Fall 10-16-2021

Forecasting mortality rates with Recurrent Neural Networks: A preliminary investigation using Portuguese data

Jorge M. Bravo
Universidade Nova de Lisboa NOVA IMS & Université Paris-Dauphine PSL & MagIC & CEFAGE-UE, jbravo@novaims.unl.pt

Forecasting mortality rates with Recurrent Neural Networks: A preliminary investigation using Portuguese data

Jorge M. Bravo, NOVA IMS, Universidade Nova de Lisboa & Université Paris-Dauphine PSL & MagIC & CEFAGE-UE, Portugal, jbravo@novaims.unl.pt

Abstract

Forecasts of age-specific mortality rates are a critical input in multiple research and policy areas such as assessing the overall health, well-being, and human development of a population and the pricing and risk management of life insurance contracts and longevity-linked securities. Model selection and model combination are currently the two competing approaches when modelling and forecasting mortality, often using statistical learning methods. This paper empirically investigates the predictive performance of Recurrent Neural Networks (RNN) with Long Short-Term Memory (LSTM) architecture in jointly modelling and multivariate time series forecasting of age-specific mortality rates across the entire lifespan. We empirically investigate different hyperparameter choices in three hidden layers LSTM models and compare the model's forecasting accuracy with that produced by classical age-period and age-period-cohort stochastic mortality models. The empirical results obtained using data for Portugal suggest that the RNN with LSTM architecture can outperform traditional benchmarking methods. The LSTM architecture generates smooth and consistent forecasts of mortality rates at all ages and across years. The predictive accuracy of the LSTM network is higher for both sexes, significantly outperforming the benchmarks in the male population, an interesting result given the added difficulties posed by the mortality hump and higher variability in male survival functions. Further investigation considering other RNN architectures, calibration procedures, and sample datasets is necessary to confirm the robustness of deep learning methods in modelling human survival.

Keywords: Mortality forecasting; RNN; LSTM; deep learning; pensions; insurance.

1. INTRODUCTION

Forecasts of age-specific mortality rates are a critical input in the computation of period and cohort life tables, in estimating life expectancy and other longevity markers, in producing annual and infra-annual forecasts of the resident population, in pricing life insurance and retirement income contracts, in the pricing and risk management of longevity-linked securities and their derivatives (Hyndman & Booth, 2008; Bravo & El Mekkaoui, 2018; Bravo & Coelho, 2019; Bravo, 2016, 2019, 2020; Simões et al., 2021; Bravo & Silva, 2006). Period life expectancy computed from age-specific survival probabilities is one of the most used markers to assess the overall health, well-being, and human development of a population. Its level and trends are used broadly by national and supra-national health organizations and researchers to assess the impact of health policies or health shocks (e.g., epidemics), to identify the lost years due to the leading causes of death, to estimate and forecast lifespan inequality, to compare countries and regions (Luy et al., 2019; UN, 2020).

More recently, life expectancy was used to automatically link earnings-related pension benefits to longevity developments observed at retirement ages, as part of broader pension reforms aiming to minimize the impact of population aging and economic shocks on the financing of pension schemes. The link has been established in multiple ways (Ayuso et al., 2021a,b; Bravo & Herce, 2020; Bravo & Ayuso, 2021a,b): (i) by indexing the

normal and the early retirement ages to life expectancy (e.g., Portugal, UK, Slovakia, Finland, Denmark, The Netherlands, Italy, Estonia, Greece, Cyprus); (ii) by linking the entry pension benefits to the so-called sustainability factors (e.g., Finland, Portugal, Spain), to old-age dependency ratios (e.g., Germany, Japan) or life annuity factors (e.g., Sweden); (iii) by indexing the eligibility requirements to the contribution length (e.g., France, Italy); (iv) by conditioning the annual pension indexation (e.g., Luxembourg); (v) by indexing the pension penalties (bonuses) for early (late) retirement to the contribution length (e.g., Portugal); (vi) by introducing longevity-linked life annuities in public and private pension schemes (e.g., USA); (vii) by conditioning the uprating of pension entitlements (e.g., The Netherlands).

In the actuarial and demographic literature, the traditional approach to age-specific mortality rate forecasting is to select a single or a multi-population discrete-time or continuous-time stochastic mortality model from a set of candidate statistical learning approaches using some method or criteria (e.g., in-sample BIC, cross-validation, forecasting accuracy metric), often neglecting model uncertainty for statistical inference purposes (see, e.g., Lee & Carter (1992); Currie (2006); Pascariu et al. (2020), Bravo & Nunes (2021); Li & Shi (2021); Carbonneau (2021), Bravo (2021); Li et al. (2021) and references therein). Empirical studies suggest that there is no single forecasting method that performs consistently well across all data sets and time horizons. Recently, model combinations of heterogeneous generalized age-period-cohort (GAPC) stochastic mortality models, principal component methods, and smoothing approaches were proposed to address model uncertainty and to improve the forecasting accuracy (Kontis et al., 2017; Bravo et al., 2021; Ayuso et al., 2021b; Ashofteh & Bravo, 2021a,b; Bravo & Ayuso, 2020; 2021a,b; Ashofteh et al., 2021). Projections often assume independence between subpopulations (e.g., men and women) (Hyndman et al., 2013). The use of machine learning and deep learning techniques in mortality forecasting is recent. Deprez et al. (2017) use machine learning techniques to investigate the fitting accuracy of mortality models. Hainaut (2018) proposes a neural network to predict and simulate mortality rates. Nigri et al. (2019) integrate a Recurrent Neural Network (RNN) within the Lee-Carter model to improve its predictive capacity. Richman and Wüthrich (2019a,b) propose a multiple-population extension of the Lee-Carter model where parameters are estimated using neural networks. Perla et al. (2021) propose a convolutional network model to forecast mortality rates. Bravo (2021c) and Bravo & Santos (2021) investigate the forecasting accuracy of RNN with LSTM and/or GRU networks using Italian and Chilean data.

Against this background, this paper empirically investigates the forecasting accuracy of RNN with an LSTM architecture to jointly model and predict age-specific mortality rates for male and female populations in a coherent way over a medium-term (10-year) forecasting horizon. We empirically investigate different choices of the hyperparameters of the three hidden layers LSTM model (units for each layer, epochs, optimizer, learning rate, activation function) considering alternative graduation intervals across ages and different combinations of the number of hidden neurons. Next, the forecast results are used to compute the period life expectancy at all ages from birth to a pre-set highest attainable age, defined using a life table closing method. We compare the empirical performance of the RNN-LSTM architecture with that obtained using two traditional GAPC stochastic mortality models, namely the Lee and Carter (1992) model and the age-period-

cohort model developed by Currie (2006). The empirical strategy uses mortality data for Portugal from 1950 to 2018 (the latest available year) disaggregated by sex. The data is from the Human Mortality Database.

The empirical results obtained in this sample dataset show that the RNN with LSTM architecture designed for jointly processing the multivariate time series data of male and female mortality rates disaggregated by age can outperform traditional generalized age-period-cohort stochastic mortality models. The forecasting accuracy of the LSTM network in the male and female Portuguese populations is higher than that of both the LC and APC models, by a significant margin in the male population, an interesting result since the behaviour of male mortality across the entire lifespan is generally harder to model and forecast because of both the mortality hump at teenage and younger ages and of the higher variability of mortality at all ages. However, further investigation considering other RNN networks, different training sets, forecasting horizons, and alternative small and big populations is required to confirm the consistency of these results.

The rest of the paper is structured as follows. In Section 2, we describe the materials and methods used in this study, namely the RNN with LSTM architecture, the GAPC models used as the benchmark, the methods used to compute life expectancy, and the datasets. Section 3 presents and briefly discusses the empirical results. Section 4 concludes and sets up the agenda for further research.

2. MATERIALS AND METHODS

2.1. Recurrent Neural Networks with Long Short-Term Memory architecture

This paper empirically investigates the predictive performance of RNN with LSTM architecture in multivariate time series forecasting. RNN is an extension of Feedforward Neural Network (FNN) incorporating an internal state (memory) to process time sequences of data (inputs), where the output from the previous iteration is used as an input to the current step. Plain vanilla RNN faces problems when they must deal with time series with long-term dependency in the input sequence because of the vanishing effect of gradients when training the network using back-propagation. This reduces the ability of the network to learn from the long-term trends in the data and make accurate forecasts. To overcome the short-term memory problem of vanilla RNN, Hochreiter and Schmidhuber (1997) developed RNN with an LSTM network or architecture considering the short and long-term dependencies in the data sequences. The network blocks incorporate internal mechanisms (gates) which regulate the flow of information, learning which data in a time series is worthy of keeping or of discarding at the time of producing forecasts.

