Leveraging deep neural networks to estimate age specific mortality from life expectancy at birth

May 30, 2021

Abstract

Life expectancy is one of the most informative indicators of population health and development. Its stability, which has been observed over time, has made the prediction and forecasting of life expectancy an appealing area of study. However, predicted or estimated values of life expectancy do not tell us about age-specific mortality. Reliable estimates of age-specific mortality are essential in the study of health inequalities, well-being and to calculate other demographic indicators. However, this task comes with several difficulties, including a lack of reliable data in many populations. Models that relate levels of life expectancy to a full age-specific mortality profile are therefore important but scarce. We propose a deep neural networks (DNN) model to derive age-specific mortality from observed or predicted life expectancy by leveraging deep learning algorithms akin to demography's indirect estimation techniques. Out-of-sample validation was used to validate the model, and the predictive performance of the DNN model was compared with two state-of-the-art models designed to do the same thing. Out-of-sample validation indicates that the DNN model provides reliable estimates of age-specific mortality for the USA, Italy, Japan and Russia using data from the Human Mortality Database. Furthermore, we show how the DNN model could be used to estimate age-specific mortality for countries without age-specific data using neighbouring information or populations with similar mortality dynamics.

Keywords: Life expectancy, Forecasting, Death 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 at birth (hereafter referred to as 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 a lack of reliable data or stochastic variation in death counts. Because of these difficulties in several countries and sub-populations, the regularities observed in trends of life expectancy have made this indicator appealing to model and predict. A key advantage of modelling life expectancy at birth, or at any age, is that the predictive model deals only with a single indicator that summarises the overall level of mortality over time, instead of modelling single time series of death rates for each age simultaneously.

Approaches that forecast life expectancy consider past trends in this indicator and its regularities, such as the linear increase of the best practice life expectancy (Vaupel, 2002). Best practice life expectancy refers the highest sex-specific national life expectancy observed in a given year. Lee (2006) exploits this regularity and models the changes in life expectancy as a linear

function of the gap with the best practice trend, allowing countries to exceed the best practice levels. In contrast, Torri (2012) model life expectancy linearly by including a smooth function that accounts for the gap with the best practice life expectancy, which is constrained to not allow countries to overtake the best practice line. Raftery et al. (2013) introduce a Bayesian hierarchical model to obtain joint probabilistic projections of life expectancy in an international context. This model is currently used by the United Nations (2019). More recently, Nigri and Marino (2021) propose forecasting life expectancy based on recurrent neural networks. These approaches forecast males and females independently. Pascariu and Vaupel (2018) further include the well documented female advantage on longevity (Luy, 2003) to forecast life expectancy for both sexes simultaneously. Although the use of life expectancy as an indicator to forecast is appealing, estimating age-specific death rates is needed to analyze patterns of mortality at different ages and to calculate other indicators, such as lifespan inequality, as well as for estimating insurance pricing and pension liabilities. This has become even more important with the recent patterns of stalls in longevity improvements, or temporary reversals, observed in several countries, including the USA and the UK (Mehta and Abrams, 2020; Aburto et al., 2021; Hendi, 2018), but also around the globe in contexts where timely data is needed and often reported with significant delays, such as Mexico or Venezuela (Aburto et al., 2016; García and Aburto, 2019).

Here, we propose a model to derive age-specific mortality from observed or predicted life expectancies. The model leverages deep learning algorithms based on neural networks to uncover age specific mortality based on past trends. Our model is akin to Demography's long tradition of developing formal demographic methods and using statistical approaches to indirectly estimate indicators, such as model life tables (United Nations, 1955, 1967; Murray et al., 2003), Brass' relational model (Brass et al., 1968; Brass, 1971), the two-dimensional mortality model (Wilmoth et al., 2012), and more recently the singular value decomposition component mortality model (SVD-Comp) (Clark, 2019). Two approaches to deriving age-specific mortality from values of life expectancy that are more closely related to ours, which we describe in depth in the next section, were recently proposed by Ševčíková et al. (2016) and Pascariu et al. (2020). Ševčíková et al. (2016) adopts a reverting process based on the Lee-Carter model, while Pascariu et al. (2020) follow a similar strategy expressing the logarithm of age-specific deaths as a linear function of the logarithm of life expectancy.

In this article, we take advantage of deep neural network models to derive the full age-specific mortality profile from values of life expectancy overcoming the linearity assumption and data requirements from past methods and provide new insights into the indirect approaches. We also provide a further step ahead by extending our DNN model to multiple populations (i.e., countries and both sexes). The 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.

2 Models to derive age specific mortality from life expectancy

Ševčíková et al. (2016) and Pascariu et al. (2020) proposed two models aimed at deriving age-specific mortality from values of life expectancy at birth in line with the functional form of the well known Lee-Carter model (Carter, 1992) given by

$$\log\left(m_{a,t}\right) = \alpha_a + \beta_a \kappa_t + \epsilon_{t,a}, \tag{1}$$

where $m_{a,t}$ is the central death rate at age a and time t, α_a captures the log-mortality average by age, κ_t is the level of mortality in year t, β_a is an age-pattern of mortality change at age a, and $\epsilon_{t,a}$ is the error term. The following constraints on κ_t and β_a avoid identifiability problems with the parameters:

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

Carter (1992) found that κ_t changes linearly and can be forecasted using a random walk with drift or other time series methods.

2.1 Ševčíková and colleagues model

The first method proposed to estimate an age specific mortality profile from a projected or forecasted value of life expectancy at birth was developed by Ševčíková et al. (2016). Their method consists on calibrating the parameter that reflects the level of mortality in the Lee-Carter model (κ_t) to derive a desired level of life expectancy, similarly to the ideas proposed by Lee (2001) and Li and Lee (2013).

Let $t \in \{1, \ldots, T\}$ and $\tau \in \{T+1, \ldots T_p\}$ denote the observed and projected time periods, respectively. Ševčíková et al. (2016) estimate the Lee-Carter parameters α_a, k_t and β_x using the observed death rates $m_{a,t}$ independently by sex. For a given value of projected life expectancy at birth $e_0(\tau)$, the method solves for future k_τ based on the previously estimated parameters $\hat{\alpha}_a$ and $\hat{\beta}_a$ using life tables. Finally, the age specific log-death rates are derived as follows:

$$\log\left(m_{a,\tau}\right) = \hat{\alpha_a} + \hat{\beta}_a \hat{k_\tau}.$$

2.2 Linear-Link model

The Linear-Link model proposed by Pascariu et al. (2020) derives specific death rates at time t and age a, $m_{a,t}$, as a linear function of the logarithm of life expectancy at birth ($e_{0,t}$) and at time t given by:

$$\log\left(m_{a,t}\right) = \beta_a \log(e_{0,t}) + \nu_a k + \varepsilon_{a,t}. \tag{2}$$

The Linear-Link model is based on the least squares estimation of the slope β_a over the observation period. β_a can be regarded as an age-specific parameter and $\varepsilon_{a,t}$ can denote a set of normally distributed errors with mean zero and variance σ^2. The model specification involves a second step to compute the singular value decomposition (SVD) of the matrix of regression residuals to obtain the parameter ν_a. To avoid projecting age specific noise, Pascariu et al. (2020) smooth the parameters β_a and ν_a using splines. Finally, parameter k is optimised to achieve the value of a projected life expectancy. The model can also be estimated by assuming that deaths follow a Poisson distribution with maximum likelihood estimation.

3 Data

We use high quality 1×1 life tables from the Human Mortality Data Base (HMD, 2021) categorised by sex with a focus on Japan, the USA, Italy and Russia from 1950 to 2015 to test the accuracy of our model. This set of countries covers a range of longevity trajectories with Japan having one of the highest life expectancies in the world, the USA with stagnation and slow improvements in life expectancy, Italy with its late demographic transition and rapid increase in life expectancy, and Russia with the highest mortality at younger ages and lower life expectancy within the HMD.

4 Method: Deep Neural Networks

Deep learning techniques, including deep neural networks (DNN), have become important in a wide range of applications, such as image classification or speech recognition with high predictive accuracy, often on par with human performance [1]. However, its applications in demographic research are still scarce. A DNN is a collection of neurons organised in a sequence of multiple layers, where the input is the neuron activation from the previous layer that performs a weighted sum of the input followed by a nonlinear activation (Montavon and Müller, 2018). The neurons then implement complex nonlinear mapping from the input to the output. This mapping is learned from the data by adapting the weights of each neuron performing a technique known as error back-propagation (Rumelhart and Williams, 1986). The general idea is that for a given set of training data $\{(x_1, y_1) \dots (x_n, y_n)\}$ sampled according to an unknown probability distribution $\mathrm{P}(\mathbf{x}, \mathbf{y})$, we find a function $\mathrm{f}(\cdot)$ that minimises the expected error on a new test set of data:

$$\int \mathrm{L}(\mathrm{y}, \mathrm{f}(\mathbf{x}))\mathrm{P}(\mathbf{x}, \mathbf{y})\mathrm{dxdy},$$

where $\mathrm{L}(\mathbf{y}, \mathrm{f}(\mathbf{x}))$ is the loss function that measures the prediction error for a given x against the actual value y. We propose a model based on DNN that assigns to life expectancy at birth at time t a vector of age specific death rates with the structure shown in Fig. 1.

