Life Expectancy and Lifespan Inequality forecasting. A Deep Learning approach.

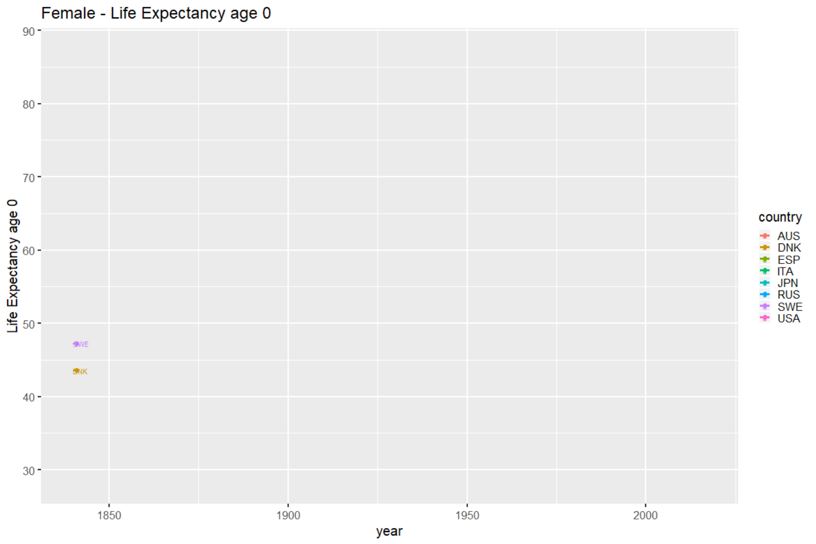
Andrea Nigri, Mario Marino, Susanna Levantesi

13.09.2019

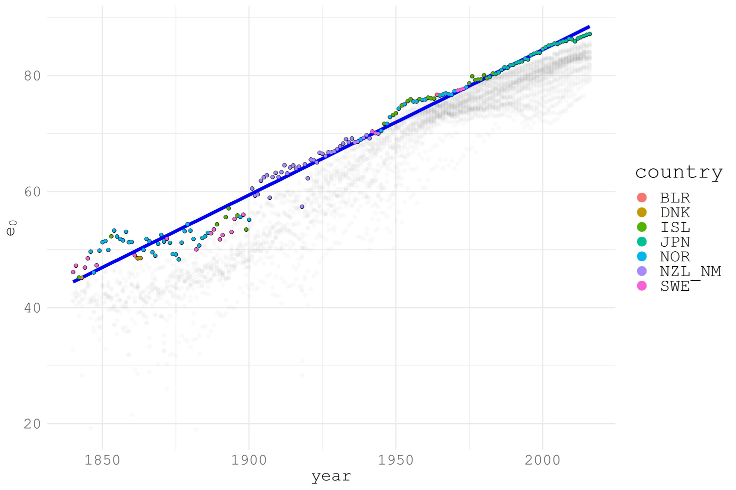
Sapienza University of Rome

Background

Steady improvement in mortality level



Oeppen & Vaupel 2002: Best Practice LE (BPLE)



Life Expectancy Forecasting: State of Art

Life Expectancy Forecasting: State of Art

- **Lee 2006:** Stochastic behavior of the changes in life expectancy trend, assuming that the average changes are functions of the gap with the BPLE.
- **Torri and Vaupel 2012:** Geometric-Brownian motion model.
- **Raftery et al. 2014 :** Females projections using Hierarchical Bayesian model from Raftery et al. 2013. Male projection by using the gender gap in life expectancy.
- **Pascariu et al. 2018:** Double Gap Model using for women the gap between country female life expectancy and BPLE, for men the gap with female life expectancy.

Our Contribution

Our Contribution

- Deep learning approach: Recurrent Neural Network with Long Short-Term Memory (LSTM) architecture.
- ARIMA (p,d,q) and the Double Gap (Pascariu et al.2018) for model comparison (for each country and both genders).
- Forecasting of life expectancy and life disparity (life expectancy lost due to death).
- Coherent prediction with historical trend and bio-demographic criteria.

Our Research Approach

The constant improvement of BPLE suggests:

- Links between events.
- Regular flow of continuous progress (Oeppen and Vaupel 2006).
- Mortality is linked to social progress in term of health, nutrition, education and medicine (Riley 2001)
- Life expectancy as a latent variable encompassing many factors that affect mortality dynamic.

