



HAL
open science

Mortality data correction in the absence of monthly fertility records

Alexandre Boumezoued, Amal Elfassihi

► **To cite this version:**

Alexandre Boumezoued, Amal Elfassihi. Mortality data correction in the absence of monthly fertility records. 2020. hal-02634631

HAL Id: hal-02634631

<https://hal.science/hal-02634631>

Preprint submitted on 27 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mortality data correction in the absence of monthly fertility records

Alexandre Boumezoued^a Amal Elfassihi^b

Milliman R&D^c

January, 2020

Abstract

Since the conjecture of [Richards \(2008\)](#), the work by [Cairns et al. \(2016\)](#) and subsequent developments by [Boumezoued \(2016\)](#), [Boumezoued et al. \(2018\)](#) and [Boumezoued et al. \(2019\)](#), it has been acknowledged that observations from censuses have led to major problems of reliability in estimates of general population mortality rates as implemented in practice. These issues led to mis-interpretation of some key mortality characteristics in the past decades, including “false cohort effects”. To overcome these issues, the exposure estimates for a given country can be corrected by using monthly fertility records. However, in the absence of birth-by-month data, the recent developments are not applicable. Therefore, this paper explores new solutions regarding the construction of mortality tables in this context, based on machine learning techniques. As a main result, it is demonstrated that the new exposure models proposed in this paper allow to provide correction with high quality and to improve the fitting of stochastic mortality models without cohort component, as it is the case for the existing correction method based on monthly fertility data.

Keywords — Human Mortality Database, cohort effect, anomalous mortality data, stochastic mortality models, machine learning, neural network

^aEmail: alexandre.boumezoued@milliman.com

^bEmail: amal.elfassihi@milliman.com

^cMilliman, 14 Avenue de la Grande Armée, 75017 Paris, France.

1 Introduction

Studies that are concerned with mortality modelling usually make the assumption that the underlying data (e.g. deaths and exposures) are accurate. However, as shown by the recent February 2018 update of the Human Mortality Database (HMD, 2018), the database which contains original life tables for almost 40 countries or areas, proposing high quality mortality rate estimates remains a challenging topic. Awareness of some data anomalies emerged following Richards' conjecture (Richards, 2008), which had suggested that some cohort effects in England might be due to the sensitivity to fertility shocks of the method of calculating mortality rates. This conjecture was confirmed by Cairns et al. (2016) who proposed a method for correcting mortality data for England & Wales based on quarterly birth rate data. The *Convexity Adjustment Ratio* introduced in their work has then been adapted by Boumezoued (2016) who focused on HMD data and showed that these anomalies are universal. To build corrected mortality tables, a link with the Human Fertility Database (HFD, 2018), the HMD counterpart for fertility, has been made to correct such anomalies in a systematic way. Boumezoued et al. (2018) then introduced an inference strategy from a deterministic population dynamics model in order to compute mortality rates based on information extracted from censuses. Their work confirmed from a mathematical perspective the efficiency of using additional monthly fertility data for appropriately computing annual mortality tables. Moreover, a general inference theory in a stochastic setting has been proposed in Boumezoued et al. (2019).

Finally, the latest version of the HMD Methods Protocol (Wilmoth et al., 2019) introduces a change to the way mortality rates are constructed in the database. Birth-by-month data that have been collected are now used to more accurately estimate population exposures.

However, these methods are not applicable to countries for which birth-by-month data are not available, so that until now, most studies dedicated to stochastic mortality modelling including a cohort component had focused on reproducing what is known now to be data anomalies and false cohort effects. The purpose of this paper is therefore to propose a model to produce corrected mortality tables for countries for which birth-by-month data are not available, by learning the features of the correction process where it has already been applied.

In the literature on mortality modelling, machine learning approaches have recently emerged. The work of Deprez et al. (2017) has shown that machine learning algorithms are useful for assessing the quality of fit of mortality estimates provided by standard stochastic mortality models. The authors applied a gradient boosting model to analyze how modelling should be improved based on an individual's characteristics, such as age or birth cohort. This regression approach (non-parametric) then makes it possible to detect the weaknesses of different stochastic mortality models. Hainaut (2018) used neural networks to find the latent mortality factors of a Lee-Carter model and predict them according to a random walk with drift. Richman and Wüthrich (2019) extended the Lee-Carter model to multiple populations using deep neural networks. Ludkovski et al. (2018) used Gaussian processes to build a one-population model on mortality data from the United States, Japan, and England & Wales, their main objective being to obtain short-term mortality projections. The work of

Levantesi and Pizzorusso (2019) examined the ability of machine learning to improve the accuracy of some standard stochastic mortality models, both in estimating and predicting mortality rates. The authors used tree-based machine learning techniques to calibrate a parameter (the machine learning estimator) to be applied to mortality rates adjusted by the standard mortality model. Following this line of research, Levantesi and Nigri (2019) proposed an approach for the machine learning estimator forecasting, based on the combination of a random forest algorithm and two-dimensional P-splines. They used mortality data of a set of developed countries and showed how machine learning algorithms can bring effective benefits to the study of mortality, both in the fitting and forecast phases.

In the present paper, focus is rather on past mortality data correction. For such purpose, a refined feature engineering is developed by studying the observables available in the Lexis diagram. Based on the new features, a neural network approach is implemented as well as a baseline linear regression. To the best of our knowledge, the only attempt to provide automatic correction has been proposed by Cairns et al. (2016). One of their contributions was to develop a Bayesian model to quantify the magnitudes of the errors on the exposures, relying on original death counts and exposures estimates for England & Wales. The approach we propose will rather benefit from the recent correction methods developed and especially the HMD version 6 update to learn the correction process from a variety of countries, which is a core feature of our work. As far as we know, no other predictive study focused on correcting population exposures estimates without monthly fertility data.

The remainder of this paper is organised as follows. Section 2 specifies the correction problem and the observables in the Lexis diagram. Section 3 presents the correction method of exposure estimates. Section 4 exposes results of the method and Section 5 applies the method to correct mortality data for West Germany, Ukraine, Belarus and Poland, for which birth-by-month data is lacking for part of the cohorts and which therefore remain currently uncorrected for both demographic and actuarial studies. Finally, Section 6 concludes with a discussion and avenues for future research.

2 The correction of mortality estimates

2.1 The correction problem

A cohort effect appears when a generation has very different longevity characteristics from adjacent generations, for instance when different distributions of disease arise from a changing or new environmental cause affecting age groups differently. In epidemiology, a cohort effect is also conceptualised as a period effect that is differentially experienced through age-specific exposure or susceptibility to that event or cause. In other words, a cohort effect could arise when a population-level environmental cause is unequally distributed in the population (Keyes et al., 2010).

On the contrary, sudden variations in births can produce anomalous cohort effects. In fact, the mortality rate $\mu(a, t)$ at age a and at year t is estimated in the so-called period mortality table by comparing the number of deaths $D(a, t)$ occurring in the year t , with the

corresponding exposure to risk $E(a, t)$. Exposure to risk represents the quantity of individuals at risk of death, *i.e.* the total time lived by the population in the period considered. The mortality rate estimator is therefore given by:

$$\hat{\mu}(a, t) = \frac{D(a, t)}{E(a, t)}. \quad (1)$$

Let us define the improvement rates by:

$$r(a, t) = \frac{\mu(a, t + 1) - \mu(a, t)}{\mu(a, t)}, \quad (2)$$

which are often used to observe particular patterns. Then, if the mortality improvement rates matrix is represented for West Germany (see Figure 15), as extracted from the latest Human Mortality Database (HMD) update, cohort effects can be observed for specific generations (born around 1915, 1920 and 1940).

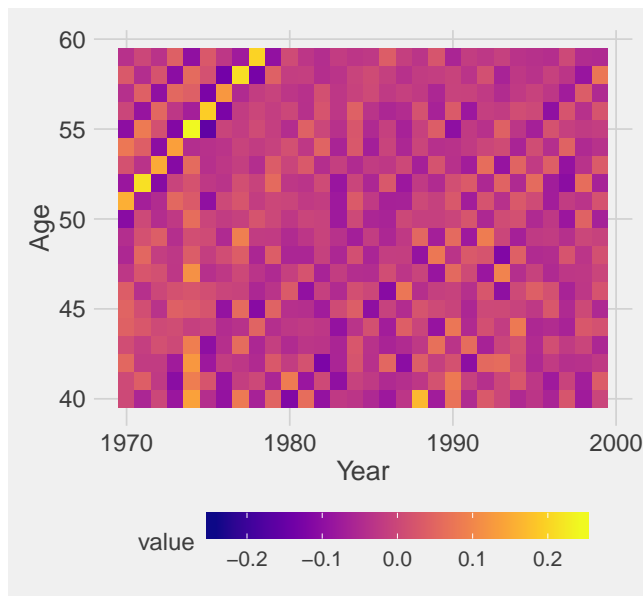


Figure 1: Mortality improvement rates — West Germany raw data

As age-specific exposure to risk in the period mortality table is approximated by the average between the population at the beginning and end of the year, under the assumption of uniform distribution of births, birth fluctuations at these times make the approximation inaccurate. Such issues were noticed in the HMD as early as 2007 by [Wilmoth et al. \(2007\)](#). Based on monthly fertility data, it is possible to improve the approximation made to the level of exposure to risk, and thus correct the death rate estimates, see again [Boumezoued \(2016\)](#) and [Wilmoth et al. \(2019\)](#). This process eliminates the anomalous cohort effects and produces more precise mortality tables, which we call “corrected mortality tables”.

However, birth-by-month data is not available for many countries for which the HFD monthly fertility records are provided starting at years greater than 1918 (see Figure 2). It

thus limits the set of cohorts for the correction method to be applicable. Hence the need to find a new correction method to correct mortality tables for a larger set of countries.

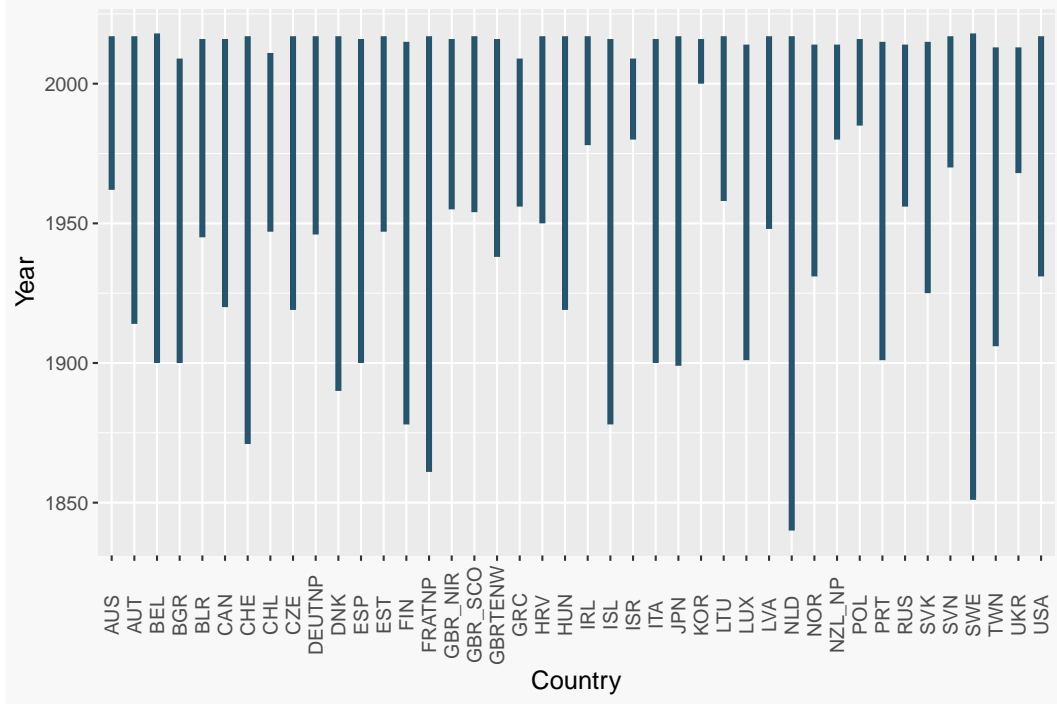


Figure 2: Birth-by-month data availability in the HMD/HFD

2.2 Observables in the Lexis diagram

Let us consider the Lexis diagram (Lexis, 1875), see Figure 3, consisting of a Cartesian coordinates system (t, a) , where t denotes the year at which an individual is observed and a denotes the age of the individual. Let us also denote by $g(a, t)$ the population density at (a, t) .