[Figure 1 about here.]

Let $\mathbf{M} = \log(m_{a,t})_{a \in A, t \in T}$ be a matrix with death rates, where rows denote age and columns calendar years, and $\mathbf{e_0} = (e_{0,t_1}, e_{0,t_2}, ..., e_{0,t_n})$ a vector with corresponding values of life expectancy at birth over time. We set $A = \{0, 1, ..., 100\}$ and $T = \{t_1, t_2, ..., t_n\}$. Then, for a hidden layer $\mathbf{H}^{(k)}$, the specific neural network structure illustrated in Fig. 1 for the logarithm of death rates is given by the following:

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

where $f^{(k)}$ is the activation function, $\mathbf{W^{(k)}}$ is the matrix of weights, $\mathbf{H^{(k-1)}}$ is the hidden layers, and $\mathbf{b^{(k)}}$ is the bias used to control the triggering value of the activation function. We use the rectified linear unit function given by $f(z) = \max(z, 0)$ (Glorot and Bengio, 2011). This function ensures faster learning in networks with many layers. In the feed-forward architecture, each hidden layer involves the previous one as shown below:

$$\mathbf{H^{(k)}} = f^{(k)}(\mathbf{W^{(k)}}\,\mathbf{H^{(k-1)}} + \mathbf{b^{(k)}}),$$

where $\mathbf{H^{(k-1)}}$ can be expressed as a function of a vector with life expectancy values as follows:

[1] Conventional machine-learning techniques were limited in their ability to process data, requiring careful engineering. DNNs provide higher flexibility, relying on the paradigm of representation learning, with multiple levels of representation, obtained by composing non-linear modules. From the input data, they build layer by layer, new sets of features, to make optimal predictions of target variables (for more details see:(Lecun and Bengio, 2015))

$$\mathbf{H^{(k-1)}} = f^{(k-1)}(\dots f^{(1)}(\mathbf{W^{(1)}}\,\mathbf{e_0} + \mathbf{b^{(1)}})\dots).$$

For example, for standard architecture consisting of three hidden layers $k = 3$ (input, hidden and output layer, respectively), the theoretical relationship defining the matrix of mortality $\mathbf{M}$ given the vector of life expectancy at birth $\mathbf{e_0}$ is represented by:

$$\mathbf{M} = f^{(3)}(\mathbf{W^{(3)}}(f^{(2)}(\mathbf{W^{(2)}}f^{(1)}(\mathbf{W^{(1)}}\,\mathbf{e_0} + \mathbf{b^{(1)}}) + \mathbf{b^{(2)}}) + \mathbf{b^{(3)}}), \tag{4}$$

where $f^{(1)}(\mathbf{W^{(1)}}\,\mathbf{e_0} + \mathbf{b^{(1)}}) = \mathbf{H^{(1)}}$ is the first hidden layer that accepts the vector $\mathbf{e_0}$ as input.

The DNN model is based on a training algorithm that involves an unconstrained optimisation problem aiming to minimise the prediction error. The idea is to adjust the weights of the network connections to minimise a measure of the difference between the actual and the desired output ($\mathbf{M}$ and $\hat{\mathbf{M}}$), respectively, known as the loss function $\mathcal{L}$. We use the Mean Square Error (MSE) as a loss function given by:

$$\mathcal{L}[\mathbf{M}, \hat{\mathbf{M}}] = \frac{1}{a \cdot t} \sum_{a,t} \left[\log(m_{a,t}) - \log(\hat{m}_{a,t})\right]^2$$

We chose the MSE because it is the benchmark in neural network regression problems (Lecun and Bengio, 2015), and was the best performer compared to other suitable loss functions with our dataset. To minimise the loss function, we use gradient descent optimisation. Gradient descent is one of the most popular algorithms used to perform optimisation and is the most common way to optimise neural networks. It consists of minimising the loss function by updating the weights in the opposite direction of the gradient ($\nabla\mathcal{L}$) with respect to the weights. For a generic set of weights $w_{n,n}^{(k)}$ and the k-th layer, using the chain rule, the gradient is given by:

$$\nabla\mathcal{L} = \frac{\partial\mathcal{L}[\mathbf{M}, \hat{\mathbf{M}}]}{\partial w_{n,n}^{(k)}} = \frac{\partial\mathcal{L}[\mathbf{M}, \hat{\mathbf{M}}]}{\partial H_n^{(k)}} \frac{\partial H_n^{(k)}}{\partial z_n^{(k)}} \frac{\partial z_n^{(k)}}{\partial w_{n,n}^{(k)}}, \tag{5}$$

where $z_n^{(k)} = w_n^{(k)}\, H_n^{(k-1)} + b_n^{(k)}$. The gradient encodes the relative importance of each weight and bias. The algorithm for efficiently computing the gradient in Eq. (5) is known as back-propagation. Back-propagation consists of a recursive algorithm. In the forward step, the prediction is computed by fixing the weights; subsequently, in the backward step, the weights are adjusted by back-propagating the gradient of the loss function to reduce the error. As a result of these adjustments, the internal hidden layers, which are not part of the input or output, are able to represent and capture important features of age specific mortality. To update the weights ($\tilde{\mathbf{W}}$), the gradient of the loss function is multiplied by a scalar, η, often called learning rate, according to the following scheme:

$$\tilde{\mathbf{W}} = \mathbf{W} - \eta\nabla\mathcal{L}\left[\mathbf{M}, \hat{\mathbf{M}}\right]. \tag{6}$$

The learning rate η determines the size of the step taken to reach a global or local minimum. In other words, gradient descent is similar to 'climbing down a hill' until a global or local minimum is reached. For this stage, we implement the Root Mean Square Propagation (RMSProp) algorithm proposed by Hinton and Swersky (2013).

Implementation

Our model requires an input vector with the time series of life expectancy at birth and a matrix with the corresponding age specific death rates over columns and time periods over rows. Each

data series is split into a training-validation set with which the network is trained, and a test set to check the accuracy of the model's prediction. The scheme is presented in Fig. 2.

[Figure 2 about here.]

To select the optimal number of hidden layers, neurons and parameters used in the DNN, it is common practice to perform a fine-tuning phase. The aim of this phase is to choose an appropriate structure of the model (number of hidden layers, neurons) according to the training error minimisation. This choice depends on the type of data that remains a heuristic problem in the field of neural networks. The process consists of feeding the model the training set and subsequently assessing its accuracy on the validation set. The test set stands for the unobserved time horizon, on which the model comparison will be performed. In practice, the test set would be the forecasted or projected life expectancy, for which the age-specific mortality profile is unknown. Formally, let t_τ, with $t_0 < t_\tau < t_s$, be the calendar year hat corresponds to the last realisation in the train-validation set. The values of life expectancy in the period (t_0, t_τ), $(e_{0,t})_{t\in[t_0,t_\tau]}$, represent the input for train-validation, while the corresponding output is $(\hat{m}_{a,t})_{a\in A,t=[t_0,t_\tau]}$. During the train-validation phase, the neural network weights are estimated and subsequently used in the test phase concerning the period $[t_{\tau+1}, t_s]$, which starts from $t_{\tau+1}$. The resulting DNN estimate is a point estimation that does not provide any information on the uncertainty given by $\hat{\mathbf{W}}$. This is one of the main limitations of deep learning algorithms, where the estimation of prediction intervals is still considered a big challenge (Khosravi et al. (2011) and Keren and Schuller (2018)). However, in our problem, an alternative solution is to perform the model on the lower and upper bound of the forecasted/projected life expectancy as pseudo confidence intervals to provide a measure of uncertainty.

5 Results

To assess the robustness of our method, we performed an out-of-sample test over three time windows (1950–1980, 1960–1990, and 1970–2000) used as train-validation sets (split according to the 80%–20% splitting rule randomly sampled), and the subsequent years of each time window (1981–1995, 1991–2005 and 2001–2015 respectively) as test sets. The model was applied to data from Italy, Japan, Russia and the USA by sex. We use a six hidden-layer architecture following the fine-tuning and compared the results from the DNN model with those obtained in Ševčíková et al. (2016)'s and the Linear-Link models. To ensure comparability, the results were smoothed using P-splines (He, 1999)[2]. We focus on the results pertaining to females during the period from 2001–2015 in this section; the results related to males in the same years and in the other study periods for both sexes are reported in Appendix Section A.

Age-specific mortality estimates

Figure 3 shows age specific death rates (in log scale) for females in Russia, Japan, Italy, and the USA. The observed (target) profile is shown with dots and estimated values from the models using the training period from 1970–2000, which correspond to DNN (red line), Linear-Link (green line) and Ševčíková et al. (2016) (blue line). The three models capture the general pattern of mortality, with a decreasing trend from birth to around age 15, and increasing linearly from around age 30. For Italy and a recent period in Russia, the Ševčíková et al. (2016) model tends to overestimate mortality at young ages, while the opposite is true for the Linear-Link in the

[2]Unsmoothed results are shown in the additional material. The smoothing step does not affect the models' accuracy ranking: the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) improvements after the smoothing are on average 1.55% and 1.57%, respectively.

case of Japan. The DNN model adequately captures the mortality patterns. However, the three models fail to accurately capture the sharp decrease from infancy in the cases of Italy and Russia.

