Actuar. J. 2019 (8): 661-685.

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

Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. Psychological Review 65: 386-408.

Ševčíková, H., N. Li, V. Kantorová, P. Gerland, and A. E. Raftery (2016). Age-specific mortality and fertility rates for probabilistic population projections. In Dynamic demographic analysis, 285-310. Springer.

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

Vallin, J., Meslé, F. (2009). The Segmented Trend Line of Highest Life Expectancies. Population and development review, 35 (1): 159–187.

White, K. M. (2002). Longevity advances in high-income countries, 1955-96. Population and Development Review 28 (1): 59-76.