Following estimates can then be observed in the Lexis diagram:

- The population at (exact) time t , with (integer) age a at its last birthday: $P(a, t) = \int_a^{a+1} g(x, t) dx$,
- The number of individuals who attained (exact) age a during the year $[t, t + 1)$: $N(a, t) = \int_t^{t+1} g(a, s) ds$,
- The annual birth counts $B(t) = \int_t^{t+1} g(0, s) ds$.

Also, death counts are provided on the upper (T_U) and lower (T_L) triangles of the Lexis diagram:

$$D_U(a, t) = \iint_{T_U(a,t)} \mu(x, s)g(x, s) dx ds \text{ and } D_L(a, t) = \iint_{T_L(a,t)} \mu(x, s)g(x, s) dx ds, \quad (3)$$

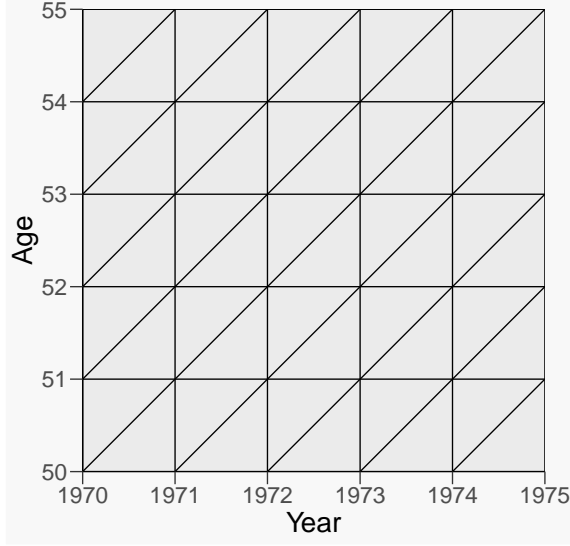


Figure 3: Lexis diagram

where:

$$T_U(a, t) = \{(x, s) \mid x \in [a, a + 1) \text{ and } s \in [t, t - a + x)\},$$

$$T_L(a, t) = \{(x, s) \mid x \in [a, a + 1) \text{ and } s \in [t - a + x, t + 1)\}.$$

Thus, the death counts for 1×1 Lexis squares is given by $D(a, t) = D_U(a, t) + D_L(a, t)$.

Data for Males from the HMD version 5 (see <https://v5.mortality.org>) and the most recent version 6 of the HMD (see <https://www.mortality.org>) are considered. Our choice for the HMD version 6 as the reference corrected tables is driven by the fact that those are publicly available in their final form, so that results of our paper can be reproduced. Note that alternative (and more general) formulas for the death rate estimator have been proposed by Boumezoued et al. (2018); this choice would lead to different correction results compared to those presented in our paper.

Let k be the country variable, $\mu^{(5)}$ and $\mu^{(6)}$ the period mortality rates from the version 5 and the version 6 of the HMD respectively, defined by:

$$\mu^{(v)}(a, t, k) = \frac{D^{(v)}(a, t, k)}{E^{(v)}(a, t, k)}, v \in \{5, 6\},$$

with $E^{(v)}(a, t, k)$ and $D^{(v)}(a, t, k)$ denoting respectively the exposure to risk and the number of deaths relatively to the version $v \in \{5, 6\}$ of the HMD for the country k . Let us denote by $I(a, t, k)$ the so-called correction indicator of mortality data, defined as follows:

$$I(a, t, k) := \frac{\mu^{(5)}(a, t, k)}{\mu^{(6)}(a, t, k)}. \quad (4)$$

As we have: $D^{(5)}(a, t, k) = D^{(6)}(a, t, k)$ for a fixed country k , the following equality is

satisfied:

$$I = \frac{\mu^{(5)}(a, t, k)}{\mu^{(6)}(a, t, k)} = \frac{E^{(6)}(a, t, k)}{E^{(5)}(a, t, k)}. \quad (5)$$

The HMD Methods Protocol corresponding to the HMD version 6 (Wilmoth et al., 2019) uses birth-by-month data when it is available in order to more accurately estimate population exposures. Therefore, it is assumed in the remainder of the paper that if $I \neq 1$ for a fixed 3-tuple (a, t, k) , the quantity $\mu^{(5)}(a, t, k)$ corresponds to a raw mortality rate which has not been corrected by using birth-by-month data, whereas $\mu^{(6)}(a, t, k)$ corresponds to the corrected mortality rate.

Boumezoued (2016) suggested that the ratio between corrected and crude estimates in the period table is the same for each age within a given cohort; this assumption has been relaxed in Boumezoued et al. (2018) where the correction approach has been reworked and made age-dependent. Here the value of the variable I is analysed for ages 18-70, cohorts 1915, 1919, 1940 and 1941 and following countries: Spain, Italy, France, Austria, Denmark, Sweden, Finland, and Norway. These results are depicted in Figures 4 to 7. At first sight, the figures illustrate that the variable I is roughly stable over ages within a given cohort. A global stability is observed for most countries, although this general remark does not hold for cohorts 1940 and 1941, which shows some variations of I for Spain and Italy. The analysis of I in terms of formulas and age variability is discussed in Appendix C. Overall, one can remark the significant discrepancies between the raw and corrected mortality rates.

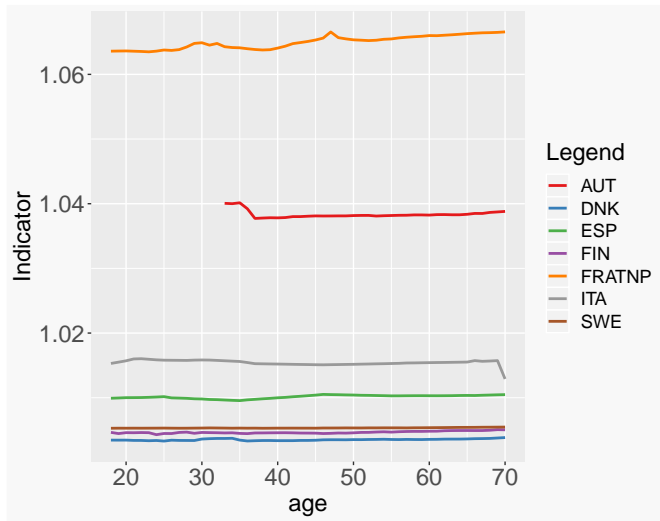


Figure 4: Values of I for cohort 1915 (Males)

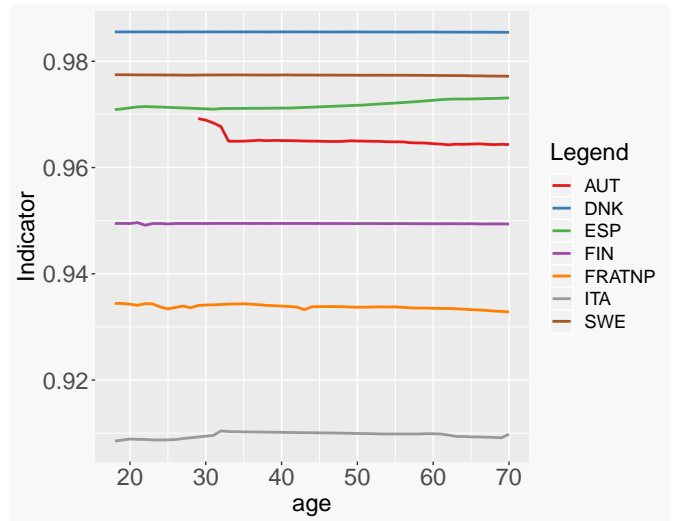


Figure 5: Values of I for cohort 1919 (Males)

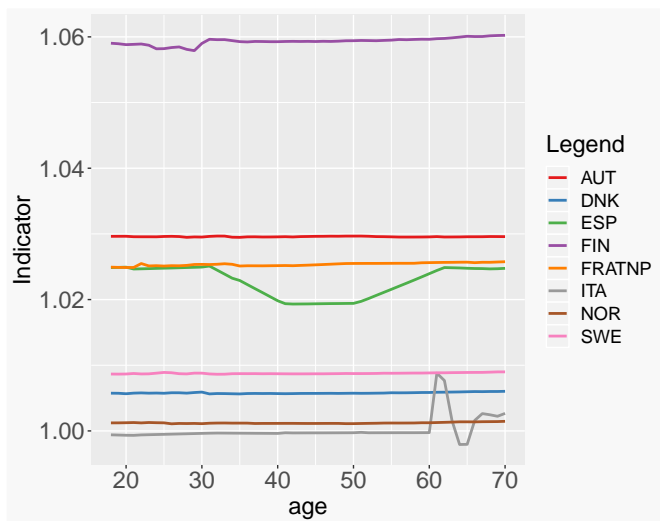


Figure 6: Values of I for cohort 1940 (Males)

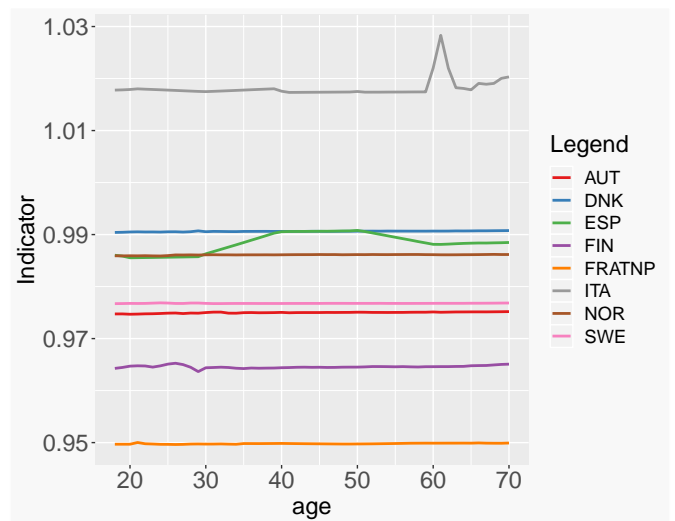


Figure 7: Values of I for cohort 1941 (Males)

3 The correction method to improve mortality exposure estimates

3.1 Neural networks and Representation Learning

Machine learning is programming computers to optimise a performance criterion using example data or past experience. A model is defined up to some parameters, and learning is the execution of a computer program to optimise the parameters of the model using the training data or past experience. The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data, or both.

Neural networks refer to a system whose design is originally schematically inspired by the functioning of neurons in the human brain (McCulloch and Pitts, 1943; Wiener, 1948). Its architecture includes neurons, synaptic connections that connect neurons to each other, and learning algorithms. A neural network is composed of three types of layers, respectively called input layers, hidden layers and output layers, each of which includes one or more neurons. Each unit of a network obtains “weighted” information through the synaptic links of the other connected units and returns an output using an activation function transforming the weighted sum of the input signals.

Neural networks aim to approximate non-linear functions defined on finite-dimensional space, relying on composition of layers of simple functions. The relevance of neural networks comes from the universal approximation theorem and the Kolmogorov-Arnold representation theorem (Arnold, 2009; Kolmogorov, 1956; Cybenko, 1989; Hornik, 1991).