Latent Behavior

- This latent behavior should be stressed in the forecasting phase.
- Considering the entire life expectancy series is crucial in order to catch either short and long term factors on mortality improvements.
- To this purpose, we need models able to catch more in-depth features into the historical data of life expectancy.

Model

Model: Recurrent NN (LSTM)

Long Short Term Memory: suitable for sequential data structure.

- **Preserves the significant information** over the time, thus preventing older signals from vanishing during processing.
- **It is capable to capture the noise** of the past mortality trend and repeat it in forecasting trend.

Model: Recurrent NN (LSTM)

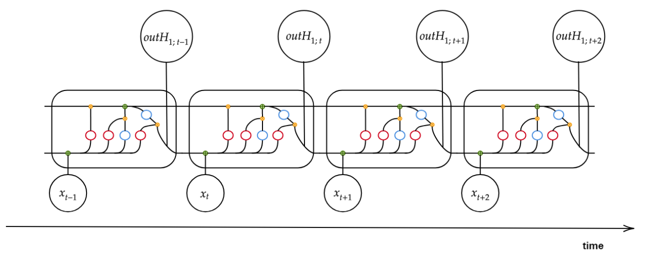
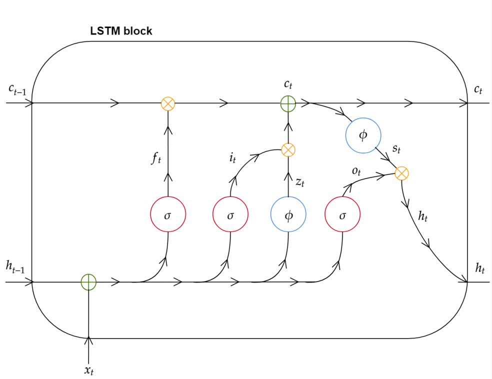
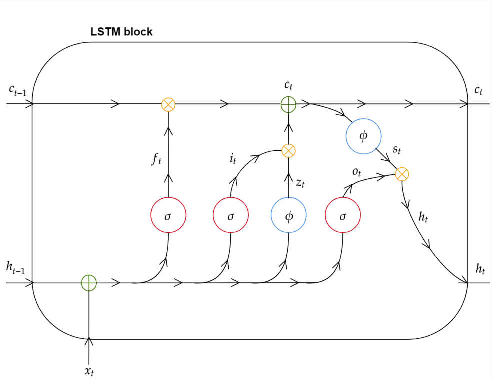


Figure 1: Source: Nigri et al. 2019

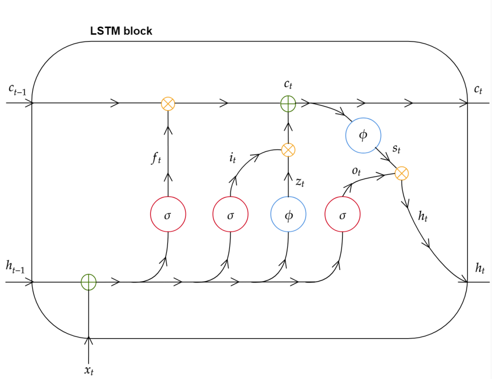
- The horizontal line which passes over the diagram. It works like a transporter belt on which information flows.



- forget f_t gate layer that decides what information to reject from the cell.



- To store, two layers are used: a sigmoid layer (the input gate layer i_t), which decides which values we will update and the tanh layer that creates a vector of new candidate values (z_t).



- We then update C_{t-1} to obtain the new candidate values.
- Through the layer sigmoid, we decide which part of the state of the cell we will return as output (o_t). Then we have the new short-term h_t and long term c_t results.

Numerical application

Numerical application

Autoregressive patterns

$$e^{g}_{x,t_k} = f\left(e^{g}_{x,t_{k-1}}, e^{g}_{x,t_{k-2}}, \cdots, e^{g}_{x,t_{k-J}}\right) + \epsilon_{t_k}, \qquad (1)$$

where $J \in \mathbf{N}$ is the number of time lags, ϵ is a homoschedastic error term and f is a neural network.