Assume we have as input (explanatory) variables (features) a time series of data $(x_1, \dots, x_T)$ with components $x_t \in \mathbb{R}^{\tau_0}$ observed at times $t = 1, \dots, T$. Our goal is to use this data to describe (e.g., forecast) a given output data $y \in \mathcal{Y} \subset \mathbb{R}$, specifically age-specific mortality rates by sex. To provide some intuition on RNN, we follow Nigri et al. (2019) and plot in Figure 1 a schematic representation of an LSTM unit or block structure. A typical LSTM block comprises three gates (an input gate, an output gate, and a forget gate) that control the flow of information into and out of the cell, their interactions, and the subsequent memory cell. The unit receives as

initial information flow the current input $x_t \in \mathbb{R}^{\tau_0}$, the output from the previous (short-term memory) LSTM unit $h_{t-1} \in \mathbb{R}^h$, with $h \in \mathbb{N}$ denoting the number of LSTM blocks in a hidden layer, and the (long-term) memory of the previous unit $c_{t-1} \in \mathbb{R}^h$. The data is processed by three gates (red circle) named, respectively, forget gate f_t, input gate i_t and output gate o_t, and auxiliary neural networks (blue circle), which assist in the normalization of the information flow. The forget gate (loss of memory rate) combines the previous unit state h_{t-1} and the current input x_t in a nonlinear way using the sigmoid activation function $\sigma(\cdot)$ to determine how much information from the past c_{t-1} is carried forward to the following units. The input gate (memory update rate) uses the sigmoid function to decide how much input data should be used to update the memory of the network, with the hyperbolic tangent activation function $\phi(\cdot)$ controlling for the importance of the values passed. Finally, in the output gate (release of memory information rate) the data input x_t, the new (updated) memory c_t, the previous output h_{t-1} and a bias vector is used to determine the output to the next LSTM unit.

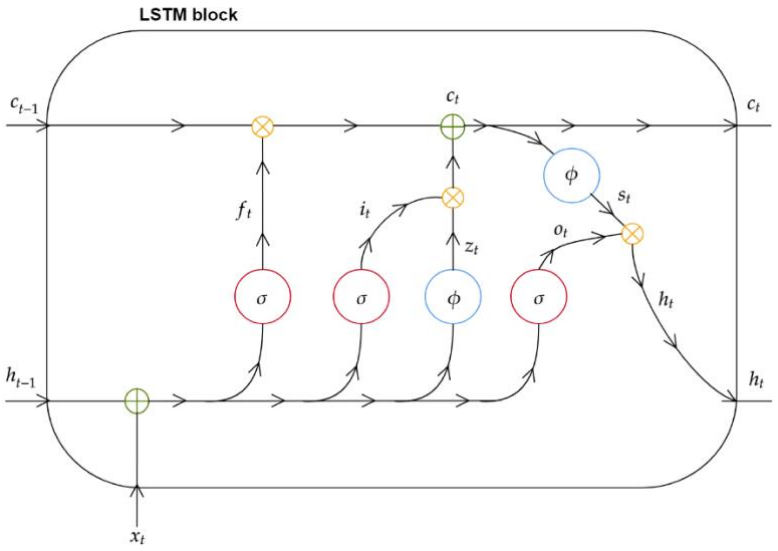

Figure 1 – Schematic representation of a Long Short-Term Memory (LSTM) block structure.
Source: Nigri et al. (2019).

Let $W \in \mathbb{R}^{\tau_0 \times h}$ and $U \in \mathbb{R}^{h \times h}$ denote the weight matrices for the input and the previous short-term result gates, respectively. The RNN with LSTM architecture can be formally described by the following set of equations:

$$f_t = \sigma\big(W_f x_t + U_f h_{t-1} + b_f\big), \tag{1}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i), \tag{2}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \tag{3}$$

$$z_t = \phi(W_z x_t + U_z h_{t-1} + b_z), \tag{4}$$

$$c_t = c_{t-1} \circ f_t \circ i_t \circ z_t, \tag{5}$$

$$h_t = \phi(c_t) \circ o_t, \tag{6}$$

$$\sigma(x) = \frac{1}{1+e^{-x}} \in (0,1) \tag{7}$$

$$\phi(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \in (-1,1) \tag{8}$$

with initial values $c_0 = 0$ and $h_0 = 0$, and where $f_t \in \mathbb{R}^h$, $i_t \in \mathbb{R}^h$, $o_t \in \mathbb{R}^h$ and $z_t \in \mathbb{R}^h$ represent the outputs of the forget gate, input gate, output gate, and the auxiliary-output gate, respectively. The symbol $\circ$ denotes the Hadamard product (element-wise product). The output $h_t \in \mathbb{R}^h$ of the LSTM block is passed to the next layer and became the short memory input for the next instance. We empirically investigate different choices of the hyperparameters of the three hidden layers LSTM model (units for each layer) considering for 10-year forecasting horizons, 500 epochs, the Adaptive Moment Estimation (Adam) optimizer, and the mean squared error as loss function.

2.2. Generalized Age-Period-Cohort stochastic mortality models

This paper considers two widely adopted and performing GAPC models as benchmarks for assessing the forecasting accuracy of the RNN with an LSTM architecture: the standard age-period (LC) Lee and Carter (1992) model under a Poisson setting and the age-period-cohort (APC) model proposed by Currie (2006). Following Ayuso et al. (2021b) and Bravo (2020, 2021a,b,c), let $D_{x,t,g}$ denote the number of deaths recorded at age x during calendar year t from the population g initially ($E^0_{x,t,g}$) or centrally ($E^c_{x,t,g}$) exposed-to-risk. GAPC models link a response variable ($q_{x,t}$ in APC; $\mu_{x,t}$ in LC) to an appropriate linear predictor $\eta_{x,t}$, capturing the systematic effects of age x, time t and year-of-birth (cohort) $c = t - x$, defined as

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \kappa_t^{(i)} + \beta_x^{(0)} \gamma_{t-x}, \tag{9}$$

where $exp(\alpha_x)$ denotes the general shape of the mortality schedule across age, $\beta_x^{(i)}\kappa_t^{(i)}$ is a set of N age-period terms describing the mortality trends, with each time index $\kappa_t^{(i)}$ specifying the general mortality trend and $\beta_x^{(i)}$ capturing its specific effect across ages, the term γ_{t-x} models cohort effects, with $\beta_x^{(0)}$ capturing its effect across ages. The substructure corresponding to the LC model is given from (9) by:

$$\eta_{x,t} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)}, \tag{10}$$

with identifiability parameter constraints

$$\sum_{x=x_{min}}^{x_{max}} \beta_x^{(1)} = 1\,, \qquad \text{and} \sum_{t=t_{min}}^{t_{max}} \kappa_t^{(1)} = 0. \tag{11}$$

The substructure corresponding to the LC model is given from (9) by

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \gamma_{t-x}, \tag{12}$$

with identifiability parameter constraints

$$\sum_{t=t_{min}}^{t_{max}} \kappa_t^{(1)} = 0\,, \qquad \sum_{c=t_{min}-x_{max}}^{t_{max}-x_{min}} \gamma_c = 0\,, \qquad \text{and} \sum_{c=t_{min}-x_{max}}^{t_{max}-x_{min}} c\gamma_c = 0. \tag{13}$$

The model specification is complemented with assumptions regarding the statistical distribution of the number of deaths: Poisson distribution in the LC model, i.e., $D_{x,t} \sim \mathcal{P}(\mu_{x,t} E_{x,t}^c)$ with $\mathbb{E}(D_{x,t}/E_{x,t}^c) = \mu_{x,t}$; Binomial distribution in the APC model, i.e., $D_{x,t} \sim \mathcal{B}(q_{x,t} E_{x,t}^0)$ with $\mathbb{E}(D_{x,t}/E_{x,t}^0) = q_{x,t}$. The parameter estimates in (10) and (12) are obtained through maximum-likelihood estimation methods. To calibrate the models, we use the same training and test sets as the previous section. To forecast mortality rates, we follow the standard approach and assume that vectors $\hat{\alpha}_x$ and $\hat{\beta}_x^{(1)}$ remains constant over time and model $\hat{\kappa}_t^{(1)}$ and $\hat{\gamma}_{t-x}$ with univariate ARIMA(p,d,q) time series methods. We make multi-step 10-year out-of-sample forecasts. For each population and model, we evaluate the out-of-sample forecasting accuracy using the mean squared error metric, defined as:

$$MSE_g = \frac{1}{N} \sum_{t=t_{min}}^{t_{max}} \sum_{x=x_{min}}^{x_{max}} \left(\mu_{x,t,g} - \hat{\mu}_{x,t,g}\right)^2. \tag{14}$$