In this paper, a feedforward neural network is considered, also called multi-layered perceptron (Alpaydin, 2014). Note that this choice is driven by the complexity/non-linearity of the prediction problem. Other machine learning techniques could have been considered; however, we prefer to focus on this approach in this work due to its flexibility and recent popularity in the field of mortality prediction. Moreover, we benchmark this approach to a standard linear regression, which will benefit from some feature engineering, as described in the next sections.

3.2 Splitting the dataset into train set and test set

Our dataset \mathcal{D} consists of N_{obs} observations:

$$\mathcal{D} = \{(\mathbf{x}_i, I_i) \mid I_i \neq 1 \text{ and } i \in \llbracket 1, N_{\text{obs}} \rrbracket\}, \quad (6)$$

where \mathbf{x} denotes the input variables vector of dimension D and the correction indicator of mortality I is the scalar output or target. Note that the number of observations can be large (correction indicator for several countries, ages and years).

The column vector inputs for all N_{obs} observations are aggregated in the $D \times N_{\text{obs}}$ design matrix X , and the targets are collected in the vector \mathbf{I} , so it can be written $\mathcal{D} = (X, \mathbf{I})$. In the regression setting, the targets are real values. We are interested in making inferences about the relationship between inputs and targets, *i.e.* the conditional distribution of the targets given the inputs (but we are not interested in modelling the input distribution itself).

In a realistic correction scenario where we would like to correct the mortality data of a country k for which fertility data is unavailable, we would use a set of countries for which such fertility data is available as the train set. We would then train our model on this set and compute the predictions of the correction indicator for the country k .

Following this observation, instead of separating the dataset \mathcal{D} randomly between a train set $\mathcal{D}_{\text{train}}$ and a test set $\mathcal{D}_{\text{test}}$ in order to determine the performance of our model, we thus decide to use each country as a test set. More precisely, let us denote by K the number of considered countries in the correction method. The dataset \mathcal{D} is divided in K subsamples, each subsample corresponding to the available data of a fixed country. One of the subsamples is then selected as the test set $\mathcal{D}_{\text{test}}$ and the union of the $K - 1$ other subsamples is the train set $\mathcal{D}_{\text{train}}$. The process is then repeated by selecting another subsample among the $K - 1$ subsamples which have not been used. Thus, the process is repeated K times such that each country is used exactly once as a test set $\mathcal{D}_{\text{test}}$.

The case where the information on years, ages and cohorts available is similar for any country is considered, leading to restrict information to the intersection of all individual countries data. The common available data for all countries is therefore made of a collection of death counts and population estimates for years between t_{\min} and t_{\max} , ages in the range a_{\min} to a_{\max} and cohorts between c_{\min} and c_{\max} .

Finally, our model is trained and tested K times on the K train sets and test sets. The average of the K mean squared errors (MSEs) of the target variable I (correction indicator) is then calculated to estimate the prediction error of the model. The MSE is a measure frequently used for assessing the accuracy of prediction obtained by a model. It measures the differences or residuals between actual and predicted values, respectively \hat{I}_i and I_i . The formula for computing MSE is as follows:

$$\text{MSE} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} (\hat{I}_i - I_i)^2, \quad (7)$$

where N_{test} is the number of data points in the considered test set, given by:

$$N_{\text{test}} = (t_{\max} - t_{\min} + 1)(a_{\max} - a_{\min} + 1) - \left(\mathbb{1}_{\{c_{\min} \geq t_{\min} - a_{\max} + 1\}} \sum_{j=1}^{c_{\min} - t_{\min} + a_{\max}} j + \mathbb{1}_{\{c_{\max} \leq t_{\max} - a_{\min} - 1\}} \sum_{j=1}^{t_{\max} - a_{\min} - c_{\max}} j \right).$$

3.3 Input variables and feature engineering

In machine learning, feature engineering is an informal although fundamental process which consists in creating new variables defined from existing variables in the dataset, in order to improve the predictive capacity of a considered model.

Let us again denote by k the country variable and c the cohort variable, where $c = t - a$. In the following, for simplicity of notation, it is assumed that k is fixed. We consider the following discrete second-order derivative of the mortality rate with respect to year, denoted by:

$$\Phi(a, t) = \frac{\mu(a, t + 2)}{\mu(a, t + 1)} - \frac{\mu(a, t + 1)}{\mu(a, t)}, \quad (8)$$

$\Phi(a, t)$ is thus the discrete first-order derivative of the mortality improvement rates $r(a, t)$ with respect to the year, defined by Equation (2). Similarly, we also consider the discrete second-order derivative of the mortality rate with respect to age, given by:

$$\Psi(a, t) = \frac{\mu(a + 2, t)}{\mu(a + 1, t)} - \frac{\mu(a + 1, t)}{\mu(a, t)}. \quad (9)$$

As the *Convexity Adjustment Ratio* of Cairns et al. (2016), the quantities $\Phi(a, t)$ and $\Psi(a, t)$ can be interpreted as convexity matrices of the mortality rate $\mu(a, t)$. These variables are of interest to detect potential anomalies in the raw mortality tables; however, the set of all convexity matrices is too large to ensure good predictions of the correction indicator I , especially due to sampling risk. In this context, the idea is to compute an average of these variables for each fixed cohort. In the following, an operator \mathcal{F} is defined to formalise this approach.

The following operator \mathcal{F} is defined by:

$$\mathcal{F} : V(a, t) \mapsto \mathcal{F}_V(c) = \frac{1}{N_c} \sum_{t-a=c} V(a, t), \quad \text{with } N_c = |\{V(a, t) \mid t - a = c\}|, \quad (10)$$

where V designates a variable depending on age a and year t , and $|\mathcal{E}|$ denotes the number of elements of a given set \mathcal{E} . We have for instance that $\mathcal{F}_\Psi(c) = \frac{1}{N_c} \sum_{t-a=c} \Psi(a, t)$ with $N_c = |\{\Psi(a, t) \mid t - a = c\}|$. In this case, N_c is the number of observed quantities $\Psi(a, t)$ which we have from the data for the fixed cohort c . As shown in Figure 8, $\mathcal{F}_\Psi(c)$ corresponds thus to the arithmetic mean of the discrete second-order derivatives of the quantity $\mu(a, t)$ with respect to age a for the fixed cohort $c = t - a$. As mortality rates data is available for several countries, the previous new created variables are also dependent on country k .

Final chosen input variables. Several input variables were tested in the baseline and neural network models, which include first and second-order derivatives of the annual birth counts $B(\cdot)$ and the population estimates $P(\cdot)$, the age a , as well as $\mathcal{F}_{\Phi_k}(k, c)$, and $\mathcal{F}_{\Psi_k}(k, c)$, where additional dependence to country k is highlighted in the notation. Finally, the variables $\mathcal{F}_{\Phi_k}(k, c)$ and $\mathcal{F}_{\Psi_k}(k, c)$ are kept for the baseline linear regression model and the age variable a is added for the neural network model. Therefore, the input vector is $\mathbf{x} = (x_1, x_2, x_3) = (\mathcal{F}_{\Phi_k}(k, c), \mathcal{F}_{\Psi_k}(k, c), a)$.

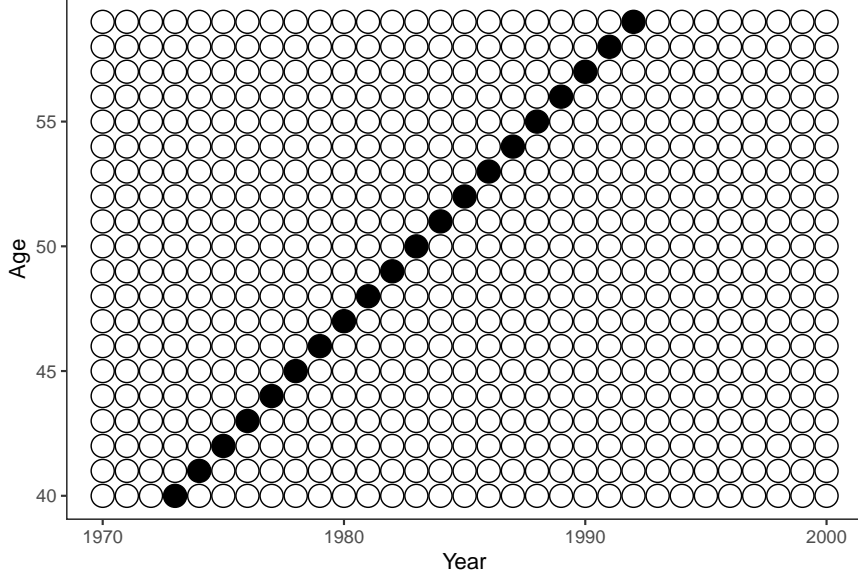


Figure 8: In this toy example, data points represent the values taken by the variable V for ages 40-59 and years 1970-2000. Black points represent the values corresponding to the 1933 cohort that are averaged to compute $\mathcal{F}_V(1933)$. In this case, $N_{1933} = 20$, which is the total number of black points.

3.4 The neural network model

As the input variables do not have the same orders of magnitude, the *Mean and Standard deviation transformation* is used to normalise them. Let ξ be a continuous input variable. The transformation for the variable ξ of the example i , denoted by ξ_i , is written as :

$$\begin{cases} \mathcal{T} : \xi_i \mapsto \check{\xi}_i = \frac{\xi_i - \bar{\xi}}{\sigma(\xi)} \\ \mathcal{T}^{-1} : \check{\xi}_i \mapsto \xi_i = \sigma(\xi) \cdot \check{\xi}_i + \bar{\xi} \end{cases} \quad (11)$$

with

$$\begin{cases} \bar{\xi} = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} \xi_i \\ \sigma(\xi) = \sqrt{\frac{1}{N_{\text{train}} - 1} \sum_{i=1}^{N_{\text{train}}} (\xi_i - \bar{\xi})^2} \end{cases} \quad (12)$$

where N_{train} is the number of data points in the train set.

Initial values of weights and biases of the neural network need to be defined in order to avoid problems relative to learning speed and generalisation capacity. In the following, the uniform Glorot initialisation ([Glorot and Bengio, 2010](#)) is used.

The architecture of a feedforward neural network with 3 hidden layers is chosen, including respectively 64, 32 and 16 neurons. A ReLU activation function is chosen for the 3 hidden layers, and a linear function as the output activation function. In order to avoid overfitting and get a better capacity of generalisation of the neural network (Hinton et al., 2012), a dropout layer is added between each hidden layer. It randomly omits each neuron with a probability p , which is a hyperparameter set to the value of 15%. Note that avoiding overfitting is of particular importance for mortality rates prediction (or here, ratios of raw and corrected mortality rates) since the observations are subject to sampling risk: due to the random realization of deaths, the death rate is not observable and we only deal with mortality rates estimators.