Splitting rule: 80% training and 20% testing. Thus, for a given age x and gender g, the series $e^{g}_{x,\mathbf{t}}$ can be written as follow:

- $e^{g,\ Train}_{x,\mathbf{t}'} = \left\{e^{g}_{x,t_1}, \cdots, e^{g}_{x,\tau}\right\}$ Training set is used for supervised learning,
- $e^{g,\ Test}_{x,\mathbf{t}''} = \left\{e^{g}_{x,\tau+1}, \cdots e^{g}_{x,t_m}\right\}$ Testing is used to test the forecasting accuracy. τ is the forecasting basis year, defined by the splitting rule.

Numerical application

For a specific $g = gender$

Output	Input		
$\hat{e}_{x,t_k}$	$e_{x,t_{k-1}}$	$e_{x,t_{k-2}}$	$e_{x,t_{k-3}}$
$\hat{e}_{x,t_{k+1}}$	$\hat{e}_{x,t_k}$	$e_{x,t_{k-1}}$	$e_{x,t_{k-2}}$
...	...	...	...
$\hat{e}_{x,t_{k+n}}$	$\hat{e}_{x,t_{k+(n-1)}}$	$\hat{e}_{x,t_{k+(n-2)}}$	$\hat{e}_{x,t_{k+(n-3)}}$

Network has learned the input-output functional relationship
...it is able to predict future values of $\hat{e}^g_{x,\mathbf{t}}$ using only the input.

Life expectancy and Life span inequality

The life expectancy: measures the **average mortality** levels

$$e_{x,t} = \frac{\int_x^\infty S(y,t)dy}{S(x,t)} \quad (2)$$

Lifespan Inequality: quantifies the inequality within a population (compression of mortality) (Vaupel et al. 2011).

$$e_0^\dagger(t) = -\int_0^\infty S(x,t)\ln S(x,t)dx \quad (3)$$

It is an index of heterogeneity in population health therefore crucial in insurance and for public provision of medical care. (van Raalte 2018).