[Figure 3 about here.]

Age-specific relative differences

We further analyse the accuracy of the models with the relative differences ($\Delta_{a,t}$) between estimates and the observed death rate by age for each model defined as (see Fig. 4):

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

Because the relative difference is calculated from the log-death rates, red hues indicate that the model underestimates mortality, while blue hues indicate overestimation. Differences with the observed mortality profile are small in general across models and countries. However, it observed a systematic underestimation in working-ages (20–50) for Russia and the USA from the DNN and Linear-Link models in the time window from 2001–2015 (the other two time windows are shown in Appendix Figs. A.4, A.5, A.6, A.7). Similarly, there is increased deviation at very old ages for recent periods from both models. In contrast, the Ševčíková et al. (2016) model tends to overestimate mortality at older ages for all countries, especially for recent periods in Russia.

[Figure 4 about here.]

Among males (Fig. A.1), the DNN and Linear-Link models tend to underestimate mortality at working-ages, which is compensated with increased mortality at younger and older ages below the age of 90. In contrast, the Ševčíková et al. (2016) model, as in the case for females, tends to overestimate mortality across all countries and is best used to capture the working-age pattern for Russia and the USA. Notably, deviations from the observed mortality increase with time across all time windows.

Mean Absolute Error (MAE) and Root Mean Square Error (RMSE)

To summarise the performance of the methods and to evaluate their accuracy, we report the MAE and RMSE on the test sets given by:

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

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

Tables 1 and 3 summarise the MAE and RMSE for the three models and four countries over the three time windows that we studied for females and males, respectively. For females, the DNN is the best performer in most cases, although the Linear-Link model showed the lowest MAE for Italy in the earliest and latest periods. The Ševčíková et al. (2016) model exhibited the lowest RMSE for Russia in the period from 1991–2005. The results for males (see Table 3) show less consistent results. For Italy and the USA, the DNN model consistently showed the lowest errors, but the Ševčíková et al. (2016) model performed the best for Japanese males.

Among males, for Italy and the USA the DNN model was the best performer in terms of MAE and RMSE. For Japan, as noted in the age-specific figures, the Ševčíková et al. (2016) model showed the lowest errors. For Russia, it was a mix between the three models depending

on the time window and summary measure. For both sexes, while the DNN model showed in the majority of cases the lowest departures from age-specific mortality, the Linear-Link model performed the best in capturing the life expectancy level.

[Table 1 about here.]

6 Multi population Model (mp-DNN)

In the previous section, we compared the performance of the DNN model with the Linear-Link and Ševčíková et al. (2016)'s models, which are the closest comparable models available. However, the DNN model can be used in a more general way in contexts for which there is an estimate of life expectancy but no available past age-specific mortality. In such contexts, information from neighbouring countries or from countries with similar mortality dynamics could be used to estimate an age-specific mortality profile. From this perspective, the time dimension is lost and not needed. Therefore, the DNN model fills the gap left by predecessor models by relying only on past data and becomes more akin to indirect methods or model lifetables.

Here, we present a simple example of how to extend the DNN model for the multiple populations case (mp-DNN). Consider the case in which the full HMD is used to train a model and then to predict the age-specific mortality profile of a country's life expectancy. We still model the functional relationship between life expectancy at birth and death rates, as numerical inputs and outputs. To extend the framework to the mp-DNN model by adding other demographic features, such as country, year and sex, we use what is known as embedding layers (Richman (2020); Bengio and Vincent (2013)). Embedding is a tool that allows for the capture of relationships that are very difficult to capture otherwise due to high dimensionality.

Results

Figures A.2 and 5 show age-specific death rates (in log scale) for females and males in Russia, Japan, Italy and the USA, which were estimated by exploiting the mp-DNN framework trained on the whole HMD. At first glance, the mp-DNN provides smoothed estimation by nature, due to a wider training sample. The model is able to describe the general mortality shape and provides a good fit for Italy and Japan, and remarkable accuracy for Russia, which represents a real challenge for the single population models. The USA provides a particular example where mp-DNN constantly underestimates mortality, at both young and older ages.

[Figure 5 about here.]

Figures A.3 and 6 provide the accuracy of the mp-DNN model with the relative differences ($\Delta_{a,t}$) between estimates and the observed death rate by age and time windows. Overall, mp-DNN shows small deviations with the observed mortality profile, across countries and periods, with the exception of the USA. In this case, the model provides inconsistent results, showing underestimations that increase over time. An alternative to treat atypical cases, such as the USA, could be used to train the model with data from countries with relatively high young mortality rates. Italy, in particular the female population, shows high age-specific sensitivity towards periods, shifting from overestimation to underestimation of the older ages. For Japan and Russia, as noted in the age-specific figures, the model provides reliable estimations, notably, deviations from the observed mortality decrease across all time windows.

Tables 2 and 4 show the MAE and RMSE for the multi-population model estimation for the four countries over the three study periods for both sexes. We can confirm the inadequacy of

this model to represent the USA mortality dynamics. However, we provide evidence of good accuracy for Italy and Japan, and accurate results in terms of errors referring to Russia. We underline that mp-DNN was able to provide reliable estimates for Russia over the first time window, where other models were not. Indeed, single population models, in order to estimate parameters, need training data that are not completely provided for Russia in the reference period from 1950–1980. The mp-DNN leverages the multi-population estimated parameters by applying them to the Russia case of the out-of-sample window (1981–1995).

[Figure 6 about here.]

[Table 2 about here.]

7 Conclusion

We presented a novel method to indirectly estimate a full mortality profile from a level of life expectancy at birth by leveraging deep neural networks using prior information on age specific mortality. When tested with state-of-the-art methodologies, the DNN model performed the best in many cases with fewer assumptions than previous methods. The method outlined here is non-parametric, data-driven and does not rely on assumptions that may not be completely accurate. Nevertheless, the results show that the three models tested here perform in a satisfactory way, with the DNN model offering the best performance in most cases. As shown, the reconstruction of an accurate mortality surface from a given level of demographic summary measure, such as life expectancy, is challenging. However, we offer a new alternative based on machine learning algorithms that complement the existing demographic toolbox. Moreover, the analysis of four countries with substantially different mortality and three sequential time windows of 30 years provided robust results. We confirm that the Linear-Link model, because of its dynamic constraint and consequential reparametrisation, assures coherence with respect to input life expectancy level (providing a negligible error), while for the Ševčíková et al. (2016) and DNN models, deviations were larger.

A substantial advantage of the DNN model appears in the multiple-population framework. While the Ševčíková et al. (2016) and Linear-Link models were designed to derive an age-specific mortality profile for a future value of life expectancy based on past data of a single population, the DNN model could be used to estimate age-specific mortality for countries where there is no data available using information from countries with similar mortality dynamics. Therefore, the DNN model fills the gap of the predecessors' models by relying only on past data and becomes more akin to indirect methods or model lifetables. We take a step forward among demographic methods, offering a multi-population indirect estimation based on a data driven approach, that can be fitted to many populations simultaneously, using DNN optimisation approaches. While we apply our methodology to country-specific scenarios, the model could be used to indirectly estimate mortality profiles for regions or subpopulations with similar mortality profiles. This characteristic makes our model appealing for countries where present information is lacking but past data are available from surrounding countries or populations, as we have shown in the Russian case.

We acknowledge that our model is subject to several limitations, including the choice of architecture (e.g., the number of hidden layers) and the parameters involved in the training phase. This remains a heuristic problem for neural network users, as indeed the choice often depends on the type of data and a preliminary round of fine-tuning, before the testing, which is highly desired, albeit, somewhat time-consuming (Nigri et al., 2019). This issue is dimmed in the case of multi-population. Indeed, this framework relies on a bigger data set, thus more

examples during the training phases are required, which has two implications. On the one hand, the model provides more robustness towards structural changes; however, some country-specific dynamics may not be captured, as in the case of the USA. Therefore, we strongly recommend the careful selection of the countries' subgroup on which the DNN model will be trained.

To conclude, here we propose a new approach that provides a valuable alternative tool to capture irregular mortality trajectories. While machine learning, deep or not, is not widely used in the field of demography, our method shows that it can be used to provide robust estimations of age-specific death rates. Due to its nature, our method can be leveraged in other demographic contexts, i.e. to derive age-specific fertility profiles from observed or predicted mean age at childbearing. This may foster new research at the frontier of demographic studies using innovative, yet simple to implement, techniques such as the DNN model. This is even more important in the context of rapid population aging and fast mortality decline, but also in contexts where lacking mortality estimates can provide crucial information for policy planning.