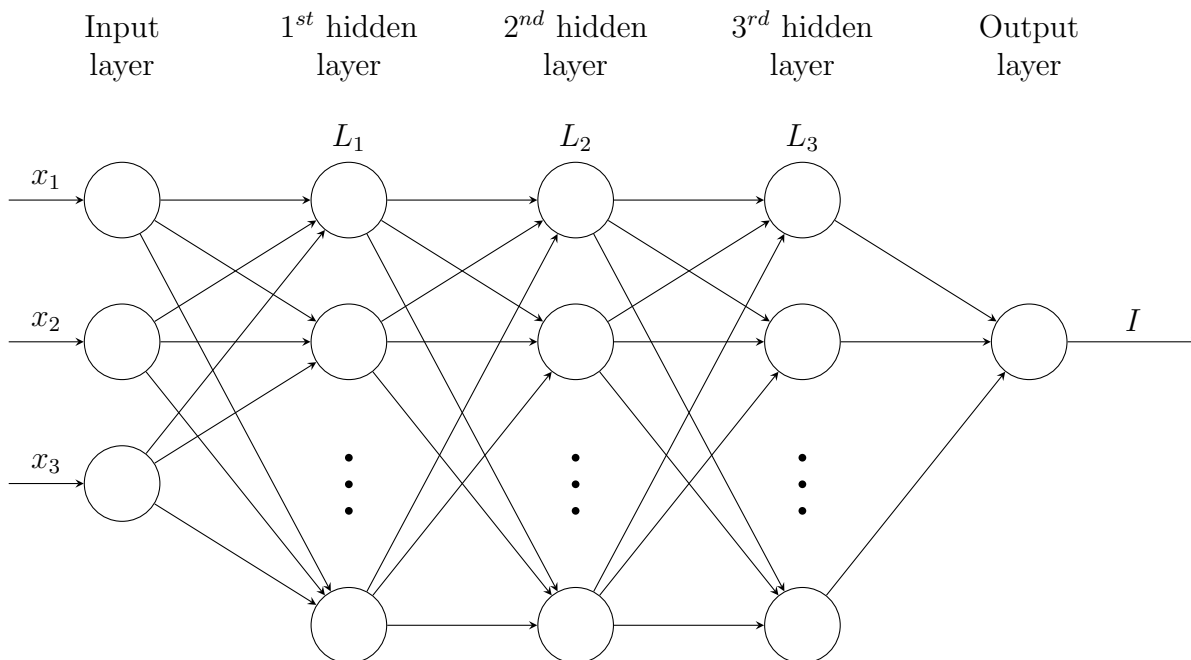


Figure 9: Architecture of the neural network. For clarity, only some neurons of the hidden layers have been shown; dropout layers are not shown.

A number of 30 epochs is chosen and it is assumed that the best model is the model which minimises the validation error. A random sample of 10% of the train set is used as a validation set. In other words, the neural network is calibrated on 90% of the training data, and its performance is assessed on a validation set, which corresponds to 10% of the train set.

In order to calibrate the model, the R package `Keras` is used (Chollet and Allaire, 2018), with `Adam` (Kingma and Ba, 2014) as the optimiser, and the mean squared error (MSE) as the cost function to minimise.

3.5 The baseline linear regression model

In order to benchmark the performance of the neural network approach, we rely on a simple linear regression which benefits from the feature engineering previously detailed, defined for each country k and cohort c :

$$\hat{I}(k, c) = \alpha_1 \mathcal{F}_{\Phi_k}(k, c) + \alpha_2 \mathcal{F}_{\Psi_k}(k, c) + \alpha_3, \quad (13)$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ is the parameter vector.

Note that the linear regression provides a correction which only depends on the cohort variable c and country k , unlike the neural network approach for which the age a is an additional input variable.

4 Results

Our dataset includes the following European countries for which corrected tables are available within the last HMD update: Spain, Italy, France, Austria, Denmark, Sweden, Finland and Norway. The data considered is made of the time period 1935-2016, cohorts 1915-1950, and ages 18-70.

It is worth pointing out that the predictions of a trained neural network are differing for each iteration. This is a consequence of (1) the optimisation algorithm which initialises the weights of the neural network to random values, (2) the random selection of batches during training to compute the gradients used in the back-propagation algorithm, and (3) dropout regularisation which is applied at random to some neurons of the neural network. As such, the test set MSEs of a neural network vary between training runs. Thus, in what follows, each neural network is fitted 10 times and the average MSE is considered to indicate the performance for each country test set.

The results are shown in Table 1, where it can be seen that the neural network performs best in terms of average MSE on 6 out of 8 country test sets. However, the results shown in Table 1 are the averages of 10 runs of a random training process. As discussed before, training a neural network once instead of 10 times may produce different results, as shown by the boxplot in Figure 10.

In order to compare the global performance of the neural network and the linear regression, the average of the Mean-Square Errors (MSE) corresponding to each test set is computed. We get an average prediction MSE error of $9,95.10^{-5}$ for the linear regression and $7,69.10^{-5}$ for the neural network. More specifically, the neural network method improved the prediction MSE error by 22.7% on average, compared to the linear regression method.

A graphical comparison of the predictions by the baseline model and the neural network model is available in Appendix A.

| Country test set | | Baseline model | Neural network | |
|------------------|----------|----------------------|----------------------|----------------------|
| Country | HMD code | MSE | Average MSE | Median MSE |
| Spain | ESP | $6,34 \cdot 10^{-5}$ | $7,45 \cdot 10^{-5}$ | $7,45 \cdot 10^{-5}$ |
| Italy | ITA | $7,42 \cdot 10^{-5}$ | $6,55 \cdot 10^{-5}$ | $6,59 \cdot 10^{-5}$ |
| France | FRATNP | $1,44 \cdot 10^{-4}$ | $8,56 \cdot 10^{-5}$ | $8,54 \cdot 10^{-5}$ |
| Austria | AUT | $1,11 \cdot 10^{-4}$ | $1,21 \cdot 10^{-4}$ | $1,20 \cdot 10^{-4}$ |
| Sweden | SWE | $6,14 \cdot 10^{-5}$ | $4,52 \cdot 10^{-5}$ | $4,52 \cdot 10^{-5}$ |
| Finland | FIN | $1,15 \cdot 10^{-4}$ | $8,48 \cdot 10^{-5}$ | $8,41 \cdot 10^{-5}$ |
| Denmark | DNK | $1,28 \cdot 10^{-4}$ | $7,66 \cdot 10^{-5}$ | $7,56 \cdot 10^{-5}$ |
| Norway | NOR | $9,81 \cdot 10^{-5}$ | $6,19 \cdot 10^{-5}$ | $6,09 \cdot 10^{-5}$ |

Table 1: Country test set MSEs of the baseline linear regression and the neural network

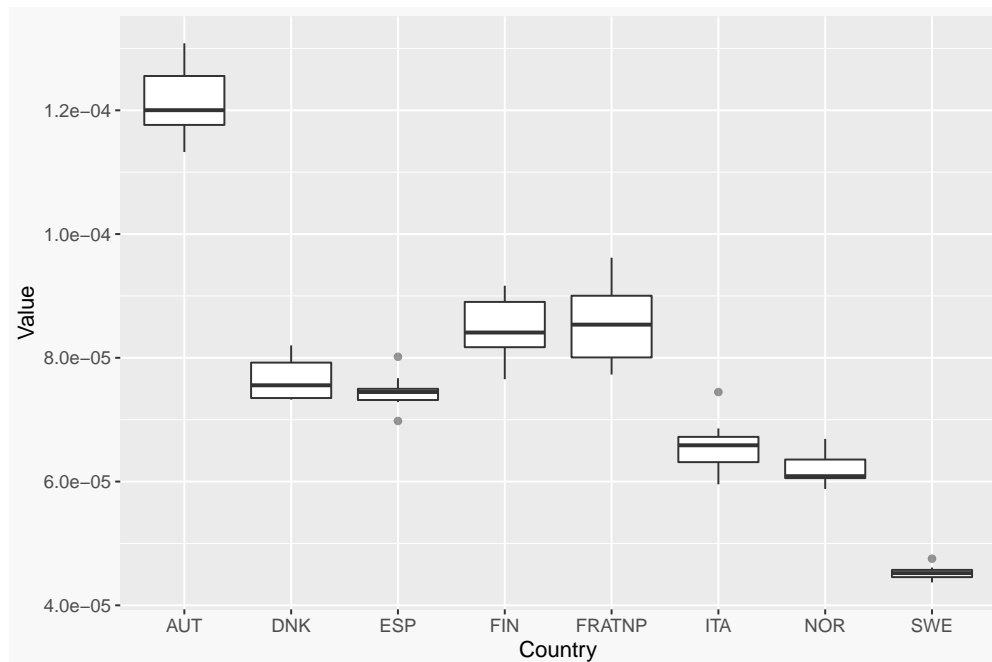


Figure 10: Boxplot of the MSE results on the country test sets. The neural network is trained for 10 runs

5 An application of the correction method

5.1 The corrected matrix of mortality improvement rates

The following will present an example of an application of our correction method to countries for which birth-by-month data is unavailable. West Germany, Ukraine, Belarus and Poland data for Males is used from the HMD version 5 for the time period 1970-2000 inclusive and ages 40-59 inclusive. Our correction method is then applied to predict the correction indicators I_{LR} and I_{NN} , respectively using the baseline linear regression and the neural network approach (see Figures 11 to 14). The mortality rates after correction μ_{LR} and μ_{NN} are then deduced, following Equation (4). Thus, the mortality improvement rates matrix $r(a, t)$ (see e.g. Figure 15 for West Germany) can be compared with the corrected ones denoted $r_{LR}(a, t)$ and $r_{NN}(a, t)$ for the linear regression and the neural network respectively (see Figures 16 and 17), defined by:

$$r_{LR}(a, t) = \frac{\mu_{LR}(a, t+1) - \mu_{LR}(a, t)}{\mu_{LR}(a, t)} \text{ and } r_{NN}(a, t) = \frac{\mu_{NN}(a, t+1) - \mu_{NN}(a, t)}{\mu_{NN}(a, t)}. \quad (14)$$

The correction method seems to be very efficient concerning the isolated false cohort effects, as the anomalous diagonals around 1915 and 1920, and well as those around 1940 disappear. It turns out that the historical volatility of mortality improvement rates is greatly reduced when using retreated mortality data given by the correction method, see also Boumezoued (2016). Moreover, from the matrix of mortality improvement rates, one can see that the period and age features are preserved, even cleaned in the corrected version. This qualitative feature of the correction process is appealing, since it allows to recover the main characteristics of mortality without distortion in areas where anomalous diagonals are not present.