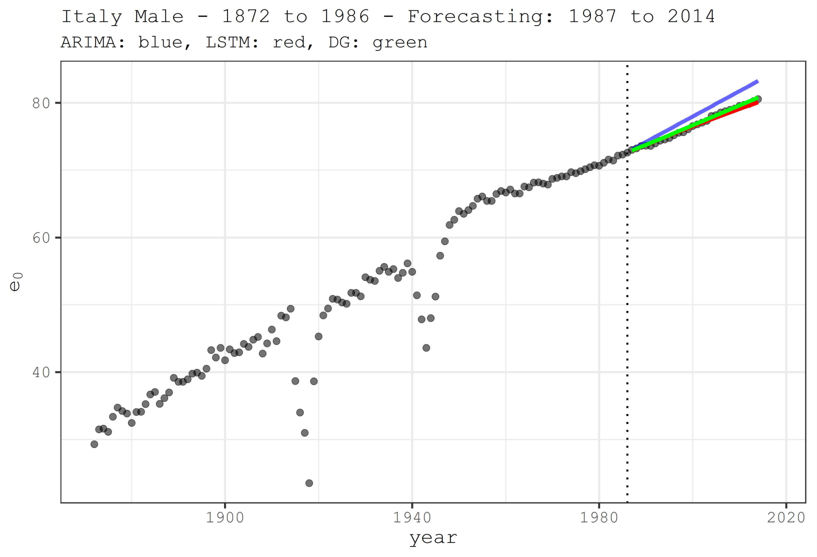
Results

Data

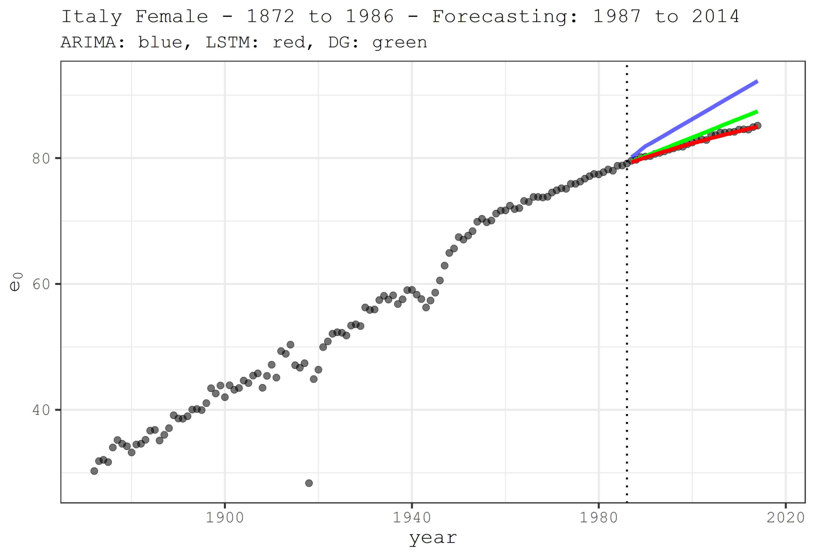
Data from **HMD**

Country	Total Years	Testing Set Years
Australia	1921–2016	1998–2016
Italy	1872–2014	1987–2014
Japan	1947–2017	2002–2016
USA	1933–2017	2001–2017
Sweden	1751–2017	1965–2017

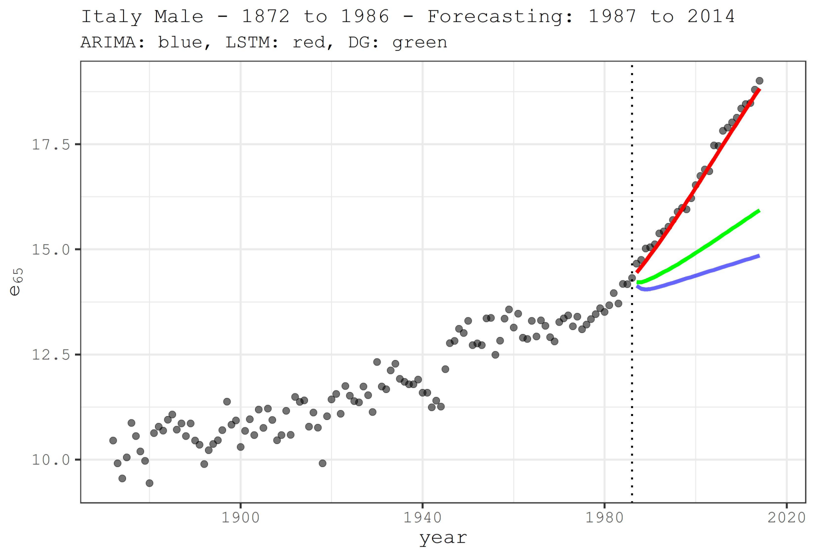
Life Expectancy at 0



Life Expectancy at 0



Life Expectancy at 65



Life Expectancy at 65

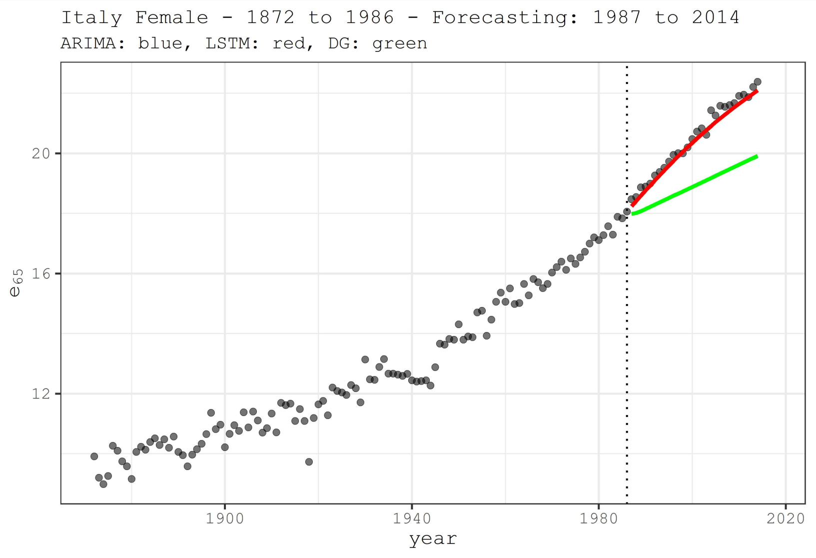


Table 1: e_0 performances of LSTM, best ARIMA and DG.