Material availability: The replication scripts, written in the R statistical programming language (R. Core Team, 2013), are hosted on the Open Science Framework (OSF) at `https://osf.io/5dynw/?view_only=badbcfa0e67b4d98a78560acb50f0b8c`

References

Aburto, J. M., Beltrán-Sánchez, H., and García-Guerrero, Victor Manuel & 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.

Aburto, J. M., Kashyap, R., Schöley, J., Angus, C., Ermisch, J., and Mills, Melinda C & Dowd, J. B. (2021). Estimating the burden of the covid-19 pandemic on mortality, life expectancy and lifespan inequality in england and wales: a population-level analysis. *Journal of Epidemiology & Community Health*.

Bengio, Y. and Vincent, A. C. . P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828.

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., and Romaniuk, A & Van de Walle, E. (1968). *The Demography of Tropical Africa*. Princeton University Press.

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

Clark, S. J. (2019). A general age-specific mortality model with an example indexed by child mortality or both child and adult mortality. 56(3):1131–1159.

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.

Glorot, X. and Bengio, A. B. . Y. (2011). Deep sparse rectifier neural networks. In Gordon, G., Dunson, D., and Dudík, M., editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 315–323, Fort Lauderdale, FL, USA. JMLR Workshop and Conference Proceedings.

He, X. & Ng, P. (1999). Cobs: Qualitatively constrained smoothing via linear programming. *Computational Statistics*, 14:315–337.

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

Hinton, G. and Swersky, N. S. . K. (2013). Neural networks for machine learning. lecture 6a: overview of mini-batch gradient descent. *Department of Computer Science University of Toronto*.

HMD (2021). University of california, berkeley (usa), and max planck institute for demographic research (germany). Data downloaded on 01/01/2019.

Keren, G. and Schuller, N. C. . B. (2018). Calibrated prediction intervals for neural network regressors. *IEEE Access*, 6:54033–54041.

Khosravi, A., Nahavandi, S., and Atiya, D. C. . A. F. (2011). Comprehensive review of neural network-based prediction intervals and new advances. *IEEE Transactions on Neural Networks*, 22(9):1341–1356.

Lecun, Y. and Bengio, Yoshua & Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

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*, 3:19–39.

Lee, R. & Miller, T. (2001). Evaluating the performance of the lee-carter method for forecasting mortality. *Demography*, 38(6):537–549.

Li, N. and Lee, Ronald & 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.

Luy, M. (2003). Causes of male excess mortality: Insights from cloistered populations. *Population and Development Review*, 29(4):647–676.

Mehta, N. K. and Abrams, Leah R & Myrskylä, M. (2020). US life expectancy stalls due to cardiovascular disease, not drug deaths. *Proceedings of the National Academy of Sciences*, 117(13):6998–7000.

Montavon, G. and Múller, W. S. . K. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15.

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

Nigri, A., Levantesi, S., Marino, M., and Perla, S. S. . F. (2019). A deep learning integrated lee-carter model. *Risks*, 7(1).

Nigri, A. and Marino, S. L. . M. (2021). Life expectancy and lifespan disparity forecasting: a long short-term memory approach. *Scandinavian Actuarial Journal*, 2021(2):110–133.

Pascariu, M. D., Basellini, U., and Canudas-Romo, J. M. A. . V. (2020). The linear link: Deriving age-specific death rates from life expectancy. volume 8.

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

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

Richman, R. (2020). AI in actuarial science – a review of recent advances – part 1. *Annals of Actuarial Science*, page 1–23.

Rumelhart, D. and Williams, G. H. . R. (1986). Learning representations by back-propagating errors. *Nature*, 323.

Torri, Tiziana & 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.

United Nations (2019). *World Population Prospects 2019: Methodology of the United Nations population estimates and projections*. Department of Social Affairs.

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

Ševčíková, H., Li, N., Kantorová, V., and Raftery, P. G. . A. E. (2016). Age-specific mortality and fertility rates for probabilistic population projections. In *Dynamic Demographic Analysis*, volume 39, pages 285–310. Springer Series on Demographic Methods and Population Analysis.

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

Figures

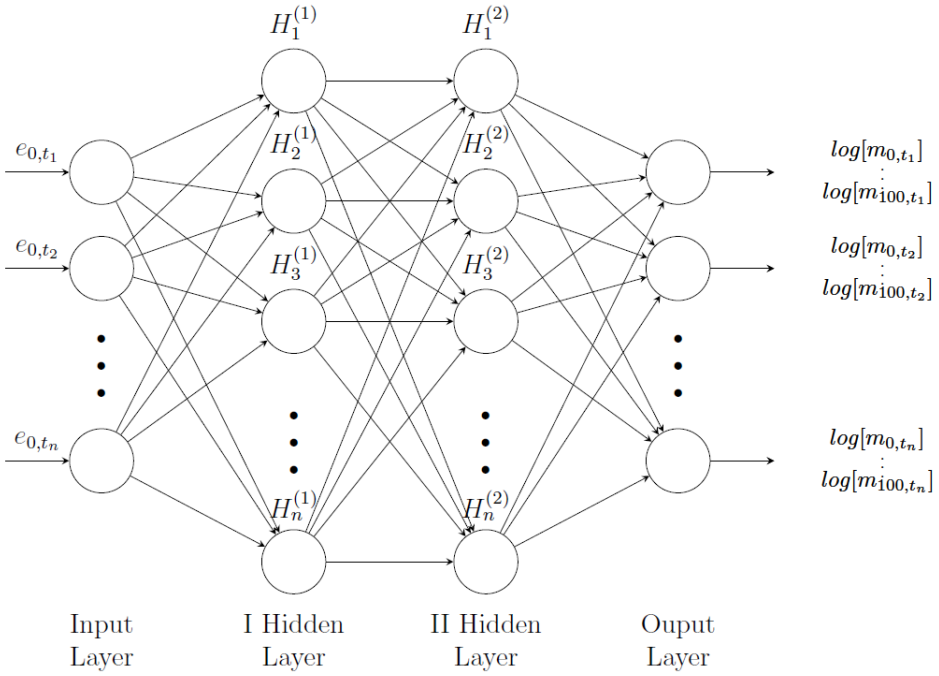

Figure 1: Graphical representation of our DNN model. The input is a vector of life expectancies over time $e_{0,t}$, which passes through the neurons and all multiple layers. The output in this diagram is a set of log-death rates at each age that correspond to each value of life expectancy, trained with observed age-specific data.

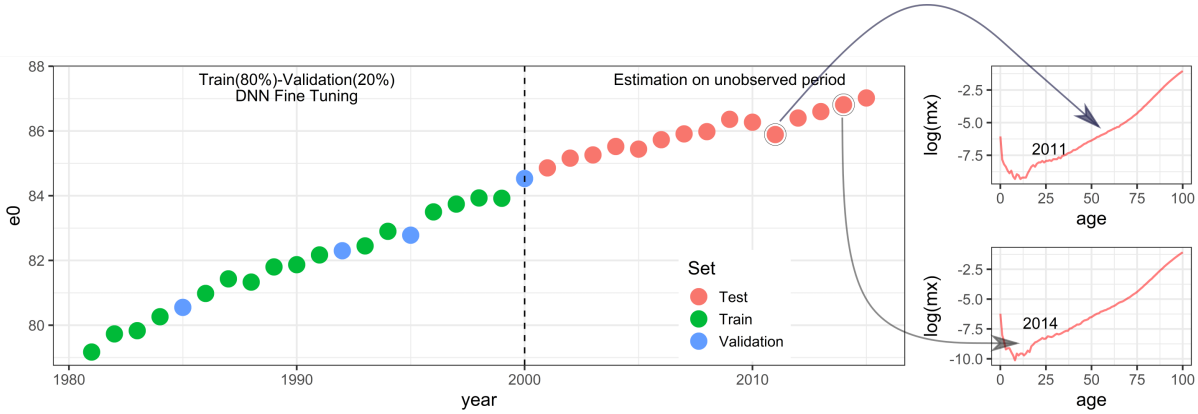

Figure 2: Implementation scheme of the DNN model. The model is trained and validated with observed age-specific death rates that are consistent with life expectancy levels from the train and validation period (green and blue dots). The results from this training phase are then applied to estimate a full age-specific mortality profile for a given value of (projected or forecasted) life expectancy (orange dots).