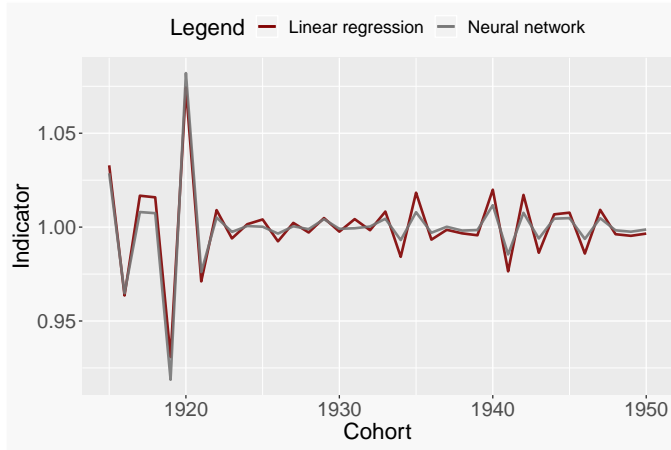


Figure 11: Predicted correction indicator
— West Germany

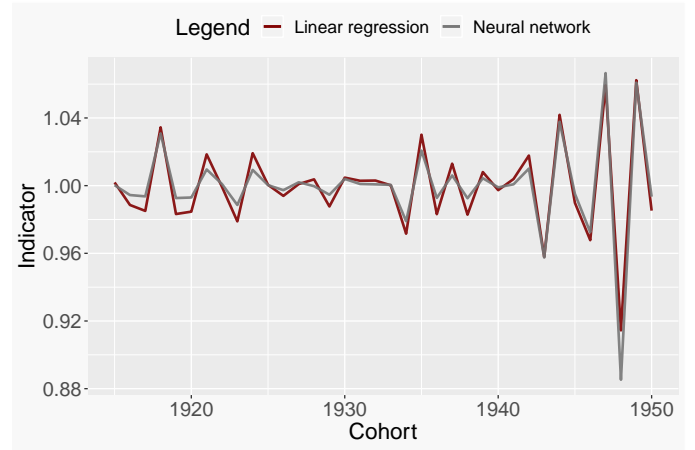


Figure 12: Predicted correction indicator
— Ukraine

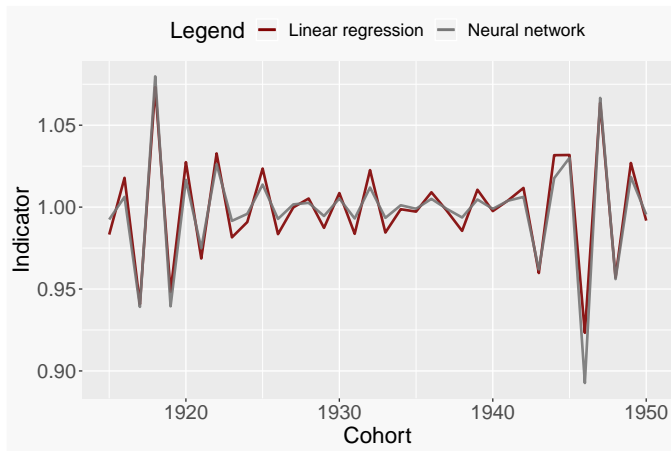


Figure 13: Predicted correction indicator
— Belarus

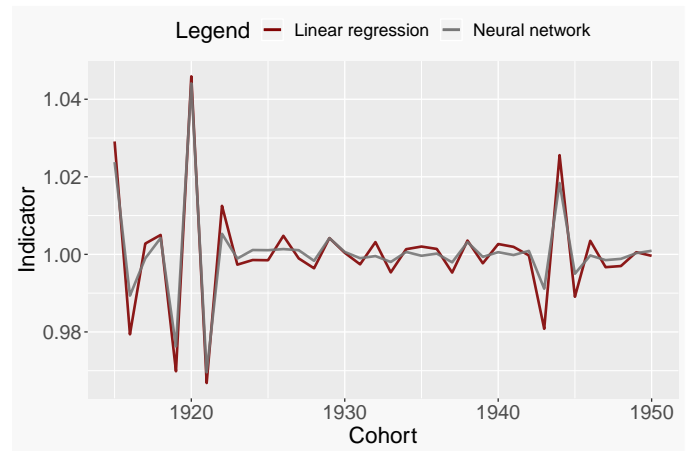


Figure 14: Predicted correction indicator
— Poland

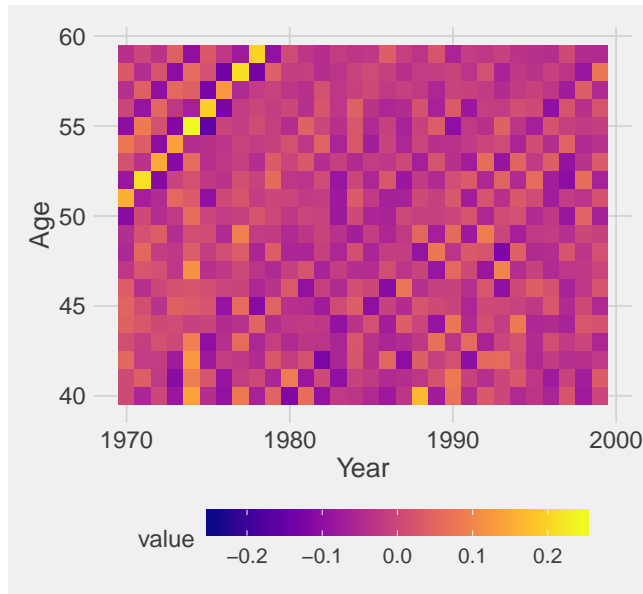


Figure 15: Mortality improvement rates — West Germany raw data

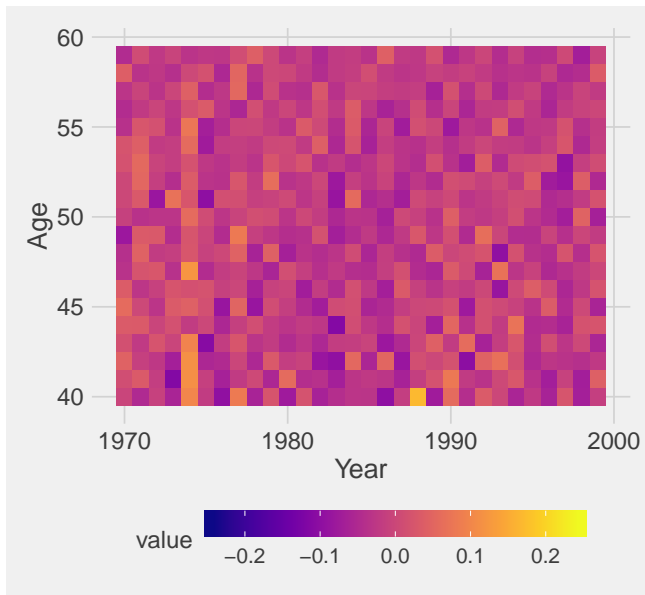


Figure 16: Mortality improvement rates — West Germany corrected data by the linear regression method

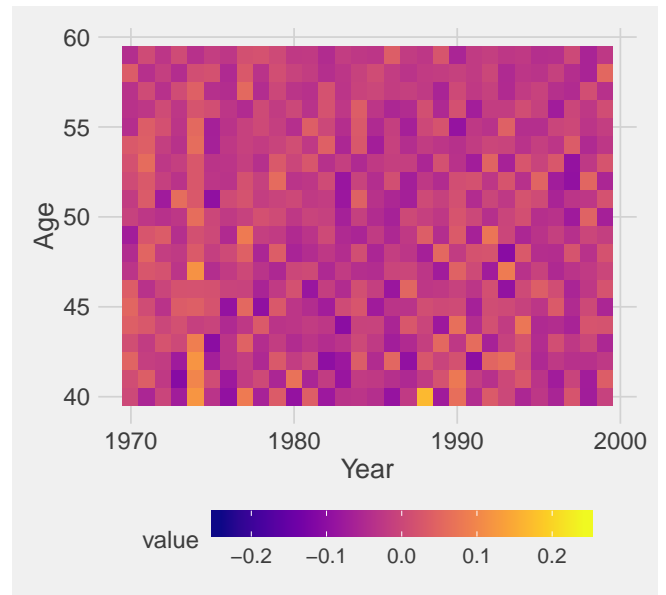


Figure 17: Mortality improvement rates — West Germany corrected data by the neural network method

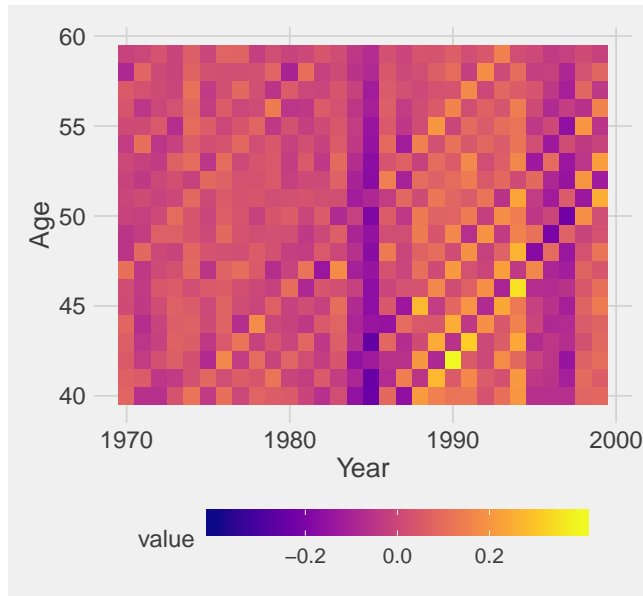


Figure 18: Mortality improvement rates — Ukraine raw data

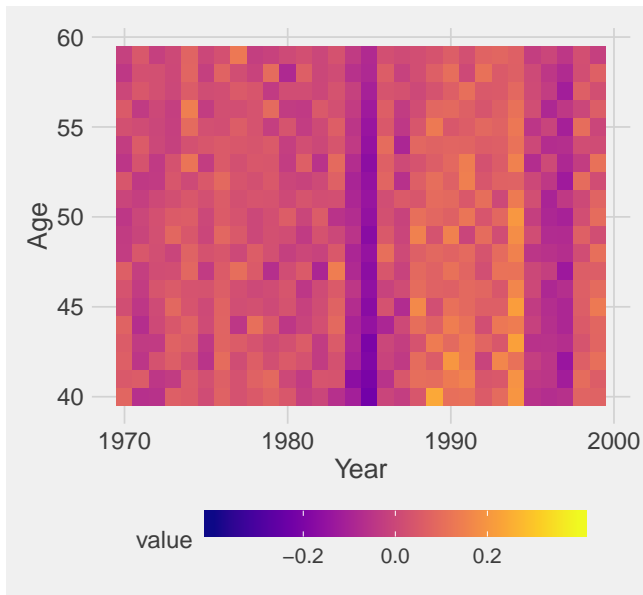


Figure 19: Mortality improvement rates — Ukraine corrected data by the linear regression method

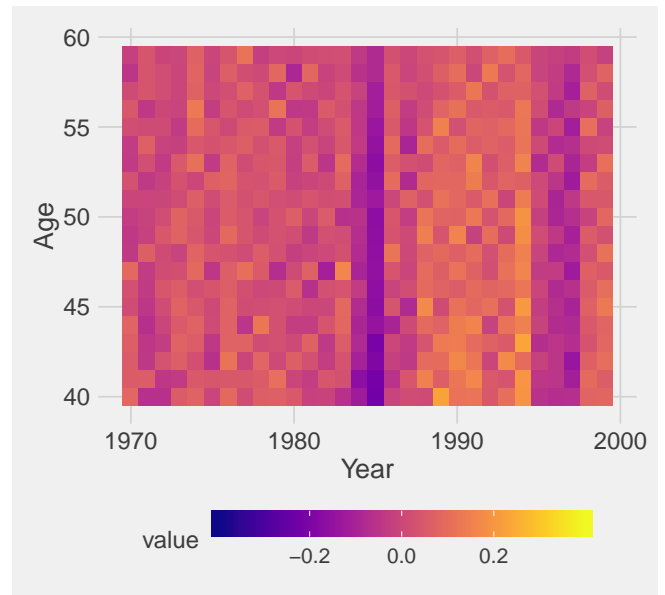


Figure 20: Mortality improvement rates — Ukraine corrected data by the neural network method