2.3. Period and cohort life expectancy

Given the stochastic force of mortality process, the period life expectancy at age x in year t for population g is computed as follows:

$$\dot{e}_{x,g}^P(t) := \int_0^{\omega-x} {}_s p_{x,g}(t) ds, \tag{15}$$

where ω denotes the highest attainable age in the life table and ${}_\tau p_{x,g}(t)$ is the τ-year survival probability for an individual aged x, computed as

$${}_\tau p_{x,g}(t) := \mathbb{E}\left[exp\left(-\int_0^\tau \mu_{x+s,g}(s) ds\right)\right], \tag{16}$$

where $\mu_x(t)$ is assumed constant within each square in a Lexis diagram. This latter assumption allows us to proxy $\mu_x(t)$ by the central death rate, computed as $m_{x,t,g} = D_{x,t,g}/E_{x,t,g}$. For both sexes and all years, we

close life tables at the age of 115 years old (i.e., $\omega = 115$), according to the current methodology adopted by Statistics Portugal (INE) for national and subnational life table estimation.

2.4. Data

The mortality data used in this study is from the Human Mortality Database (2021). The datasets comprise the number of observed deaths $D_{x,t,g}$ and the exposure-to-risk (population counts) $E_{x,t,g}$ by individual age x ($\mathcal{X} = \{x \in \mathbb{N}, 0 \leq x \leq 110+\}$, calendar year $\mathcal{T} = \{t \in \mathbb{N}, 1950 \leq t \leq 2018\}$, year of birth and sex. Figure 1 plots the raw age-specific mortality rates $\hat{m}_{x,t,g}$ (in log scale) for the Portuguese population by age in the range 0 to 90 years old and sex.

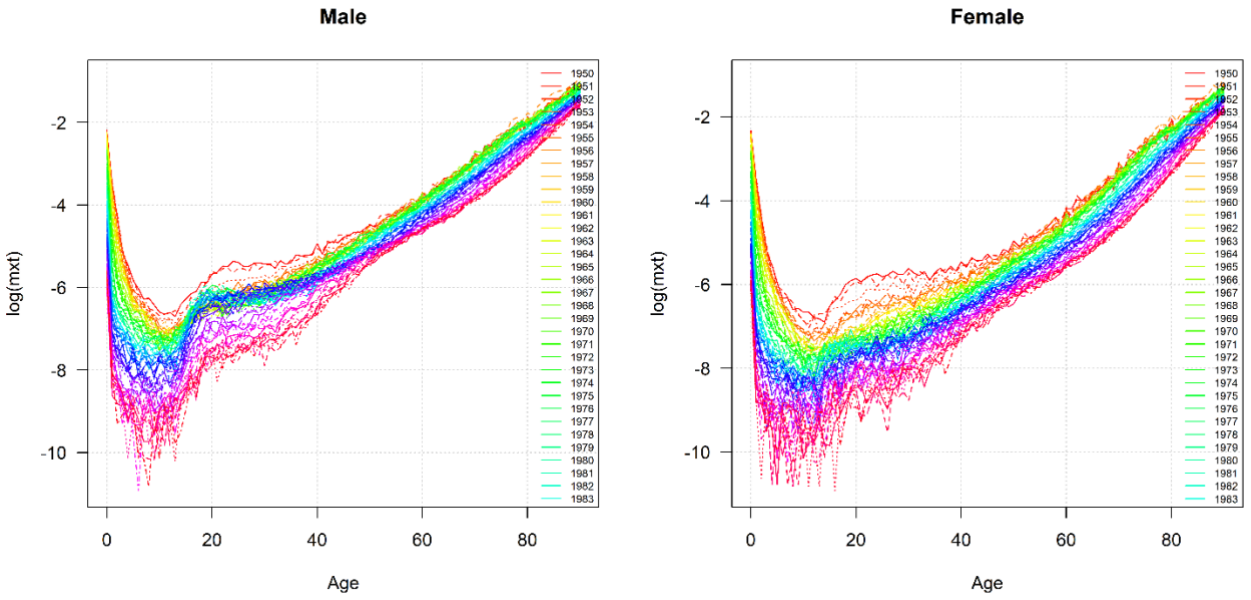

Figure 1 – Raw log-mortality rates by age and sex, Portugal, 1950-2018

A clear downward trend in the mortality rates at all ages and years is observed for both sexes, with longevity improvements more significant at younger ages and in the female population. Women have, on average, higher survival prospects than men at all ages. This is also evident in Figure 2, where we plot a heatmap and a contour plot of the raw log-mortality rates by sex.

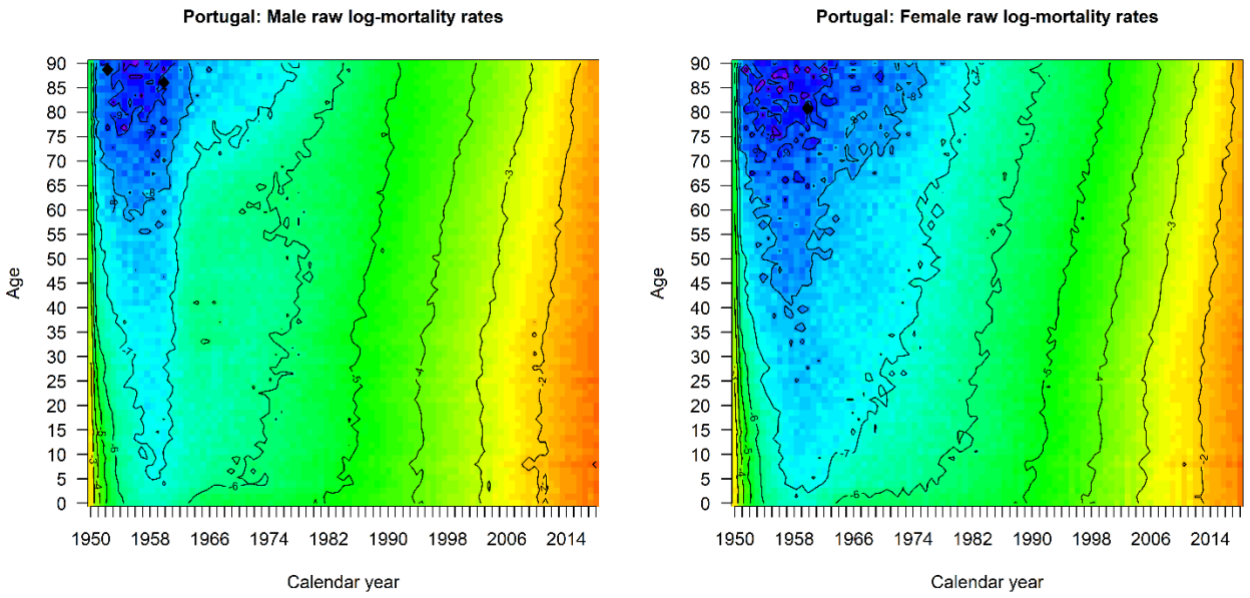

Figure 2 – Heatmap and contour plots of the raw log-mortality rates by sex, Portugal, 1950-2018

Notes: for both sexes, the orange (blue) colour represents a small (high) mortality rate.

The so-called mortality (accident) hump, which roughly affects ages in the interval [15-30], is evident in the male population but tends to smooth out in the female population. The marginally upward sloping and diagonal structure of the contour lines and colours in both heatmaps, highlight that longevity improvements have been progressively shifting from younger ages to adult and old ages.

3. RESULTS AND DISCUSSION

3.1. Hyperparameter calibration

First, we pre-process the explanatory variables to map them in the [-1;1] domain using the Min-Max Scaler and add an indicator variable for the genders (female and male) considered. To guarantee coherence and equal distribution between the estimations made for the male and female populations, the genders are alternated in the training dataset. For model learning, we follow Richman and Wüthrich (2019a,b) and partition the training data at random into a learning dataset encompassing 80% of the data and a test dataset comprising the remaining 20% of the training data. We experimented with different values but finally opted to run 500 epochs (in batches of 100) of the gradient descent algorithm on the learning dataset using the Adaptive Moment Estimation (Adam) optimizer and selecting the calibration with the lowest MSE loss in the test set using a callback procedure.