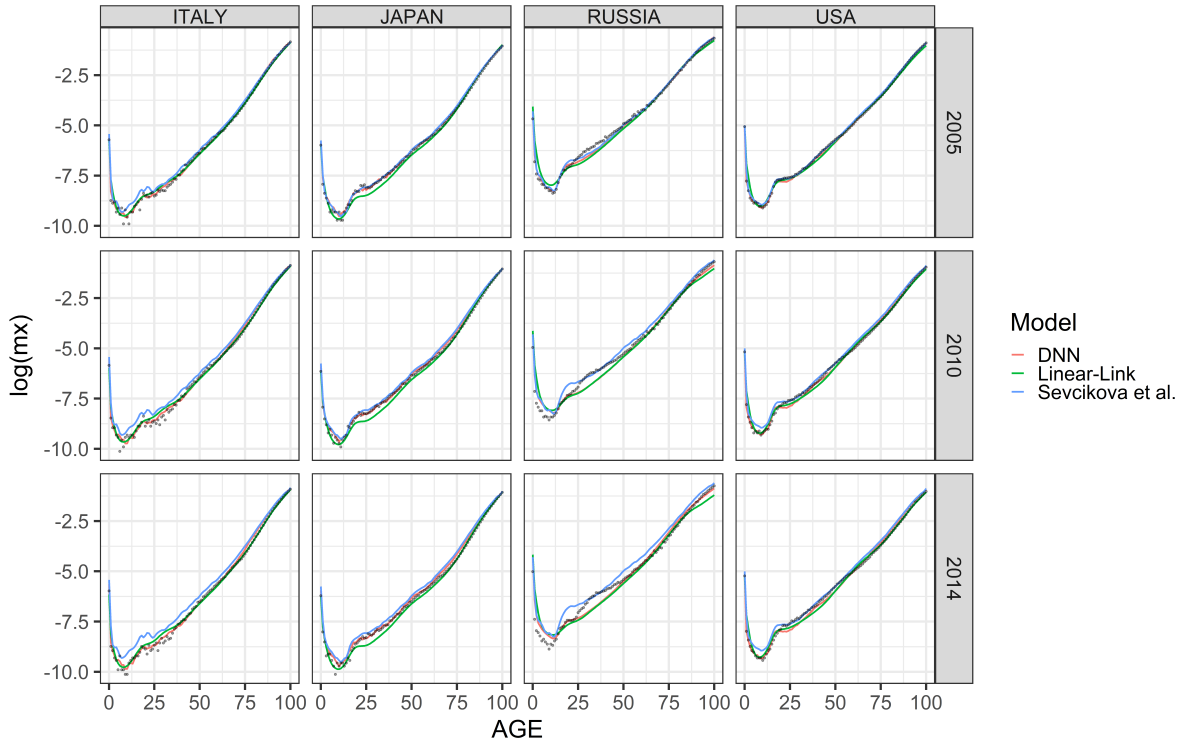

Figure 3: Estimated age-specific female log-mortality rates $\log(m_{a,t})$ for three models: DNN, Linear-Link and Ševčíková et al. (2016), by country, for 2005, 2010 and 2014 based on the training period from 1970–2000. The black dots are the observed log-mortality rates.

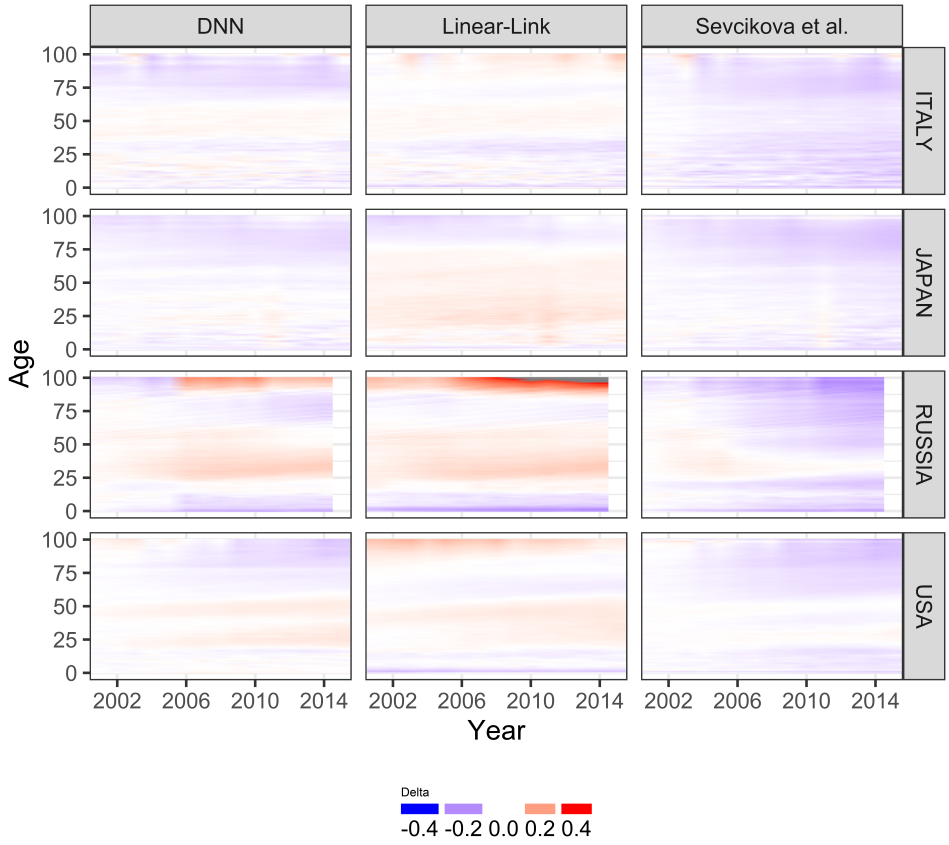

Figure 4: Relative differences ($\Delta_{a,t}$) between estimates and the observed death rate by age for each model. Red hues indicate that the model underestimates mortality, while blue hues indicate overestimation. The female test period took place between 2001–2015.

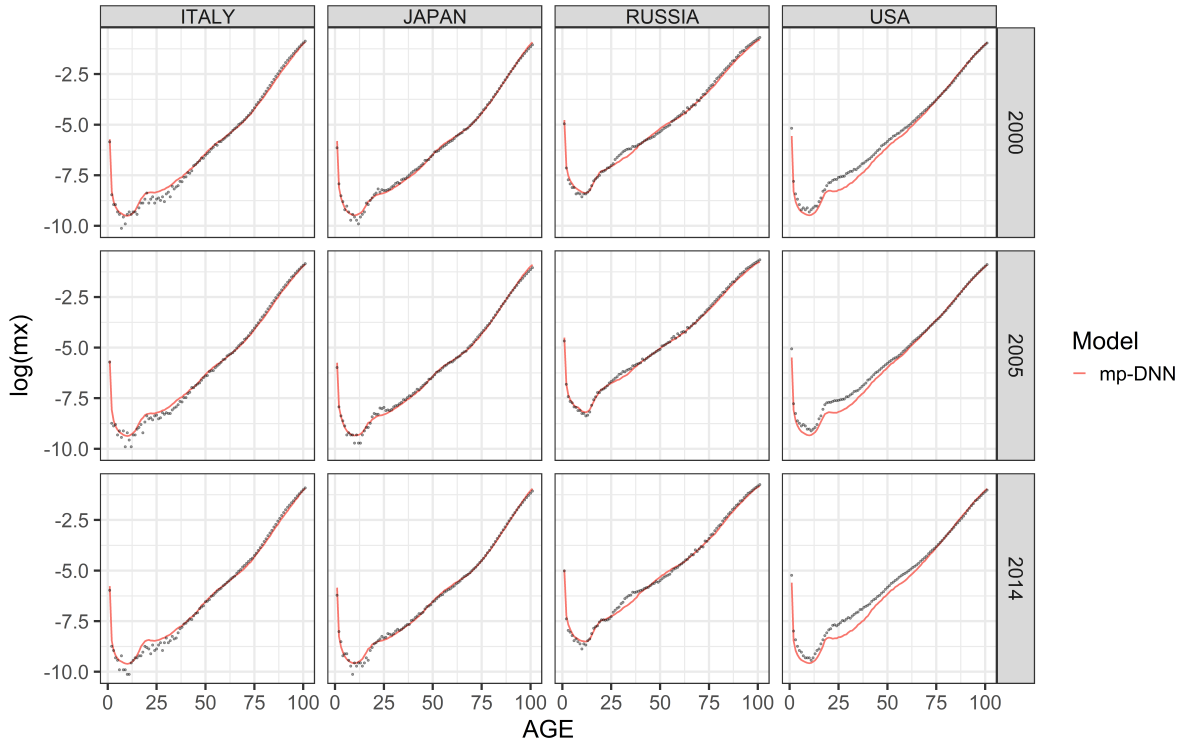

Figure 5: Estimated age-specific female log-mortality rates $\log(m_{a,t})$ for the mp-DNN model by country for 2005, 2010 and 2014 based on the training period from 1970–2000. Black dots are the observed log-mortality rates.