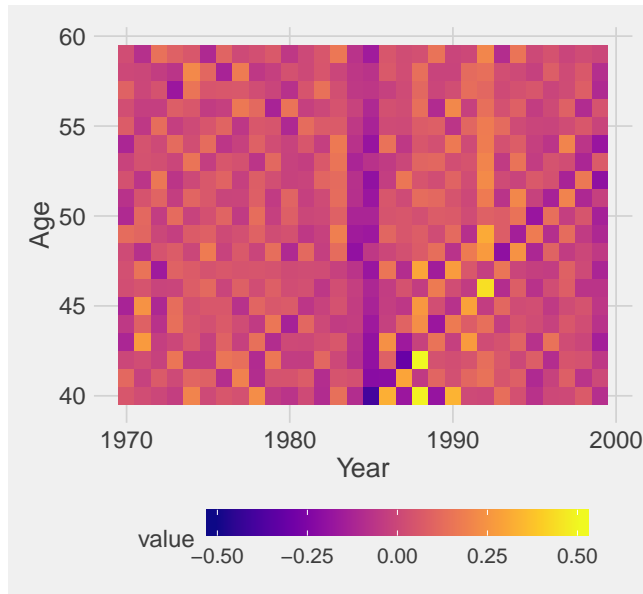


Figure 21: Mortality improvement rates — Belarus raw data

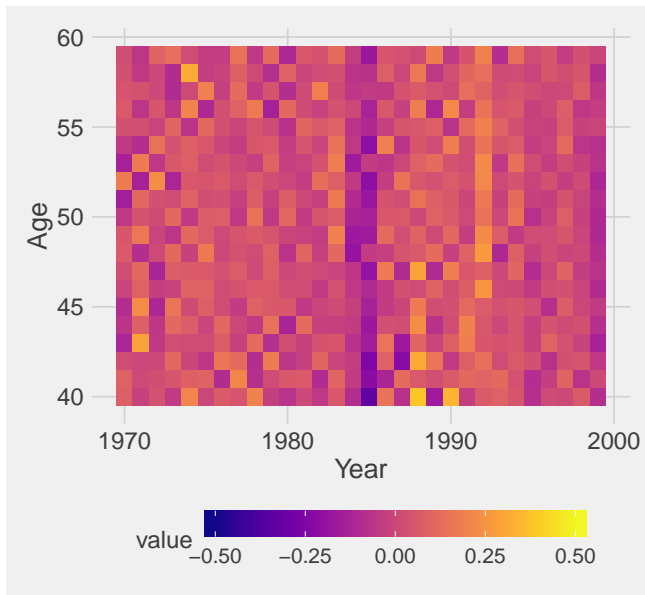


Figure 22: Mortality improvement rates — Belarus corrected data by the linear regression method

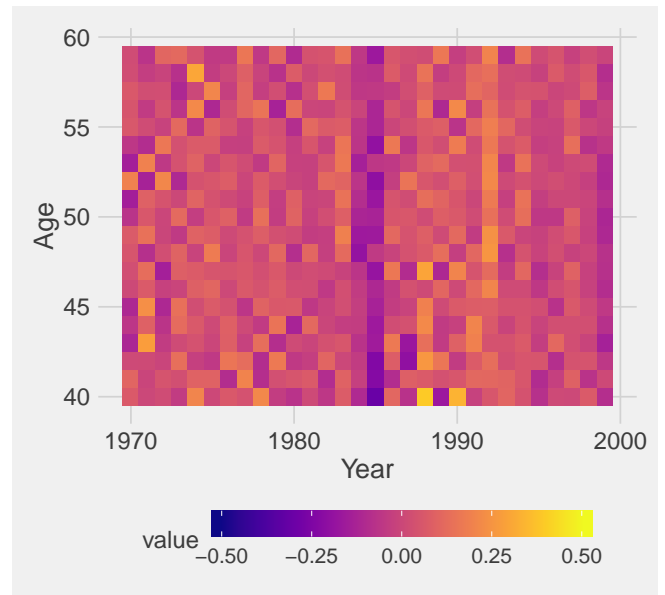


Figure 23: Mortality improvement rates — Belarus corrected data by the neural network method

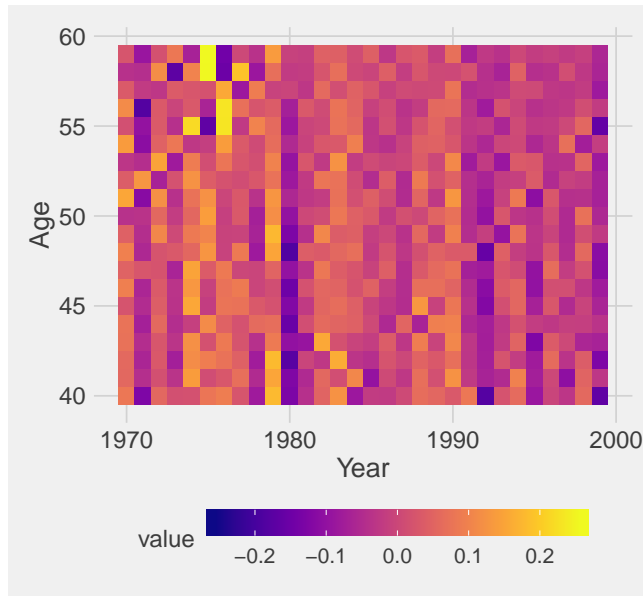


Figure 24: Mortality improvement rates — Poland raw data

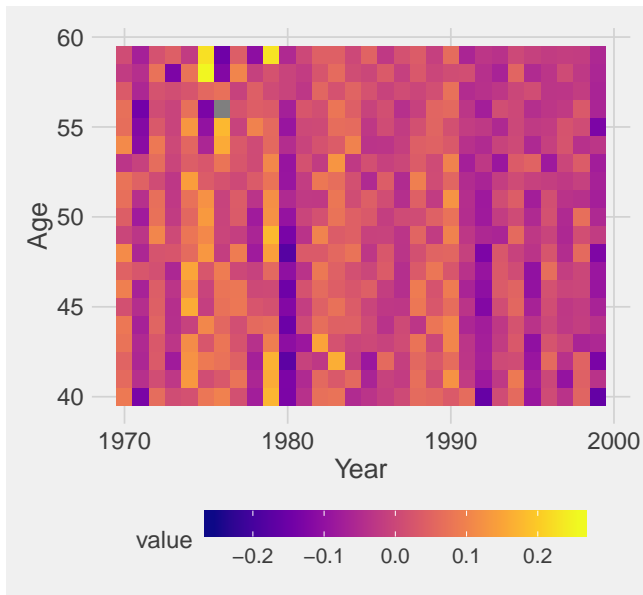


Figure 25: Mortality improvement rates — Poland corrected data by the linear regression method

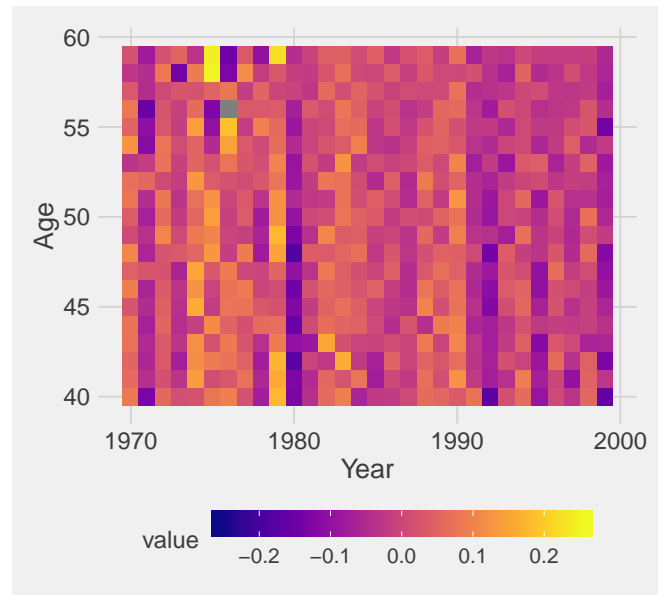


Figure 26: Mortality improvement rates — Poland corrected data by the neural network method

5.2 Impact of the correction method on stochastic mortality models

The correction of the mortality data in general, and the removal of false cohort effects in particular, has already been discussed in the context of stochastic mortality modelling by Cairns et al. (2016), Boumezoued (2016) and Boumezoued et al. (2019). As in Boumezoued et al. (2019) relying on monthly birth counts, the aim of this subsection is to show how previous findings are preserved with the proposed correction inference method without birth-by-month data. Focus is first on the classical Lee-Carter model (Lee and Carter, 1992), also referred to as M1 model. This model decomposes mortality as a static age-structure $\beta_a^{(1)}$, a general level driven by a stochastic process in time $\kappa_t^{(2)}$ and an age-specific sensitivity $\beta_a^{(2)}$ to this general level as follows:

$$\log \mu(a, t) = \beta_a^{(1)} + \beta_a^{(2)} \cdot \kappa_t^{(2)}. \quad (15)$$

Renshaw and Haberman (2006) proposed to include a cohort effect as follows:

$$\log \mu(a, t) = \beta_a^{(1)} + \beta_a^{(2)} \cdot \kappa_t^{(2)} + \beta_a^{(3)} \cdot \gamma_{t-a}^{(3)}, \quad (16)$$

with γ_c the additional cohort-related factor which allows to adjust mortality rates for a cohort which originates from year $c = t - a$. This model, known as M2 model, is a generalisation of the M1 model. As the M2 model is known to be difficult to calibrate, it is chosen to consider the Age-Period-Cohort model, also known as M3 model (Currie, 2006), which is a special case of the M2 model, where the age, period and cohort components influence mortality independently in the following form:

$$\log \mu(a, t) = \beta_a^{(1)} + \kappa_t^{(2)} + \gamma_{t-a}^{(3)}. \quad (17)$$

As an alternative, Plat (2009) proposed a model that combines the CBD model from Cairns et al. (2006), known as M5, with some features of the Lee-Carter model to produce a model that is suitable for full age ranges and captures the cohort effect. Here, as only a specific age band is of interest, a simplified version of this model is considered, and is given by:

$$\log \mu(a, t) = \beta_a^{(1)} + \kappa_t^{(2)} + \kappa_t^{(3)}(x - \bar{x}) + \gamma_{t-a}^{(4)}, \quad (18)$$

where \bar{x} denotes the mean age over the range of ages used in the calibration.

The experiment is carried out using the `StMoMo` R package (Villegas et al., 2015), and the standard log-Poisson calibration (Brouhns et al., 2002). The age range for model inference is 40-59 inclusive and the time period is 1970-2000 inclusive. Note that since M3 and simplified Plat models include cohort parameters and in agreement with the usual practice (see e.g. Cairns et al. (2009) and Haberman and Renshaw (2011)), all cohorts that have fewer than three observations are excluded.

When evaluating the goodness-of-fit of different models, it is generally anticipated that models with more parameters provide a better fit to the data. To rule out the possibility that the better fit observed in a model is the result of over-parametrisation and compare the

relative performance of several models, it has become common in the mortality literature to use information criteria which modify the maximum likelihood criterion by penalizing models with more parameters (Cairns et al., 2009; Villegas et al., 2015). Two of these criteria are the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC), with a lower value of AIC and BIC being preferable. They are defined as $AIC = 2\nu - 2 \log \mathcal{L}$ and $BIC = \nu \log \aleph - 2 \log \mathcal{L}$, where ν designates the effective number of parameters in the model, \mathcal{L} is the model likelihood, and \aleph is the number of observations in the data.

Tables 2 to 5 below show the quantities $-2 \log \mathcal{L}$, AIC and BIC for M1, M3 and simplified Plat models fitted on raw HMD data and retreated mortality data for West Germany, Belarus, Ukraine and Poland.

It is shown that the retreated mortality estimates allow the M1 model to better capture mortality dynamics embedded in the corrected tables. However, there is no change in statistical inference criteria for M3 and simplified Plat models. This seems to be explained by the fact that these models have a cohort component which captures the irregularities of the original mortality tables, leading to unchanged residuals between the raw and the retreated mortality tables. The log-likelihood, AIC and BIC stability for M3 and Plat models also show that our correction method does not disturb the mortality rates table, except for the cohort component.