We considered 72 different three hidden layers LSTM architectures by exploring all possible combinations in $\tau_0 = \{1,3,5\}$, $\tau_1 = \{3,5,10,20\}$, $\tau_2 = \{5,10,15\}$ and $\tau_3 = \{5,10\}$. In a first step, all LSTM architectures are trained on the training set years $\mathcal{T}_1^{train} = \{t \in \mathcal{T} \ , 1950 \leq t \leq 2008\}$ and predictive performance was

assessed on $\mathcal{T}_1^{test} = \{t \in \mathcal{T} \ , 2009 \leq t \leq 2018\}$. The MSE error on $\mathcal{T}_1^{test}$ measures the out-of-sample accuracy and is used as a criterion to select the best network architecture. In a second step, the best LSTM network (the model producing the lowest MSE error) was re-trained on $\mathcal{T}_1^{train}$, from which future mortality rates for years in $\mathcal{T}_1^{test}$ were predicted. Table 1 summarizes the average fitting and forecasting losses of all the LSTM hyperparameter combinations for the joint calibration of the male and female populations. The $LSTM_3(\tau_0 = 3;\ \tau_1 = 20;\ \tau_2 = 10;\ \tau_3 = 5)$ produced the best results, followed closely by $LSTM_3(\tau_0 = 1;\ \tau_1 = 3;\ \tau_2 = 10;\ \tau_3 = 10)$.

Figure 2 illustrates the learning strategy on the best LSTM architecture, plotting the (early stopping) in-sample loss on the learning dataset (cyan colour line), and the out-of-sample loss on the test dataset (coral colour dots) of the $LSTM_3(T = 10;\ \tau_0 = 3;\ \tau_1 = 20;\ \tau_2 = 10;\ \tau_3 = 5)$ architecture.

Table 2 summarizes the average fitting (Fit) and forecasting (Men, Women) losses of the best LSTM architecture for the joint calibration of the male and female populations, together with the corresponding results obtained using the Lee-Carter model and the Age-Period-Cohort model calibrated on the same training and test data sets. The results obtained for this sample and forecasting horizon show that the RNN with LSTM architecture designed for jointly processing the multivariate time series data of male and female mortality rates disaggregated by age can outperform the widely used GAPC stochastic mortality models. The results of the LSTM model calibrated to the male and female populations shows that the network outperforms both the LC and APC models by a significant margin in the male population, which is typically much harder to model and forecast due to the mortality hump and to the higher variability of mortality rates at all ages.

LSTM				MSE (in 10^{-5})			CPU
τ_0	τ_1	τ_2	τ_3	Fit	Men	Wom	time
1	3	5	5	2.91	3.35	1.49	224
3	3	5	5	1.58	1.40	0.80	234
5	3	5	5	1.95	2.38	1.37	251
1	5	5	5	1.15	1.05	0.43	242
3	5	5	5	1.34	1.31	0.85	236
5	5	5	5	1.38	1.56	0.68	251
1	10	5	5	2.30	2.25	1.27	249
3	10	5	5	1.28	1.04	0.44	263
5	10	5	5	1.26	1.22	0.69	285
1	20	5	5	1.30	0.93	0.43	301
3	20	5	5	3.22	4.34	2.38	339
5	20	5	5	1.83	2.70	1.25	303
1	3	10	5	1.32	1.12	0.54	258
3	3	10	5	1.45	1.40	0.85	259
5	3	10	5	1.96	2.22	1.18	283
1	5	10	5	1.21	1.14	0.45	265
3	5	10	5	3.54	4.15	1.75	234
5	5	10	5	1.53	2.02	1.32	237
1	10	10	5	2.91	3.35	1.49	224
3	10	10	5	1.58	1.40	0.80	234
5	10	10	5	1.27	1.15	0.50	235

LSTM				MSE (in 10^{-5})			CPU
τ_0	τ_1	τ_2	τ_3	Fit	Men	Wom	time
1	3	5	10	1.20	1.01	0.46	364
3	3	5	10	2.00	2.01	0.94	307
5	3	5	10	1.98	2.34	1.27	314
1	5	5	10	1.69	1.87	0.66	541
3	5	5	10	2.67	2.59	1.49	635
5	5	5	10	1.44	1.11	0.48	668
1	10	5	10	1.28	1.43	0.62	666
3	10	5	10	1.81	1.78	0.97	332
5	10	5	10	1.65	2.25	1.46	321
1	20	5	10	2.49	3.64	1.79	537
3	20	5	10	2.24	2.07	1.06	277
5	20	5	10	2.30	2.58	1.37	283
1	3	10	10	1.38	1.03	0.43	264
3	3	10	10	3.42	3.45	1.60	841
5	3	10	10	1.51	1.18	0.61	892
1	5	10	10	1.42	1.30	0.54	256
3	5	10	10	2.34	3.34	1.97	261
5	5	10	10	6.60	6.92	3.46	257
1	10	10	10	1.76	1.31	0.81	261
3	10	10	10	1.41	1.06	0.62	260
5	10	10	10	2.27	2.20	1.03	265

1	20	10	5	1.27	1.75	0.90	295	1	20	10	10	1.26	1.26	0.58	306
3	20	10	5	1.26	0.80	0.66	368	3	20	10	10	1.26	1.04	0.51	319
5	20	10	5	1.09	1.65	1.26	400	5	20	10	10	1.36	1.11	0.43	322
1	3	15	5	1.20	1.01	0.46	312	1	3	15	10	3.12	3.63	1.63	276
3	3	15	5	1.47	1.47	0.77	254	3	3	15	10	1.77	1.48	0.81	270
5	3	15	5	1.78	2.01	0.91	259	5	3	15	10	2.14	2.68	1.74	265
1	5	15	5	1.47	1.60	0.69	255	1	5	15	10	2.23	2.47	0.84	263
3	5	15	5	2.12	3.11	1.87	254	3	5	15	10	1.63	1.51	0.78	266
5	5	15	5	1.22	1.31	0.61	256	5	5	15	10	1.67	1.21	0.69	273
1	10	15	5	1.36	1.27	0.77	331	1	10	15	10	4.52	6.74	2.62	300
3	10	15	5	1.44	1.25	0.64	340	3	10	15	10	1.41	1.13	0.53	302
5	10	15	5	2.10	2.71	1.34	366	5	10	15	10	2.77	3.00	1.17	300
1	20	15	5	1.66	1.90	1.00	344	1	20	15	10	1.33	0.90	0.63	386
3	20	15	5	2.03	1.88	1.23	349	3	20	15	10	2.31	4.12	2.67	410
5	20	15	5	3.97	3.46	1.74	345	5	20	15	10	1.50	1.62	0.52	384

Table 1 – LSTM3: Average fitting and forecasting loss metrics with alternative hyperparameter combinations for the joint calibration of the male and female populations

Notes: τ_0, τ_1, τ_2 and τ_3 denote the number of hidden neurons in the hidden LSTM layers; Run times measured in seconds on a personal laptop with Intel(R) Core(TM) i7-10510U CPU@2.30GHz with 16GB RAM; MSE = Mean Squared Error; Results obtained considering for 10-year fitting (lookback) and forecasting horizons, 500 epochs (for batch sizes 100), and the Adaptive Moment Estimation (Adam) optimizer.