Country	Male		Female	
Australia	*MAE*	*RMSE*	*MAE*	*RMSE*
ARIMA	0.903	0.958	0.500	0.524
DG	<u>**0.229**</u>	<u>**0.265**</u>	0.523	0.706
LSTM	**0.233**	**0.268**	**0.340**	**0.449**
Italy	*MAE*	*RMSE*	*MAE*	*RMSE*
ARIMA	1.522	1.673	3.918	4.332
DG	**0.223**	**0.263**	5.647	5.688
LSTM	<u>**0.219**</u>	<u>**0.261**</u>	**0.214**	**0.275**

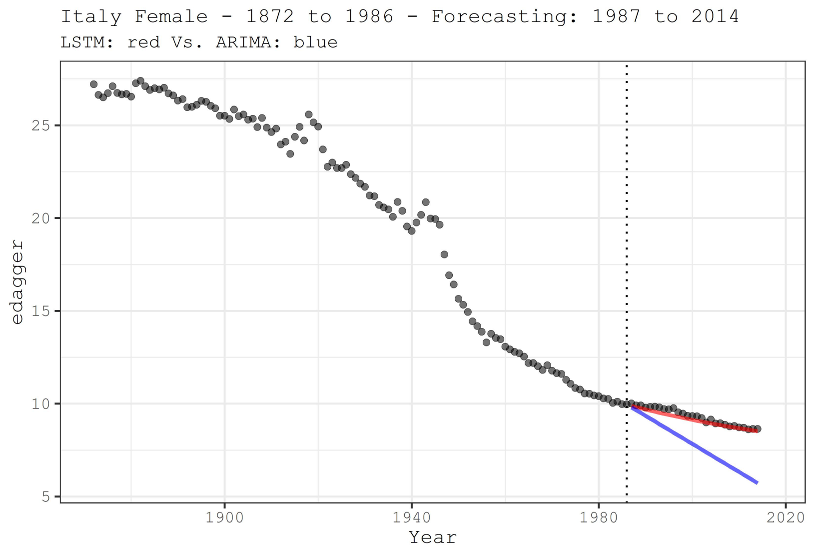
Country	**Male**		**Female**	
Japan	*MAE*	*RMSE*	*MAE*	*RMSE*
ARIMA	1.538	1.870	0.748	0.897
DG	1.877	2.119	2.163	2.495
LSTM	**0.335**	**0.372**	**0.265**	**0.208**
USA	*MAE*	*RMSE*	*MAE*	*RMSE*
ARIMA	0.389	0.592	0.799	0.904
DG	1.318	1.459	0.799	0.904
LSTM	**0.234**	**0.297**	**0.249**	**0.357**
Sweden	*MAE*	*RMSE*	*MAE*	*RMSE*
ARIMA	1.163	1.283	0.347	0.411
DG	**0.520**	**0.654**	0.485	0.623
LSTM	0.797	0.919	**0.116**	**0.152**

Table 2: e_{65} performances of LSTM, best ARIMA and DG.

Country	**Male**		**Female**	
Australia	*MAE*	*RMSE*	*MAE*	*RMSE*
ARIMA	0.587	0.615	0.845	0.891
DG	1.286	1.406	0.782	0.822
LSTM	**0.157**	**0.189**	**0.186**	**0.215**
Italy	*MAE*	*RMSE*	*MAE*	*RMSE*
ARIMA	2.292	2.536	1.569	1.683
DG	1.713	1.891	1.569	1.683
LSTM	**0.124**	**0.146**	**0.161**	**0.194**