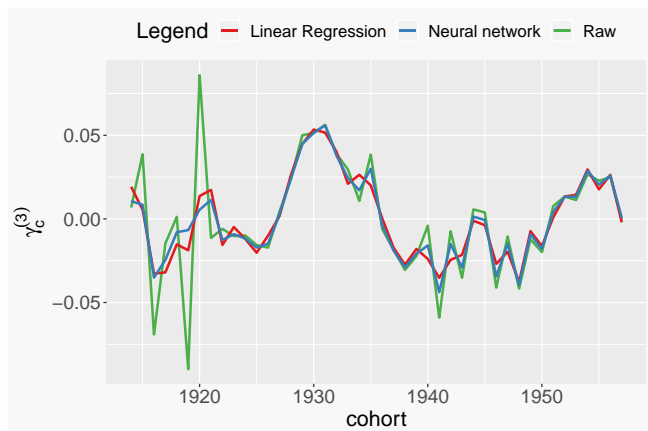


Figure 27: M3 model $\gamma_c^{(3)}$ parameter — West Germany raw and retreated mortality data

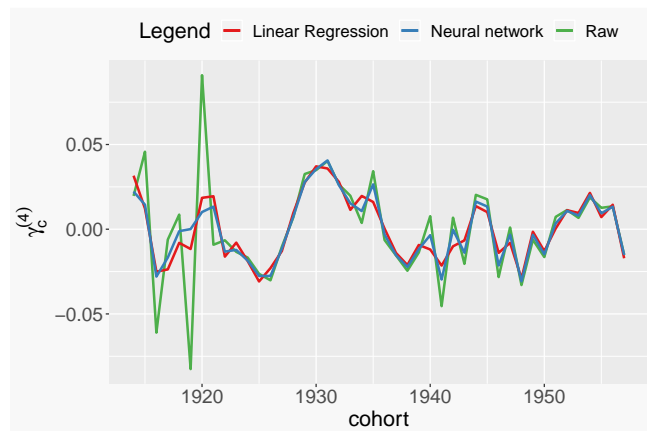


Figure 28: Simplified Plat model $\gamma_c^{(4)}$ parameter — West Germany raw and retreated mortality data

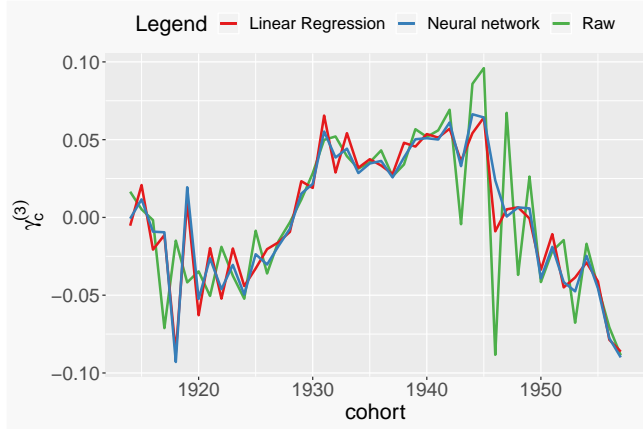


Figure 29: M3 model $\gamma_c^{(3)}$ parameter — Belarus raw and retreated mortality data

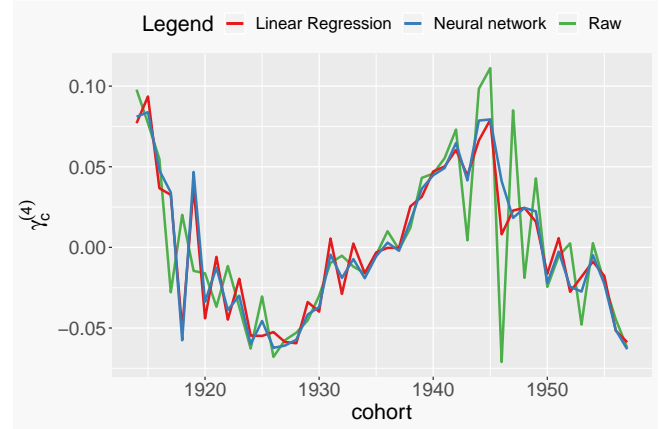


Figure 30: Simplified Plat model $\gamma_c^{(4)}$ parameter — Belarus raw and retreated mortality data

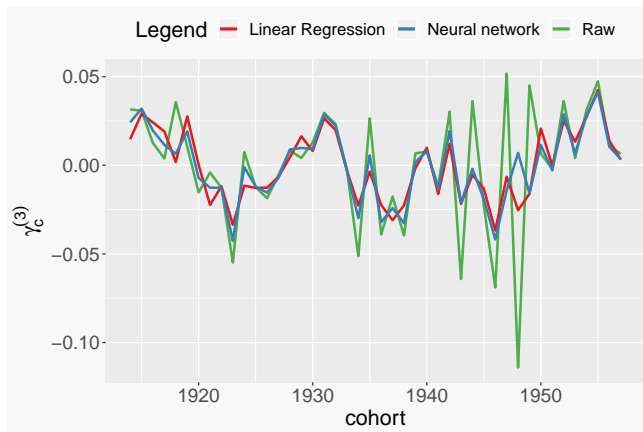


Figure 31: M3 model $\gamma_c^{(3)}$ parameter — Ukraine raw and retreated mortality data

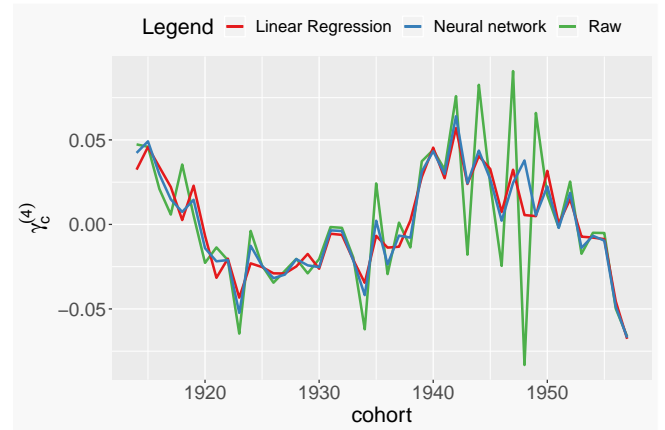


Figure 32: Simplified Plat model $\gamma_c^{(4)}$ parameter — Ukraine raw and retreated mortality data

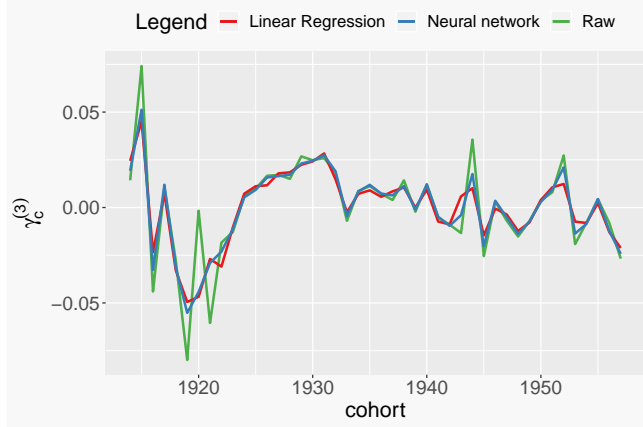


Figure 33: M3 model $\gamma_c^{(3)}$ parameter — Poland raw and retreated mortality data

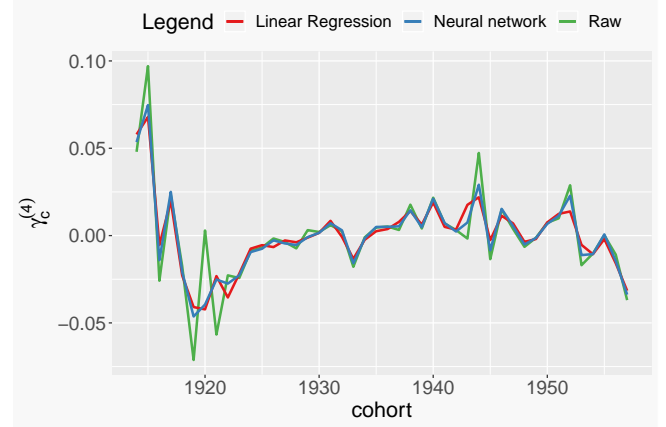


Figure 34: Simplified Plat model $\gamma_c^{(4)}$ parameter — Poland raw and retreated mortality data

More specifically, five conclusions appear in the light of the results obtained:

1. For the all three models, the log-likelihoods on retreated mortality data are close between the linear regression and the neural network.
2. The M1 model is significantly improved when calibrated with retreated mortality data. Thus, the correction method has the same effect as the latest correction method based on monthly fertility data, see the discussion in [Boumezoued et al. \(2018\)](#). Moreover, the log-likelihood is unchanged for stochastic mortality models with a cohort component (M3 and simplified Plat models), which is also in line with the findings in the cited paper.
3. The gamma parameters of M3 and simplified Plat models are much more regular when these models are fitted with retreated mortality data. It can be pointed out that the ranking between models in terms of goodness-of-fit can even be reversed when retreated mortality data is used in the calibration. In the example of Ukraine data, [Table 4](#) shows that the M1 model becomes better than the M3 model when fitted with retreated mortality data. Thus, from a statistical point of view, it can be more relevant to integrate in the model an age-specific sensitivity parameter than a cohort-related parameter.
4. The impact of calibration with retreated mortality data on future mortality projections has already been discussed by [Cairns et al. \(2016\)](#) and [Boumezoued et al. \(2018\)](#) and this is not the object of the paper. It is expected to get similar results with this new correction method which changes the dynamics of gamma parameters for mortality models with a cohort component. Let us recall that the volatility of forecasted with standard mortality models is now close to the historical volatility of corrected mortality data ([Boumezoued, 2016](#)). This is an important phenomenon as it shows that classical mortality models are able to replicate longevity and mortality risks.

5. To go further, note that even for a model which has a particularly high goodness-of-fit with raw data like the simplified Plat model, no change is observed in terms of log-likelihood when using retreated mortality data, whereas the dynamics of the cohort parameter gamma is greatly changed. On the contrary, the M1 model is unstable in terms of log-likelihood but its parameters $\beta_a^{(1)}$, $\beta_a^{(2)}$ and $\kappa_t^{(2)}$ are stable (see Appendix B). The M3 model and the simplified Plat model are stable in terms of log-likelihood but their gamma cohort-related parameters are unstable (see Figures 27 to 34).

| Mortality models | $-2 \log \mathcal{L}$ | | | BIC criterion | | | AIC criterion | | |
|------------------|-----------------------|------|------|---------------|------|------|---------------|------|------|
| | Raw | LR | NN | Raw | LR | NN | Raw | LR | NN |
| M1 | 8252 | 7556 | 7585 | 8695 | 8000 | 8029 | 8390 | 7694 | 7723 |
| M3 | 6515 | 6515 | 6514 | 7104 | 7104 | 7104 | 6699 | 6699 | 6698 |
| Simplified Plat | 6378 | 6378 | 6378 | 7153 | 7153 | 7153 | 6620 | 6620 | 6620 |

Table 2: Calibration on West Germany data

| Mortality models | $-2 \log \mathcal{L}$ | | | BIC criterion | | | AIC criterion | | |
|------------------|-----------------------|------|------|---------------|------|------|---------------|------|------|
| | Raw | LR | NN | Raw | LR | NN | Raw | LR | NN |
| M1 | 6324 | 6128 | 6113 | 6768 | 6571 | 6557 | 6462 | 6266 | 6251 |
| M3 | 5767 | 5767 | 5767 | 6357 | 6357 | 6357 | 5951 | 5951 | 5951 |
| Simplified Plat | 5583 | 5583 | 5583 | 6358 | 6358 | 6358 | 5825 | 5825 | 5825 |

Table 3: Calibration on Belarus data

| Mortality models | $-2 \log \mathcal{L}$ | | | BIC criterion | | | AIC criterion | | |
|------------------|-----------------------|------|------|---------------|------|------|---------------|------|------|
| | Raw | LR | NN | Raw | LR | NN | Raw | LR | NN |
| M1 | 10182 | 8532 | 8634 | 10625 | 8975 | 9078 | 10320 | 8670 | 8772 |
| M3 | 8547 | 8547 | 8548 | 9137 | 9137 | 9138 | 8731 | 8731 | 8732 |
| Simplified Plat | 6940 | 6940 | 6941 | 7716 | 7716 | 7717 | 7182 | 7182 | 7183 |

Table 4: Calibration on Ukraine data

| Mortality models | $-2 \log \mathcal{L}$ | | | BIC criterion | | | AIC criterion | | |
|------------------|-----------------------|------|------|---------------|------|------|---------------|------|------|
| | Raw | LR | NN | Raw | LR | NN | Raw | LR | NN |
| M1 | 7297 | 7115 | 7147 | 7741 | 7559 | 7591 | 7435 | 7253 | 7285 |
| M3 | 6752 | 6752 | 6752 | 7341 | 7341 | 7341 | 6936 | 6936 | 6936 |
| Simplified Plat | 6567 | 6567 | 6567 | 7343 | 7343 | 7343 | 6809 | 6809 | 6809 |

Table 5: Calibration on Poland data

6 Conclusion and avenues for future research

This paper has proposed to develop a correction method of mortality estimates in the absence of birth-by-month data, based on either a linear regression model or a neural network.