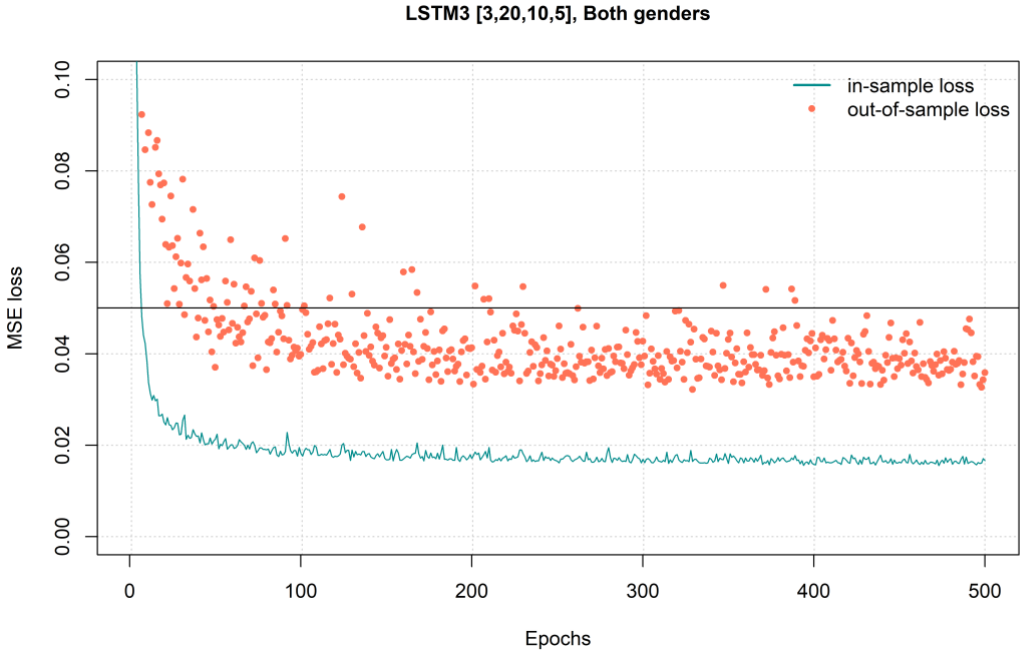

Figure 2 – Early stopping in-sample MSE loss (cyan colour) and out-of-sample loss (coral colour) of the $LSTM_3(T = 10;\ \tau_0 = 3;\ \tau_1 = 20;\ \tau_2 = 10;\ \tau_3 = 5)$ architecture

Model	MSE (in 10^{-5})			CPU time
	Fit	Men	Women	
$LSTM_3(\tau_0 = 3;\ \tau_1 = 20;\ \tau_2 = 10;\ \tau_3 = 5)$	1.26	0.80	0.66	368
Lee-Carter (LC) Model	1.31	1.44	0.70	15
Age-Period-Cohort (APC) model	2.04	1.24	1.18	17

Table 2 – Average fitting and forecasting accuracy metrics of the alternative models tested

Notes: Run times measured in seconds; MSE = Mean Squared Error; Results obtained considering for 10-year fitting (lookback) and forecasting horizons

3.2. Mortality rate and life expectancy forecasts

Figure 3 plots the age-specific forecasts of the mortality rates (in log scale) by age, year, and sex (Men, left panel; Women, right panel) produced by the best LSTM network in the test period 2009-2018. In Figure 4, a similar exercise is done for the LC and APC models by plotting the forecasts of $\dot{m}_{x,t,male}$ for the male population. Figures 7 and 8 (see Appendix), representing the raw and fitted mortality rates by individual year and sex, complement the results.

The results show that the three models produce mortality schedules that capture relatively well the dynamics of mortality rates across age and time. They also suggest that the RNN with LSTM architecture can generate smoother and consistent projections across all ages and both sexes, including at younger all where the fitting and forecasting exercises are typically more challenging. Mortality forecasting is a demanding exercise since models must capture the dynamics of individual (often one hundred) ages while preserving the relationship between mortality rates of adjacent ages and the biological impact of aging on death rates.

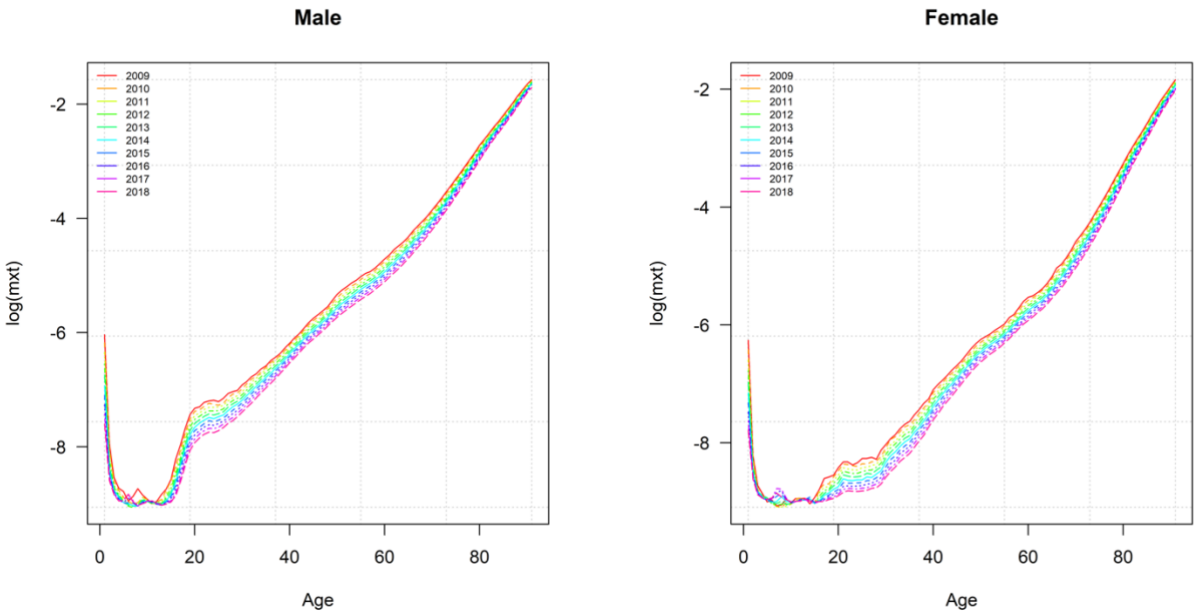

Figure 3 – Forecasts of the mortality rates by age and sex generated by the $LSTM_3(T = 10;\ \tau_0 = 3;\ \tau_1 = 20;\ \tau_2 = 10;\ \tau_3 = 5)$ network, Portugal

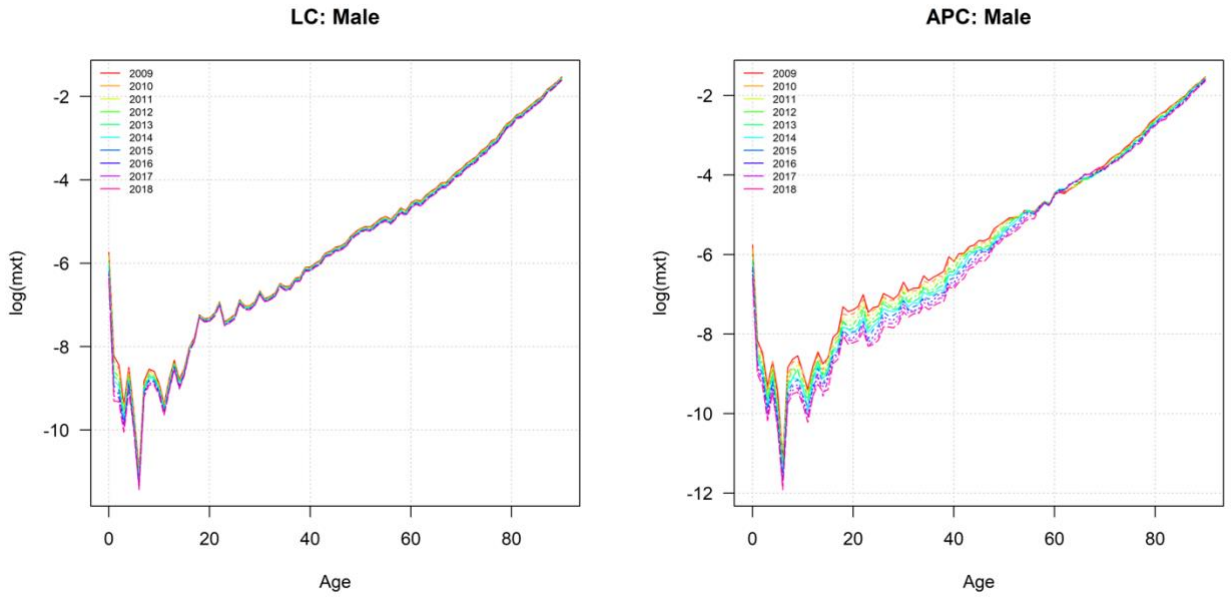

Figure 4 – Forecasts of the male population mortality rates by age generated by the LC and the APC models, Portugal

Figure 5 represents the estimates of the period life expectancy at birth ($x = 0$) and at the benchmark retirement age of 65 ($x = 65$), by sex, from 1960 to 2018, computed from the mortality rates estimated with the RNN with an LSTM architecture. The vertical dotted line marks the split between the training and test datasets. The results highlight the long-term positive trends in period life expectancy at birth and retirement ages. As of 1960, the male (female) life expectancy at birth is estimated to be 61.17 (66.79) years old, whereas the male (female) life expectancy at the age of 65 is estimated to be 12.36 (14.61) years old. The forecast results suggest that the increase in longevity at all ages will continue in an almost linear way, with period life expectancy at birth estimated to be 78.37 (83.96) years old in 2018, and the life expectancy at the age of 65 estimated to be 18.02 (21.44) years old in the same year.