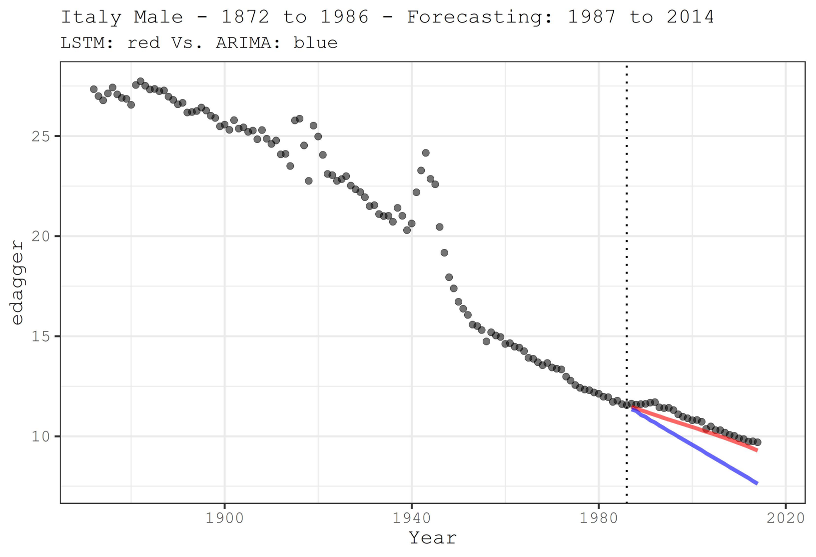
Country	**Male**		**Female**	
Japan	*MAE*	*RMSE*	*MAE*	*RMSE*
ARIMA	0.174	0.199	0.603	0.713
DG	5.804	5.818	0.849	0.973
LSTM	**0.109**	**0.142**	**0.24**	**0.279**
USA	*MAE*	*RMSE*	*MAE*	*RMSE*
ARIMA	0.971	1.058	0.345	0.398
DG	0.807	0.868	0.273	0.313
LSTM	**0.186**	**0.221**	**0.151**	**0.184**
Sweden	*MAE*	*RMSE*	*MAE*	*RMSE*
ARIMA	1.389	1.946	2.193	2.417
DG	1.187	2.3751	2.459	2.696
LSTM	**0.182**	**0.252**	**0.221**	**0.294**

e-dagger



Italy Female - 1872 to 1986 - Forecasting: 1987 to 2014
LSTM: red Vs. ARIMA: blue
edagger
25
20
15
10
5
1900
1940
1980
2020
Year

e-dagger



Italy Male - 1872 to 1986 - Forecasting: 1987 to 2014
LSTM: red Vs. ARIMA: blue
edagger
25
20
15
10
1900
1940
1980
2020
Year

e-dagger

Table 3: $e^{\dagger}$ performances of LSTM, best ARIMA.

Country	Male	Female
Australia	*RMSE*	*RMSE*
ARIMA	0.217	0.192
LSTM	**0.150**	**0.086**
Italy	*RMSE*	*RMSE*
ARIMA	1.341	1.714
LSTM	**0.392**	**0.160**
Japan	*RMSE*	*RMSE*
ARIMA	1.060	1.102
LSTM	**0.232**	**0.210**

e-dagger

Country	Male	Female
USA	*RMSE*	*RMSE*
ARIMA	1.109	0.635
LSTM	**0.172**	**0.155**
Sweden	*RMSE*	*RMSE*
ARIMA	24.752	20.012
LSTM	**15.552**	**16.553**

Summary:

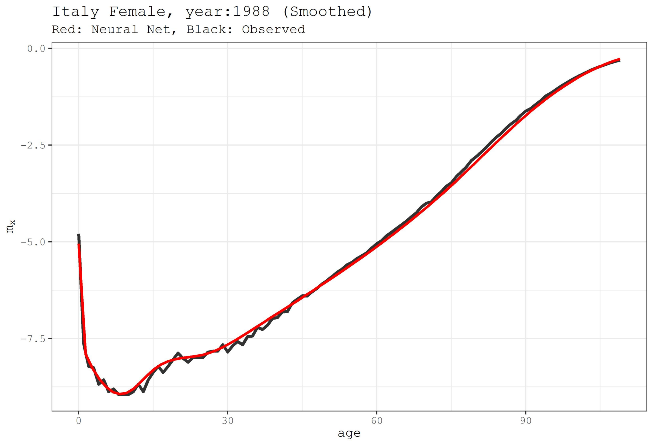
- Our approach considers forecasting of e-dagger (the life expectancy lost due to death).