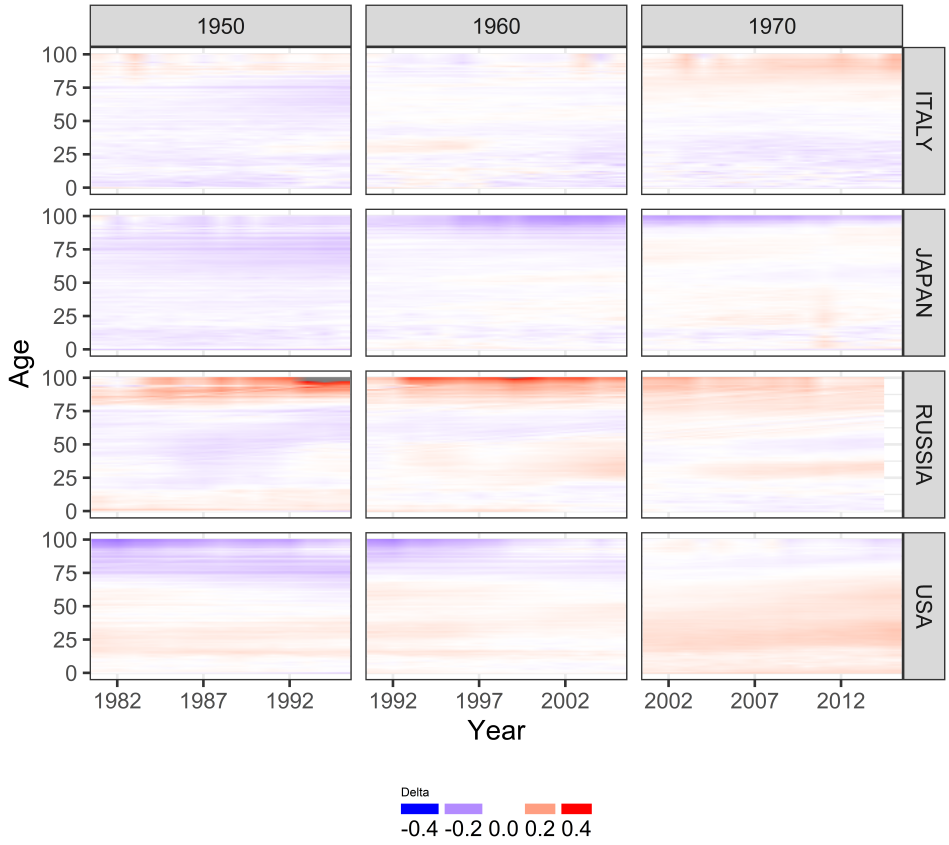

Figure 6: Relative differences ($\Delta_{a,t}$) between estimates and the observed death rate by age for the mp-DNN model. Red hues indicate that the model underestimates mortality, while blue hues indicate overestimation. The female test period took place between 2001–2015.

Tables

Table 1: Out-of-sample test: MAE and RMSE for DNN, Linear-Link, and Ševčíková et al. (2016) by country and sex. The estimation period for females took place between 1981–1995 (columns 3–4), 1991–2005 (columns 5–6) and 2001–2015 (columns 7–8).

Country	**Model**	**1981–1995**		**1991–2005**		**2001–2015**	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
Italy	DNN	0.1672	**0.2017**	**0.1038**	**0.1418**	0.1210	**0.1591**
	Linear-Link	**0.1379**	0.2325	0.1561	0.2272	**0.1154**	0.1818
	Ševčíková et al. (2016)	0.2421	0.2995	0.2042	0.2598	0.2539	0.3278
Japan	DNN	**0.1730**	**0.2055**	**0.1182**	**0.1532**	**0.1137**	**0.1436**
	Linear-Link	0.4036	0.5473	0.2602	0.3242	0.1664	0.2095
	Ševčíková et al. (2016)	0.2444	0.2785	0.1906	0.2376	0.1656	0.2014
USA	DNN	**0.0572**	**0.0788**	**0.0873**	**0.1119**	**0.0944**	**0.1206**
	Linear-Link	0.0720	0.1196	0.0949	0.1541	0.1183	0.1683
	Ševčíková et al. (2016)	0.1291	0.1642	0.1081	0.1524	0.1227	0.1643
Russia (2014)	DNN	-	-	**0.1714**	0.2544	**0.1670**	**0.2393**
	Linear-Link	-	-	0.1942	0.2736	0.2266	0.3223
	Ševčíková et al. (2016)	-	-	0.1931	**0.2510**	0.1907	0.2558

Table 2: Out-of-sample test: MAE and RMSE for mp-DNN by country and sex. The estimation period for females took place between 1981–1995 (columns 3–4), 1991–2005 (columns 5–6) and 2001–2015 (columns 7–8).

Country	**Model**	**1981–1995**		**1991–2005**		**2001–2015**	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
Italy	mp-DNN	0.1380	0.1827	0.1024	0.1592	0.1395	0.1897
Japan	mp-DNN	0.1707	0.2026	0.09544	0.1215	0.09131	0.1314
USA	mp-DNN	0.1462	0.177	0.107	0.1336	0.2397	0.3115
Russia (2014)	mp-DNN	0.144	0.1728	0.1124	0.1382	0.1003	0.1265

A Appendix

[Figure 7 about here.]

[Figure 8 about here.]

[Figure 9 about here.]

[Table 3 about here.]

[Table 4 about here.]

[Figure 10 about here.]

[Figure 11 about here.]

[Figure 12 about here.]

[Figure 13 about here.]

Figures

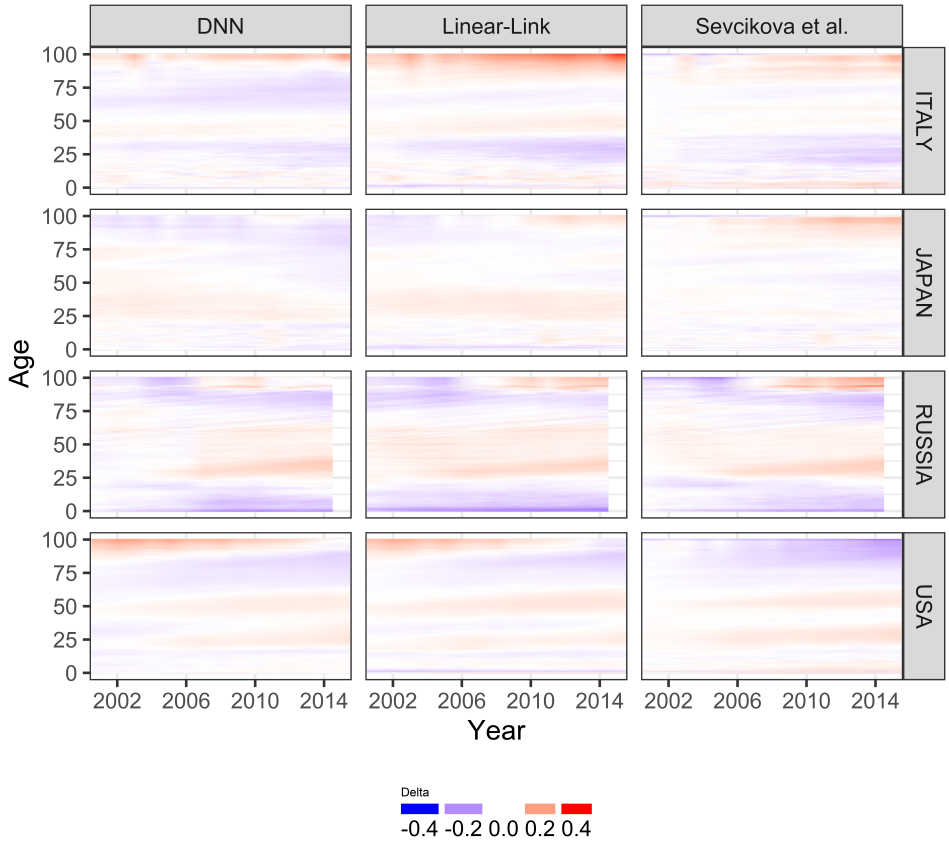

Figure A.1: Relative differences ($\Delta_{a,t}$) between estimates and the observed death rate by age for each model. Red hues indicate that the model underestimates mortality, while blue hues indicate overestimation. The male test period took place between 2001–2015.