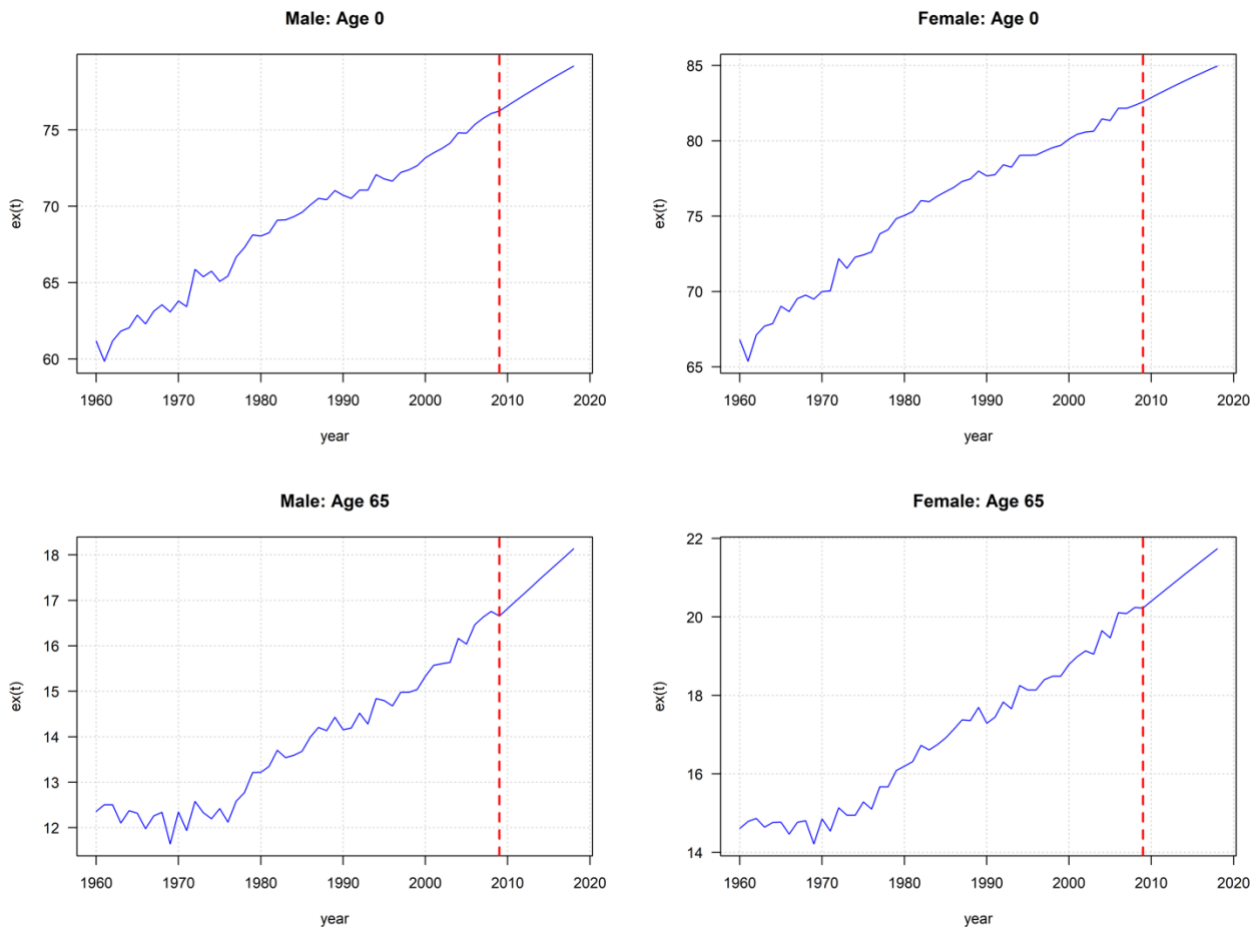

Figure 5 – $LSTM_3$: Estimates of the period life expectancy at birth and the age of 65, by sex, Portugal (values in years)

Figure 6 represents the estimates of the period life expectancy at all ages computed in 2018 (by sex) from the forecasted mortality rates generated using the LSTM network. The results highlight the well-known sex gradient in period (and cohort) life expectancy, and the higher survival prospects of women compared to that of men of all ages. The sex gradient in period life expectancy peaks at birth, with women estimated to outlive men by 5.59 years. As expected, the gradient declines steadily with age to 5.20 years at the age of 40, 3.42 years at the age of 65, 1.68 years at the age of 80, and 0.14 years for centenarians. The sex gradient in life expectancy is essential because of intragenerational fairness considerations in pension schemes.

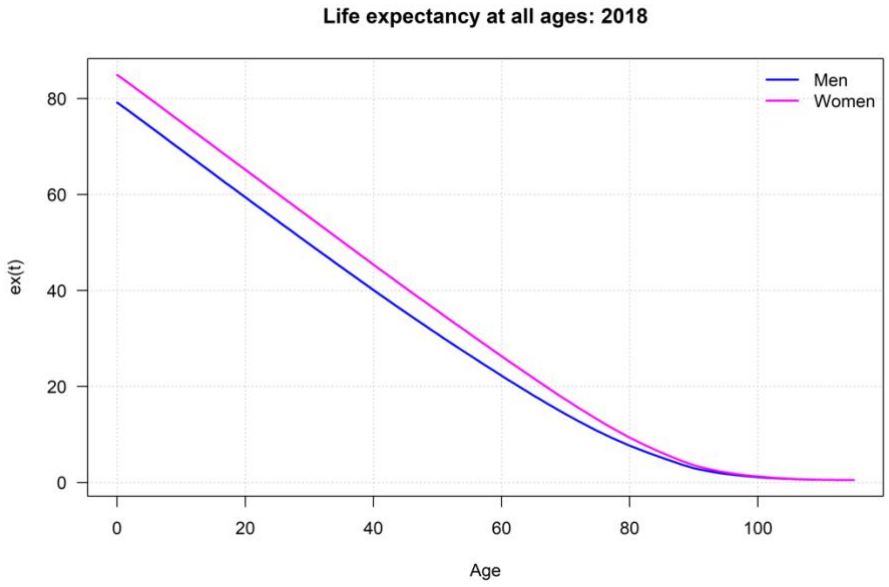

Figure 6 – $LSTM_3$: Estimates of the period life expectancy at all ages in 2018, by sex

4. CONCLUSION

RNN with an LSTM architecture is one the most popular designs of neural networks for modelling times series data. Forecasting individual age-specific mortality rates across the entire lifespan and sex or socioeconomic groups is a very demanding multivariate exercise bounded by the biological nature of the phenomena. The topic has captured and continues to capture the attention of multiple researchers and practitioners because of its importance in multiple areas, including health policy, life insurance, and capital markets instruments pricing and risk management, to retirement age estimation. In this paper, we conducted a preliminary investigation on the predictive performance of RNN with an LSTM in jointly modelling male and female mortality using the most recent Portuguese data. The empirical results show that the RNN with an LSTM architecture is suitable to model this type of dataset and provides good fitting and forecasting results compared to classical statistical learning methods. Further research investigating other RNN designs such as the Gated Recurrent Unit (GRU) or Convolution Neural Networks (CNN), other populations, alternative calibration procedures, is however needed to confirm the robustness of these methods in modelling human survival against standard approaches. Extensions to modelling mortality by socioeconomic group are critical because of the increasing interest in measuring the life expectancy gradient by income, education, or health status. This is a fundamental step in the process of correcting for the welfare redistributive impact of longevity heterogeneity in social policy, for instance, in pension design and reform (see, e.g., Bravo et al., 2021; Ayuso et al., 2021a; El Mekkaoui & Bravo, 2021).

ACNOWLEDGEMENTS

The author expresses his gratitude to the editors and to the anonymous referees for his or her careful review and insightful comments, that helped strengthen the quality of the paper.