We have shown that: without imposing model restrictions (regarding the gender gap and the BPLE). **The results:**

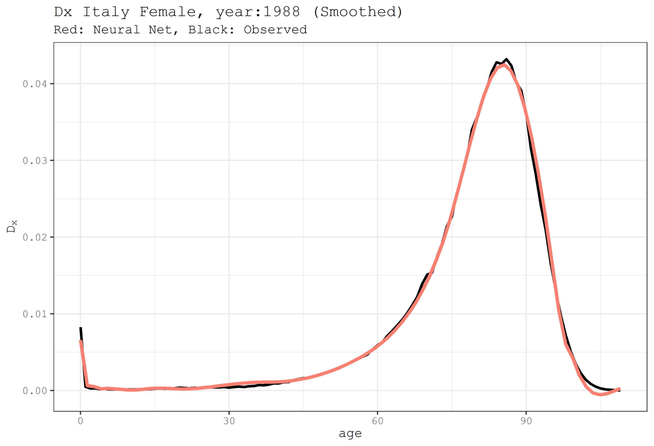
- Are coherent with the historical trend.
- Satisfy bio-demographic reasonableness criteria in particular in presence of non linearity.
- Outperform existing models in particular on life expectancy at 65 years, in which the BPLE assumption shows weakness and the gender gap is going to become a smoothed phenomenon.

Work in progress

Moreover, an accurate forecast of e_0 and e_{65} is crucial in order to get mortality surface from life expectancy.



Work in progress

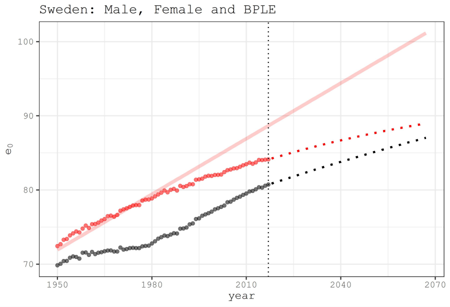


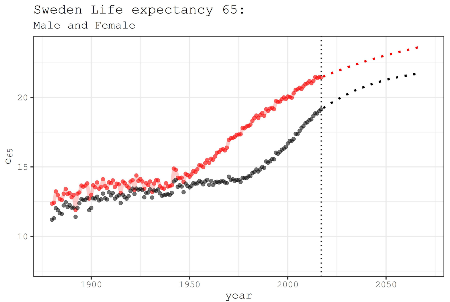
Thank you for your attention.

Questions / Suggestions ?

Appendix

Forecast over next 50 years





Recurrent Neural Network: LSTM

Formally, referring to a hidden layer composed of $n_h \in \mathbb{N}$ of LSTM blocks, letting $W \in \mathbb{R}^{d \times n_h}$ and $U \in \mathbb{R}^{n_h \times n_h}$ weight matrices (for each gate) be respectively associated to the input and to the previous short term result, the general mode of operation of a recurrent network with LSTM architecture is described by specifying the forget gate output ($\mathbf{f}_t \in \mathbb{R}^{n_h}$), input gate output ($\mathbf{i}_t \in \mathbb{R}^{n_h}$), output of output gate ($\mathbf{o}_t \in \mathbb{R}^{n_h}$), and output of auxiliary-output gate ($\mathbf{z}_t \in \mathbb{R}^{n_h}$):

LSTM

$$\mathbf{f}_t = \sigma(W_f \mathbf{x}_t + U_f \mathbf{h}_{t-1} + \mathbf{b}_f), \tag{4}$$

$$\mathbf{i}_t = \sigma(W_i \mathbf{x}_t + U_i \mathbf{h}_{t-1} + \mathbf{b}_i), \tag{5}$$

$$\mathbf{o}_t = \sigma(W_o \mathbf{x}_t + U_o \mathbf{h}_{t-1} + \mathbf{b}_o), \tag{6}$$

$$\mathbf{z}_t = \phi(W_z \mathbf{x}_t + U_z \mathbf{h}_{t-1} + \mathbf{b}_z). \tag{7}$$

LSTM

We define the processing of the entire input block, which participates in formulation of the current state of memory cell, as follows:

$$\mathbf{c}_t = \mathbf{c}_{t-1} \circ \mathbf{f}_t + \mathbf{i}_t \circ \mathbf{z}_t. \tag{8}$$

To obtain the current output, a combination between a function of $\mathbf{c}_t$, $\mathbf{s}_t = \phi(\mathbf{c}_t)$, and the upshot of auxiliary NN associated to output gate is necessary:

$$\mathbf{outH} = \mathbf{s}_t \circ \mathbf{o}_t. \tag{9}$$