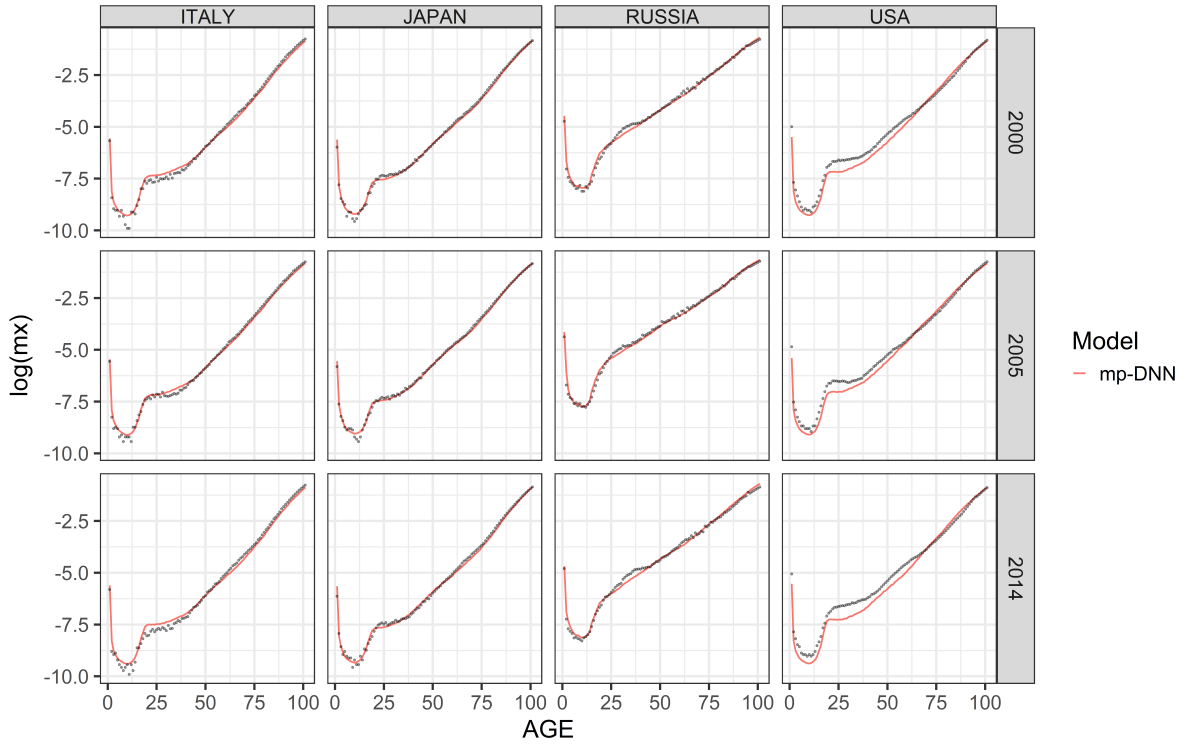

Figure A.2: Estimated age-specific male log-mortality rates $\log(m_{a,t})$ for the mp-DNN model by country for 2005, 2010 and 2014 based on the training period from 1970–2000. Black dots are the observed log-mortality rates.

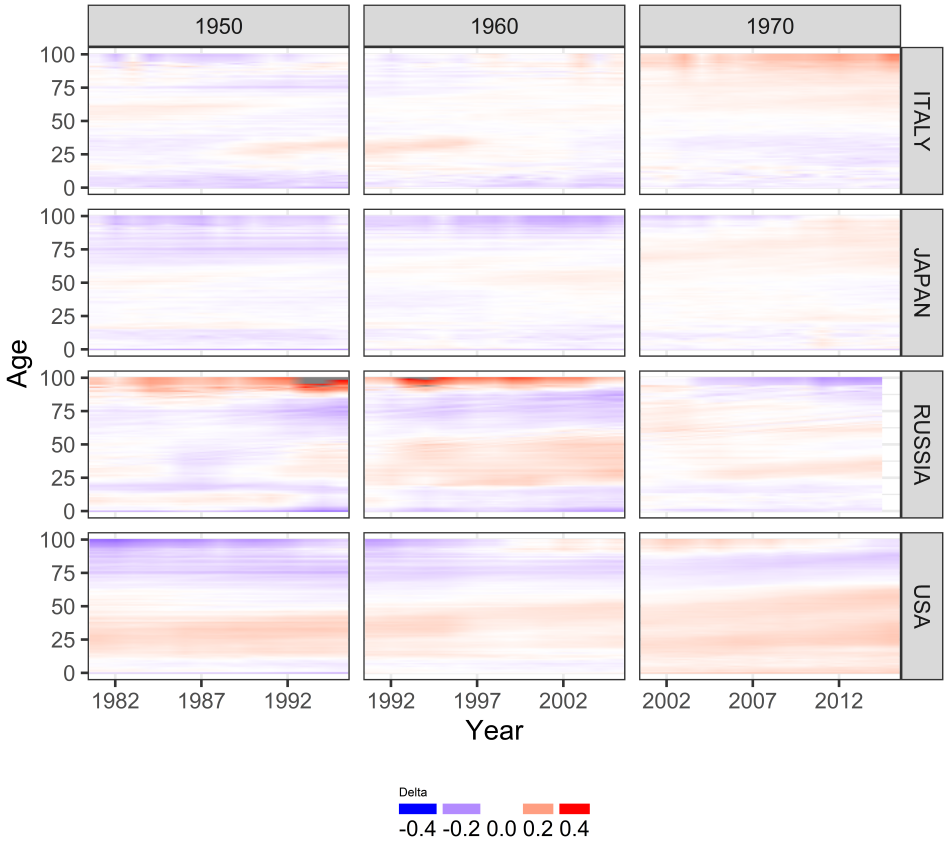

Figure A.3: Relative differences ($\Delta_{a,t}$) between estimates and the observed death rate by age for the mp-DNN model. Red hues indicate that the model underestimates mortality, while blue hues indicate overestimation. The male test period took place between 2001–2015.

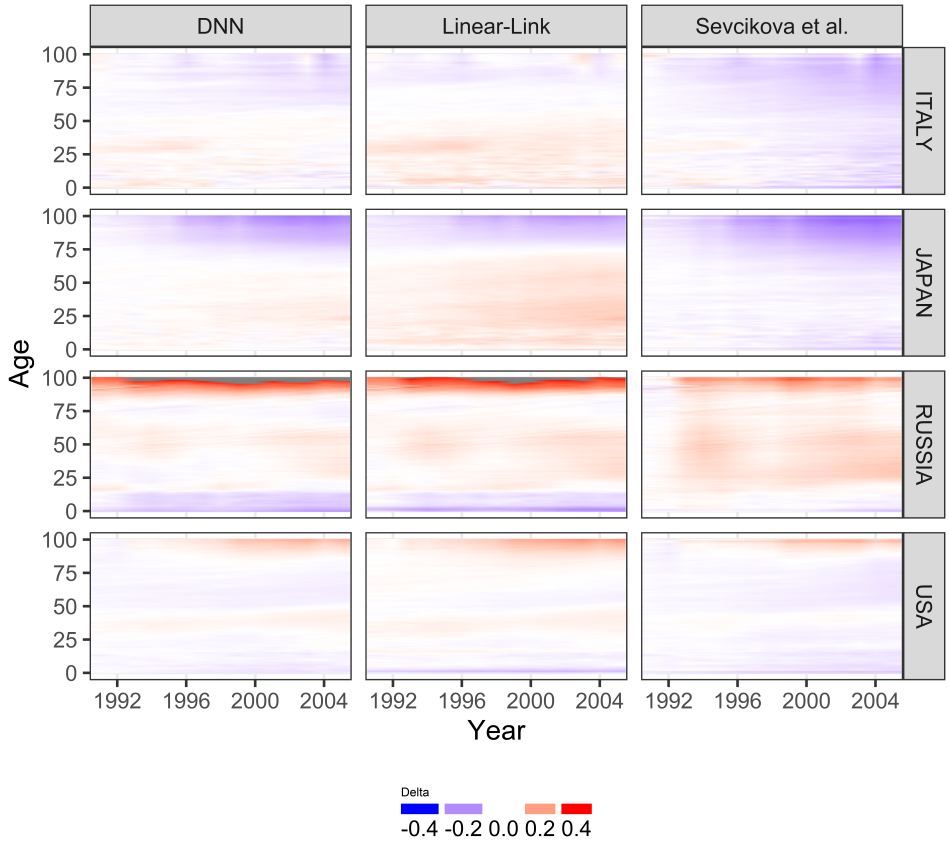

Figure A.4: Relative differences ($\Delta_{a,t}$) between estimates and the observed death rate by age for each model. Red hues indicate that the model underestimates mortality, while blue hues indicate overestimation. The female test period took place between 1991–2005.

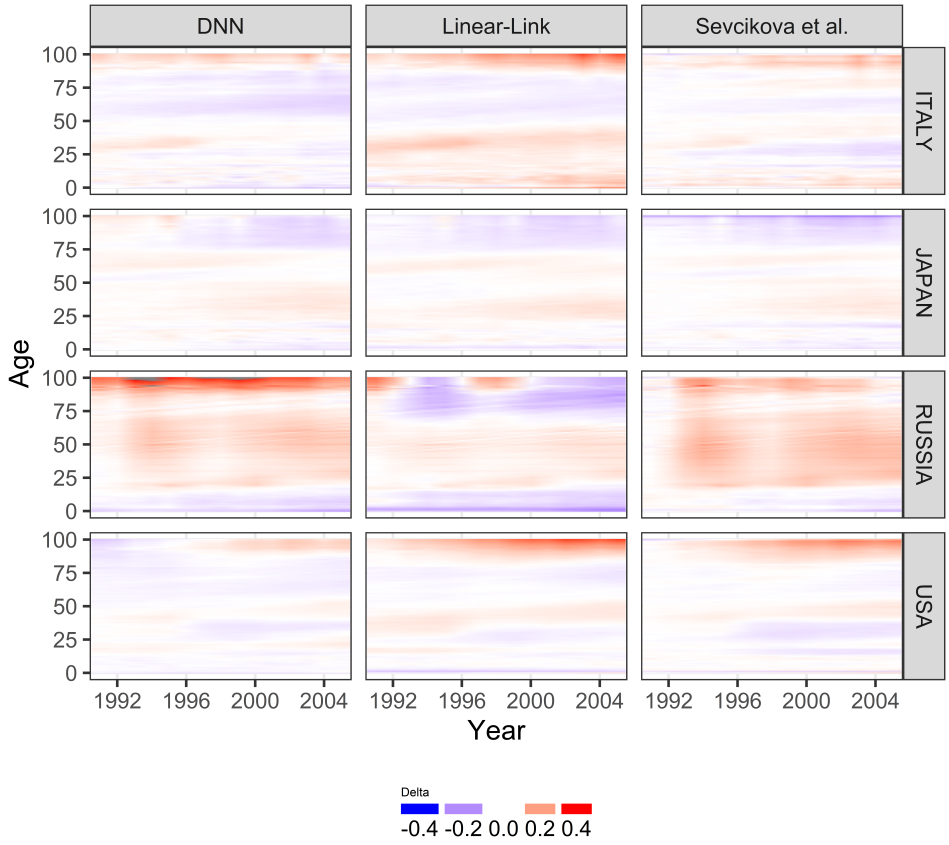

Figure A.5: Relative differences ($\Delta_{a,t}$) between estimates and the observed death rate by age for each model. Red hues indicate that the model underestimates mortality, while blue hues indicate overestimation. The male test period took place between 1991–2005.