REFERENCES

Ashofteh, A., & Bravo, J. M. (2021a). A Conservative Approach for Online Credit Scoring. Expert Systems With Applications, Volume 176, p. 1-16, 114835. https://doi.org/10.1016/j.eswa.2021.114835

Ashofteh, A. & Bravo, J. M. (2021b). Life Table Forecasting in COVID-19 Times: An Ensemble Learning Approach. 2021 16th Iberian Conference on Information Systems and Technologies (CISTI), 2021, pp. 1-6. DOI: 10.23919/CISTI52073.2021.9476583. Available: https://ieeexplore.ieee.org/document/9476583.

Ashofteh, A., Bravo, J. M. & Ayuso, M. (2021). A Novel Layered Learning Approach for Forecasting Respiratory Disease Excess Mortality during the COVID-19 pandemic. CAPSI 2021 Proceedings. (Atas da 21ª Conferência da Associação Portuguesa de Sistemas de Informação 2021), in press.

Ayuso, M., Bravo, J. M., Holzmann, R. (2021a). Getting Life Expectancy Estimates Right for Pension Policy: Period versus Cohort Approach. Journal of Pension Economics and Finance, 20(2), 212–231. https://doi.org/10.1017/S1474747220000050

Ayuso, M., Bravo, J. M., Holzmann, R., & Palmer, E. (2021b). Automatic indexation of pension age to life expectancy: When policy design matters. Risks, 9(5), 96. https://doi.org/10.3390/risks9050096

Bravo, J. M. & Coelho, E. (2019). Forecasting Subnational Demographic Data using Seasonal Time Series Methods. Atas da Conferência da Associação Portuguesa de Sistemas de Informação 2019 [CAPSI 2019 - 19th Conference of the Portuguese Association for Information Systems, Proceedings. 24].

Bravo, J. M. (2016). Taxation of Pensions in Portugal: A Semi-Dual Income Tax System. CESifo DICE Report - Journal for Institutional Comparisons. 14 (1), 14-23.

Bravo, J. M. (2019). Funding for Longer Lives: Retirement Wallet and Risk-Sharing Annuities. Ekonomiaz, 96 (2), 268-291.

Bravo, J. M. (2020). Longevity-Linked Life Annuities: A Bayesian Model Ensemble Pricing Approach. CAPSI 2020 Proceedings, 29. https://aisel.aisnet.org/capsi2020/29.

Bravo, J. M. (2021a). Pricing participating longevity-linked life annuities: A Bayesian Model Ensemble approach. European Actuarial Journal. https://doi.org/10.1007/s13385-021-00279-w

Bravo, J. M. (2021b). Pricing Survivor Bonds with Affine-Jump Diffusion Stochastic Mortality Models. ICEEG 2021: 2021 The 5th International Conference on E-Commerce, E-Business and E-Government. Association for Computing Machinery (ACM), New York, NY, USA, 91–96. https://doi.org/10.1145/3466029.3466037

Bravo, J. M. (2021c). Forecasting longevity for financial applications: A first experiment with deep learning methods. Mining Data for Financial Applications - 6th ECML PKDD Workshop, MIDAS 2021. Lecture Notes in Computer Science Series (LNCS), Springer, in press.

Bravo, J. M., & Ayuso, M. (2020). Mortality and life expectancy forecasts using bayesian model combinations: An application to the portuguese population. RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, E40, 128–144. https://doi.org/10.17013/risti.40.128–145.

Bravo, J. M., & Ayuso, M. (2021a). Forecasting the retirement age: A Bayesian Model Ensemble Approach. Advances in Intelligent Systems and Computing, Volume 1365 AIST, 123 – 135 [2021 World Conference on Information Systems and Technologies, WorldCIST 2021] Springer, Cham. https://doi.org/10.1007/978-3-030-72657-7_12.

Bravo, J. M., & Ayuso, M. (2021b). Linking Pensions to Life Expectancy: Tackling Conceptual Uncertainty through Bayesian Model Averaging. Mathematics, 9(24): 3307, p. 1-27. https://doi.org/10.3390/math9243307.

Bravo, J. M., & El Mekkaoui de Freitas, N. (2018). Valuation of longevity-linked life annuities. Insurance: Mathematics and Economics, 78, 212-229.

Bravo, J. M., & Herce, J. A. (2020). Career Breaks, Broken Pensions? Long-run Effects of Early and Late-career Unemployment Spells on Pension Entitlements. Journal of Pension Economics and Finance 1–27. https://doi.org/10.1017/S1474747220000189

Bravo, J. M., & Nunes, J. P. V. (2021). Pricing Longevity Derivatives via Fourier Transforms. Insurance: Mathematics and Economics, 96, 81-97.

Bravo, J. M., & Santos, V. (2021). Backtesting Recurrent Neural Networks with Gated Recurrent Unit: Probing with Chilean Mortality Data. Third International Conference on Computer Science, Electronics and Industrial Engineering (CSEI 2021), in press.

Bravo, J. M., & Silva, C. (2006). Immunization Using a Stochastic Process Independent Multifactor Model: The Portuguese Experience. Journal of Banking and Finance, 30 (1), 133-156.

Bravo, J. M., Ayuso, M., Holzmann, R., & Palmer, E. (2021a). Addressing the Life Expectancy Gap in Pension Policy. Insurance: Mathematics and Economics, 99, 200-221. https://doi.org/10.1016/j.insmatheco.2021.03.025.

Carbonneau, A. (2021). Deep hedging of long-term financial derivatives. Insurance: Mathematics and Economics 99, 327-340.

Currie, I. (2006). Smoothing and Forecasting Mortality Rates with P-Splines. London: Institute and Faculty of Actuaries. Available online: https://www.actuaries.org.uk/documents/smoothing-and-forecasting-mortality-rates-p-splines-handouts (accessed on 20 July 2018).

Deprez, P., Shevchenko, P., & Wüthrich, M. (2017). Machine learning techniques for mortality modeling. European Actuarial Journal, 7, 337–52.

El Mekkaoui, N., & Bravo, J. M. (2021). Drawing Down Retirement Financial Savings: A Welfare Analysis using French data. ICEEG 2021: 2021 The 5th International Conference on E-Commerce, E-Business and E-Government. Association for Computing Machinery (ACM), New York, NY, USA, 152–158. https://doi.org/10.1145/3466029.3466041

Hainaut, D. (2018). A neural-network analyzer for mortality forecast. Astin Bulletin 48, 481–508.

Hyndman, R. J., Booth, H. & Yasmeen, F. (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. Demography 50(1), 261–283.

Hyndman, R., & Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. International Journal of Forecasting 24(3), 323-342

Kontis, V., Bennett, J., Mathers, C., Li, G., Foreman, K., & Ezzati, M. (2017). Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble. Lancet 389 (10076), 1323–1335.

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

Li, H., & Shi, Y. (2021). Forecasting mortality with international linkages: A global vector-autoregression approach. Insurance: Mathematics and Economics, 100, 59-75.

Li, H., Tan, K. S., Tuljapurkar, S. & Zhu, W. (2021). Gompertz law revisited: Forecasting mortality with a multi-factor exponential model. Insurance: Mathematics and Economics, 99, 268-281.

Luy, M., Di Giulio, P., Di Lego, V., Lazarevič, P., & Sauerberg, M. (2019). Life Expectancy: Frequently Used, but Hardly Understood. Gerontology. 66(1):95-104. doi: 10.1159/000500955.

Nigri, A., Levantesi, S., Marino, M., Scognamiglio, S., & Perla, F. (2019). A Deep Learning Integrated Lee–Carter Model. Risks, 7(1), 33. http://dx.doi.org/10.3390/risks7010033

Pascariu, M., Basellini, U., Aburto, J., & Canudas-Romo, V. (2020). The Linear Link: Deriving Age-Specific Death Rates from Life Expectancy. Risks, 8(4), 109. http://dx.doi.org/10.3390/risks8040109

Perla, F., Richman, R., Scognamiglio, S., & Wüthrich, M. (2021) Time-series forecasting of mortality rates using deep learning, Scandinavian Actuarial Journal, DOI: 10.1080/03461238.2020.1867232.

Richman, R., & Wüthrich, M. (2019a). A neural network extension of the Lee–Carter model to multiple populations. Annals of Actuarial Science, 1-21. doi:10.1017/S1748499519000071

Richman, R., & Wüthrich, M. (2019b). Lee and Carter go Machine Learning: Recurrent Neural Networks. Available at SSRN: https://ssrn.com/abstract=3441030 (accessed on 10 January 2021).

Simões, C., Oliveira, L. & Bravo, J. M. (2021). Immunization Strategies for Funding Multiple Inflation-Linked Retirement Income Benefits. Risks, 9(4): 60; https://doi.org/10.3390/risks9040060

United Nations (2020). Human Development Report 2020. United Nations Development Programme, Washington.

APPENDIX

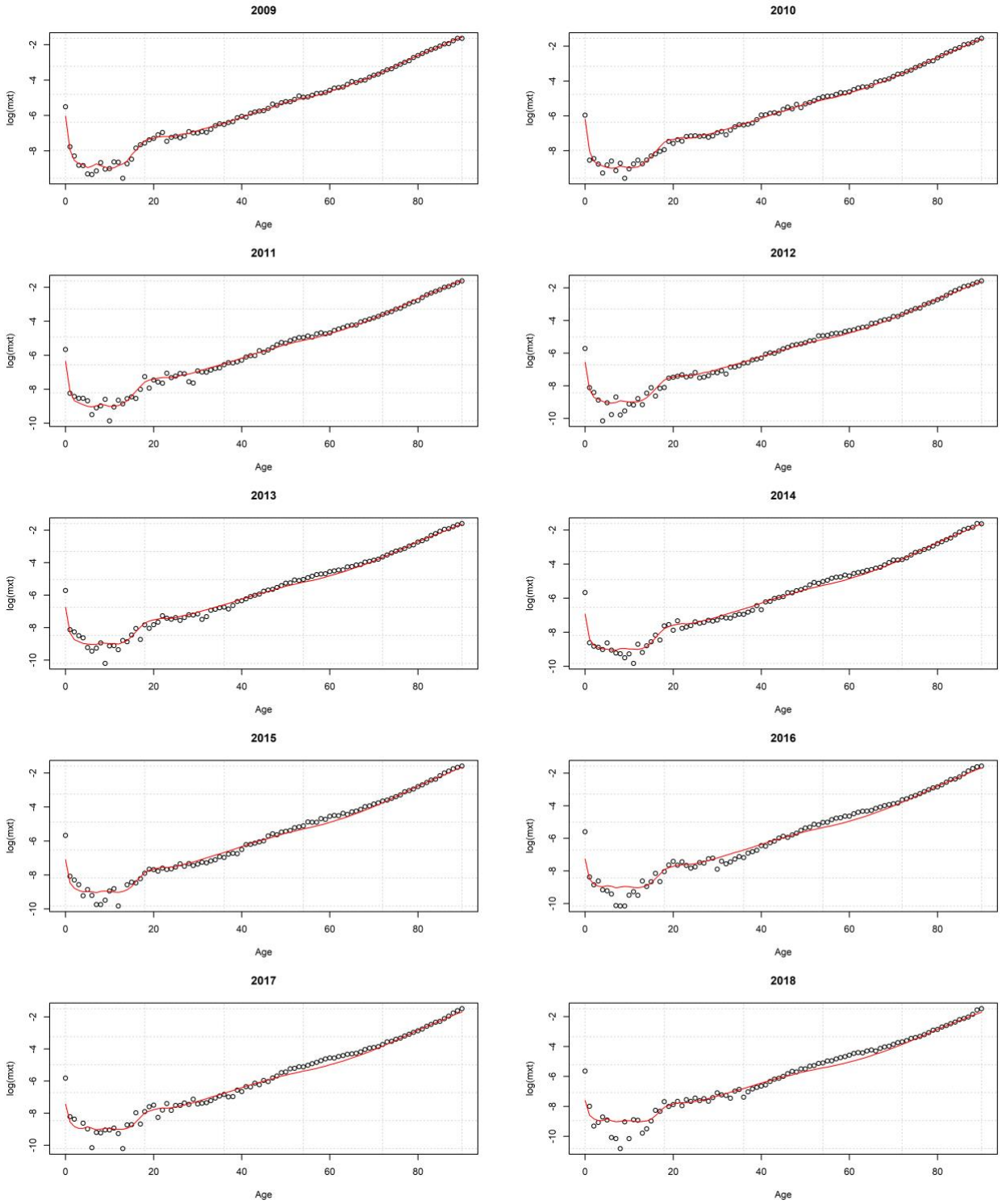

Figure 7 – $LSTM_3$: Raw and fitted log-mortality rates by individual year, Portugal, male population

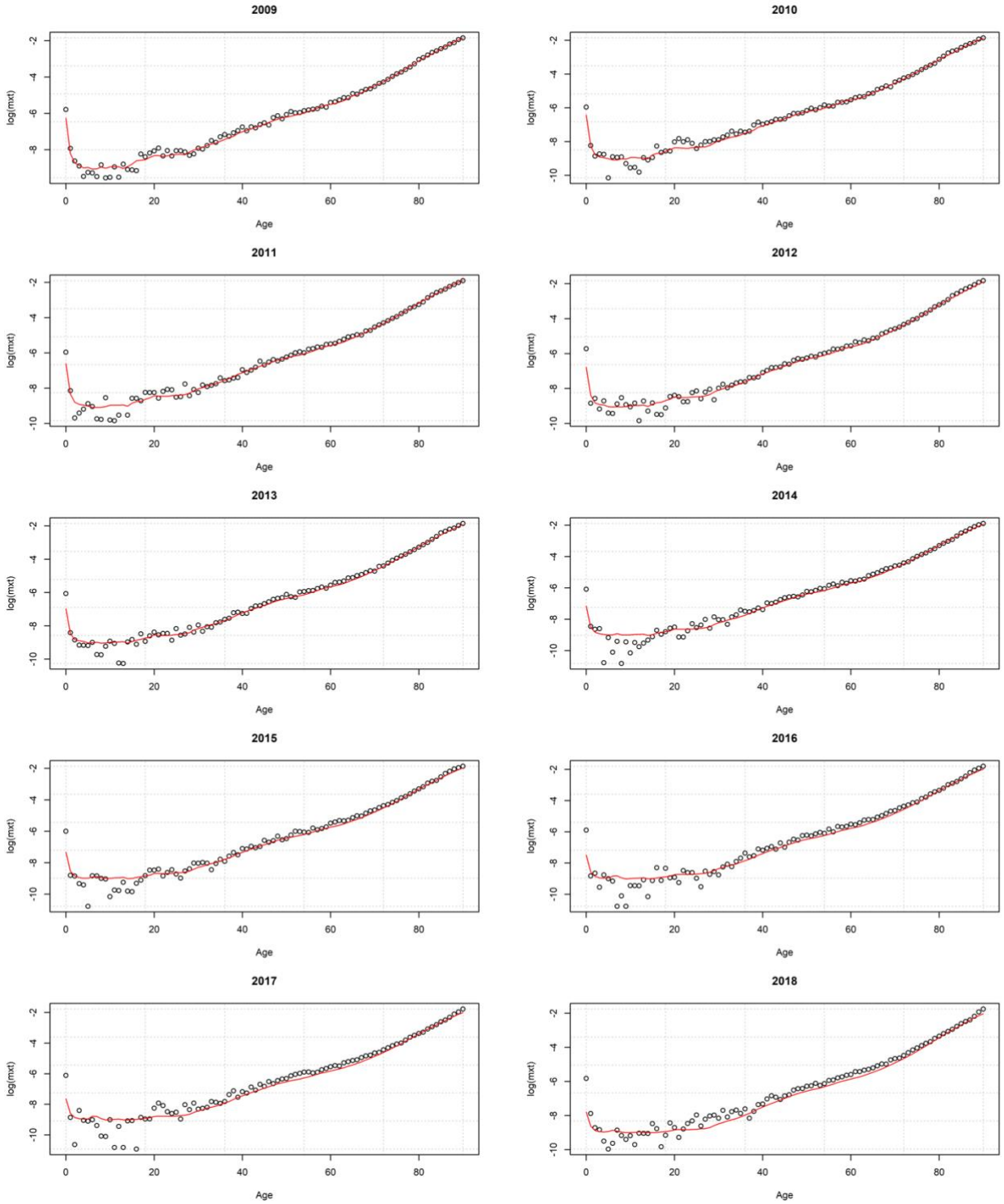

Figure 8 – $LSTM_3$: Raw and fitted log-mortality rates by individual year, Portugal, female population