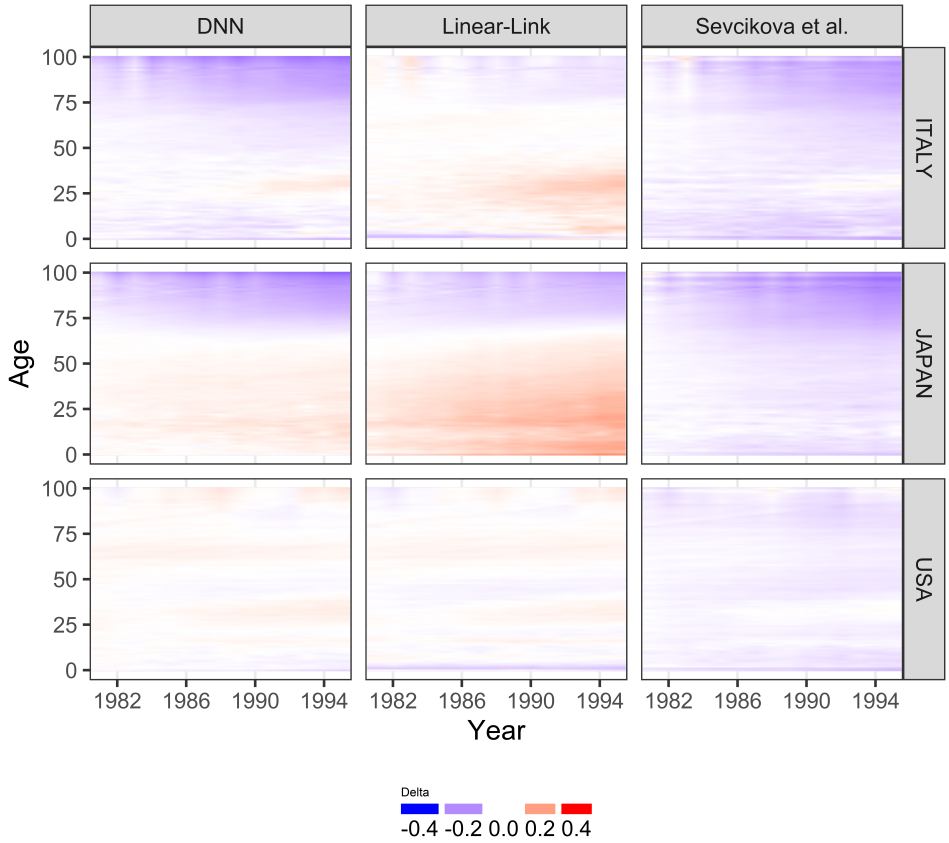

Figure A.6: Relative differences ($\Delta_{a,t}$) between estimates and the observed death rate by age for each model. Red hues indicate that the model underestimates mortality, while blue hues indicate overestimation. The female test period took place between 1981–1995.

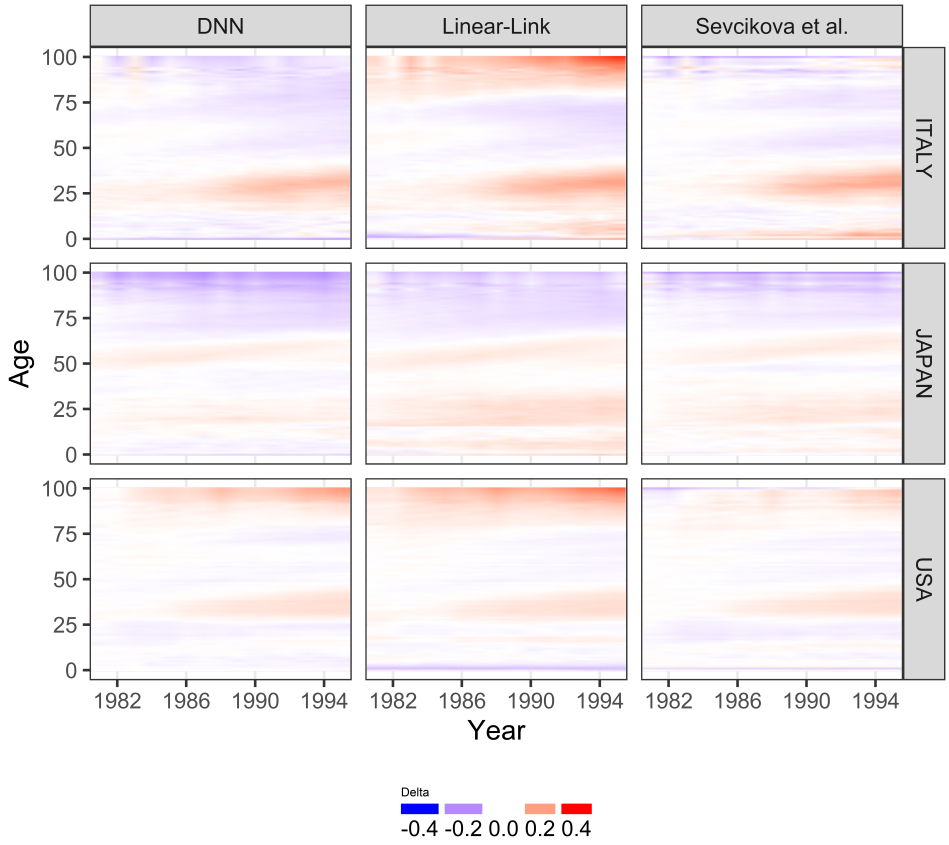

Figure A.7: Relative differences ($\Delta_{a,t}$) between estimates and the observed death rate by age for each model. Red hues indicate that the model underestimates mortality, while blue hues indicate overestimation. The male test period took place between 1981–1995.

Tables

Table 3: Out-of-sample test: MAE and RMSE for DNN, Linear-Link, and Ševčíková et al. (2016) by country and sex. The estimation period for males took place between 1981–1995 (columns 3–4), 1991–2005 (columns 5–6) and 2001–2015 (columns 7–8).

Country	**Model**	**1981–1995**		**1991–2005**		**2001–2015**	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
Italy	DNN	**0.1419**	**0.2203**	**0.1107**	**0.1493**	**0.1194**	**0.1566**
	Linear-Link	0.1840	0.2846	0.1993	0.2780	0.1541	0.2158
	Ševčíková et al. (2016)	0.1477	0.2596	0.1112	0.1701	0.1274	0.1992
Japan	DNN	0.1049	**0.1241**	0.0943	0.1264	0.0986	0.1254
	Linear-Link	0.1585	0.2088	0.1014	0.1392	0.1011	0.1480
	Ševčíková et al. (2016)	**0.1042**	0.1348	**0.0874**	**0.1205**	**0.0684**	**0.0958**
USA	DNN	**0.0746**	**0.1088**	**0.0730**	**0.0907**	**0.0955**	**0.1127**
	Linear-Link	0.1020	0.1561	0.1029	0.1437	0.1029	0.1367
	Ševčíková et al. (2016)	0.07974	0.1218	0.0907	0.1310	0.1085	0.1437
Russia (2014)	DNN	-	-	0.1989	**0.2366**	**0.1497**	0.2412
	Linear-Link	-	-	**0.1739**	0.2544	0.1951	0.3067
	Ševčíková et al. (2016)	-	-	0.2287	0.2940	0.1515	**0.2246**

Table 4: Out-of-sample test: MAE and RMSE for mp-DNN by country and sex. The estimation period for males took place between 1981–1995 (columns 3–4), 1991–2005 (columns 5–6) and 2001–2015 (columns 7–8).

Country	**Model**	**1981–1995**		**1991–2005**		**2001–2015**	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
Italy	mp-DNN	0.1276	0.1971	0.0972	0.1659	0.1257	0.1603
Japan	mp-DNN	0.117	0.1683	0.0875	0.1232	0.08049	0.1095
USA	mp-DNN	0.2093	0.2614	0.1451	0.1805	0.2535	0.3038
Russia (2014)	mp-DNN	0.1356	0.1854	0.1849	0.2267	0.08918	0.1216