The method has been backtested on countries for which birth-by-month data is available and the performance of the method has been determined by computing average prediction MSE errors, the observed correction indicator being obtained by comparing the Human Mortality Database version 5 and version 6 updates. The baseline linear regression shows good results in terms of average prediction MSE error and the results are improved by the neural network approach which includes age as an additional covariate.

The method has then been applied to West Germany, Ukraine, Belarus and Poland, countries for which birth-by-month data is not available for most cohorts of interest. The correction method appeared to be very efficient concerning the isolated false cohort effects, as the anomalous diagonals around 1915 and 1920, and well as those around 1940 disappear. Moreover, the other period and age mortality characteristics are preserved, as it is the case for the correction methods using monthly birth counts.

The impact of the correction method has also been studied on three stochastic mortality models: M1, M3 and a simplified version of the Plat model. For the all three models, the log-likelihoods on retreated mortality data are close between the linear regression and the neural network. The M1 model is significantly improved when calibrated with retreated mortality data, whereas the log-likelihood is unchanged for stochastic mortality models with a cohort component. Thus, the correction method has the same effect as the latest correction method based on monthly fertility data.

The gamma parameters of the M3 and simplified Plat models are much more regular when these models are fitted with retreated mortality data. It can be pointed out that the ranking between models in terms of goodness-of-fit can even be reversed in some cases when retreated mortality data is used in the calibration. A case has been exhibited where, from a statistical point of view, it is more relevant to integrate in the stochastic mortality model an age-specific sensitivity parameter than a cohort-related parameter. This is explained by the fact that the previous false cohort effects artificially favour mortality models with cohort parameters.

Finally, one should note that the correction method may be improved by adding other countries with available birth-by-month data in the training set. Moreover, a more comprehensive search over the architecture of the neural network (more layers, different hyper-parameters tuning), may produce results that are more optimal than those presented here. Other improvements to the model are the inclusion of additional input variables and the consideration of Females data. These aspects can be integrated in the present framework without additional modelling complexity, and are as such left for further developments.

Acknowledgements

The authors are grateful to Magali Barbieri and her colleagues for fruitful discussions on the Human Mortality Database updates.

References

- Alpaydin, E. (2014). Multilayer perceptrons. In *Introduction to Machine Learning*, chapter 11. New York: Springer.
- Arnold, V. I. (2009). *Vladimir I. Arnold-Collected Works: Representations of Functions, Celestial Mechanics, and KAM Theory 1957-1965*, volume 1. Springer Science & Business Media.
- Boumezoued, A. (2016). Improving HMD mortality estimates with HFD fertility data. <https://hal.archives-ouvertes.fr/hal-01270565>.
- Boumezoued, A., Hoffmann, M., and Jeunesse, P. (2018). A new inference strategy for general population mortality tables. <https://hal.archives-ouvertes.fr/hal-01773665>.
- Boumezoued, A., Hoffmann, M., and Jeunesse, P. (2019). Nonparametric adaptive inference of birth and death models in a large population limit. *arXiv preprint arXiv:1903.00673*.
- Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and economics*, 31(3):373–393.
- Cairns, A. J., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718.
- Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, 13(1):1–35.
- Cairns, A. J., Blake, D., Dowd, K., and Kessler, A. R. (2016). Phantoms never die: living with unreliable population data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(4):975–1005.
- Chollet, F. and Allaire, J. J. (2018). R interface to Keras. RStudio, Google. <https://cloud.r-project.org/web/packages/keras/index.html>.
- Currie, I. D. (2006). Smoothing and forecasting mortality rates with p-splines. *Talk given at the Institute of Actuaries*.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314.
- Deprez, P., Shevchenko, P. V., and Wüthrich, M. V. (2017). Machine learning techniques for mortality modeling. *European Actuarial Journal*, 7(2):337–352.

- Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256.
- Haberman, S. and Renshaw, A. (2011). A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics*, 48(1):35–55.
- Hainaut, D. (2018). A neural-network analyzer for mortality forecast. *ASTIN Bulletin: The Journal of the IAA*, 48(2):481–508.
- HFD (2018). The Human Fertility Database. Max Planck institute for demographic research (Germany), and Vienna institute of demography (Austria). <https://www.humanfertility.org>.
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.
- HMD (2018). The Human Mortality Database. University of California, Berkeley (USA), and Max Planck institute for demographic research (Germany). <https://www.mortality.org>.
- Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257.
- Keyes, K. M., Utz, R. L., Robinson, W., and Li, G. (2010). What is a cohort effect? Comparison of three statistical methods for modeling cohort effects in obesity prevalence in the United States, 1971–2006. *Social science & medicine*, 70(7):1100–1108.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kolmogorov, A. (1956). The representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables. *Doklady Akademii Nauk SSSR*, 108(2):179–182.
- Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American statistical association*, 87(419):659–671.
- Levantesi, S. and Nigri, A. (2019). A random forest algorithm to improve the Lee–Carter mortality forecasting: impact on q-forward. *Soft Computing*, pages 1–15.
- Levantesi, S. and Pizzorusso, V. (2019). Application of machine learning to mortality modeling and forecasting. *Risks*, 7(1):26.
- Lexis, W. H. R. A. (1875). *Einleitung in die Theorie der Bevölkerungsstatistik*. KJ Trübner.
- Ludkovski, M., Risk, J., and Zail, H. (2018). Gaussian process models for mortality rates and improvement factors. *ASTIN Bulletin*, 48(3):1307–1347.

- McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.
- Plat, R. (2009). On stochastic mortality modeling. *Insurance: Mathematics and Economics*, 45(3):393–404.
- Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and economics*, 38(3):556–570.
- Richards, S. (2008). Detecting year-of-birth mortality patterns with limited data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(1):279–298.
- Richman, R. and Wüthrich, M. V. (2019). A neural network extension of the Lee-Carter model to multiple populations. *Annals of Actuarial Science*, pages 1–21. <https://doi.org/10.1017/S1748499519000071>.
- Villegas, A., Kaishev, V. K., and Millossovich, P. (2015). **StMoMo**: An R package for stochastic mortality modelling. In *7th Australasian Actuarial Education and Research Symposium*.
- Wiener, N. (1948). *Cybernetics or Control and Communication in the Animal and the Machine*. 2nd ed. Cambridge : Massachusetts Institute of Technology Press.
- Wilmoth, J. R., Andreev, K., Jdanov, D., and Gleij, D. A. (2007). Methods protocol for the Human Mortality Database. *University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock*. [Last Revised: May 31, 2007 (Version 5)]. <http://v5.mortality.org>.
- Wilmoth, J. R., Andreev, K., Jdanov, D., Gleij, D. A., and Riffe, T. (2019). Methods protocol for the Human Mortality Database. *University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock*. [Last Revised: October 5, 2019 (Version 6)]. <http://mortality.org>.

Appendix A Predicted correction indicators for each country test set

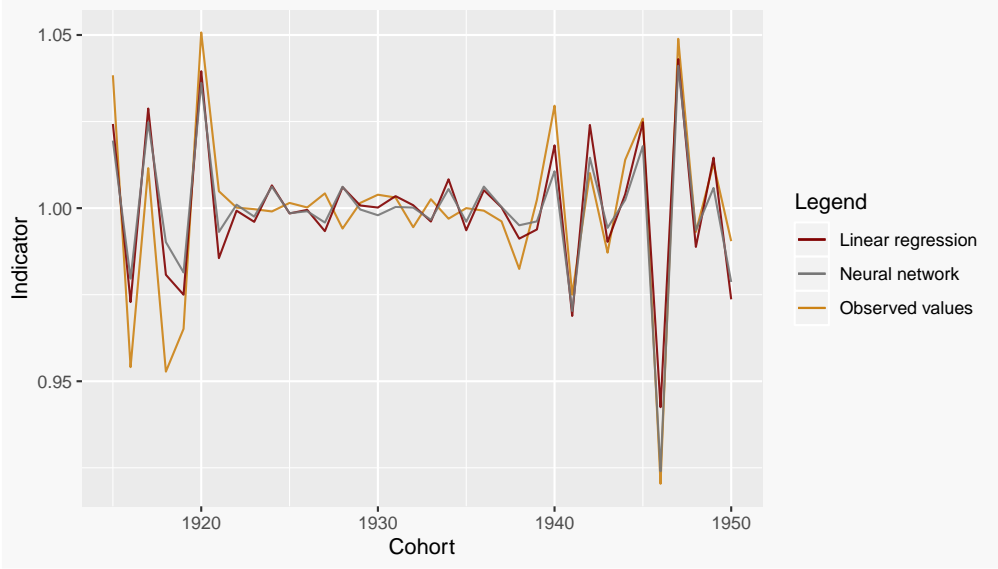


Figure 35: Austria: Predicted correction indicator (averaged by age within each cohort)



Figure 36: France: Predicted correction indicator (averaged by age within each cohort)

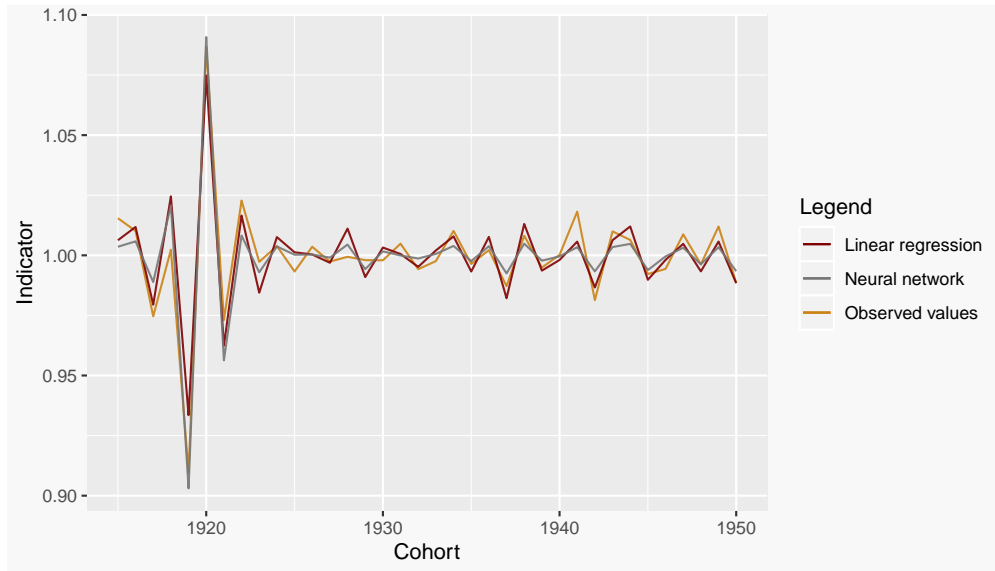


Figure 37: Italy: Predicted correction indicator (averaged by age within each cohort)

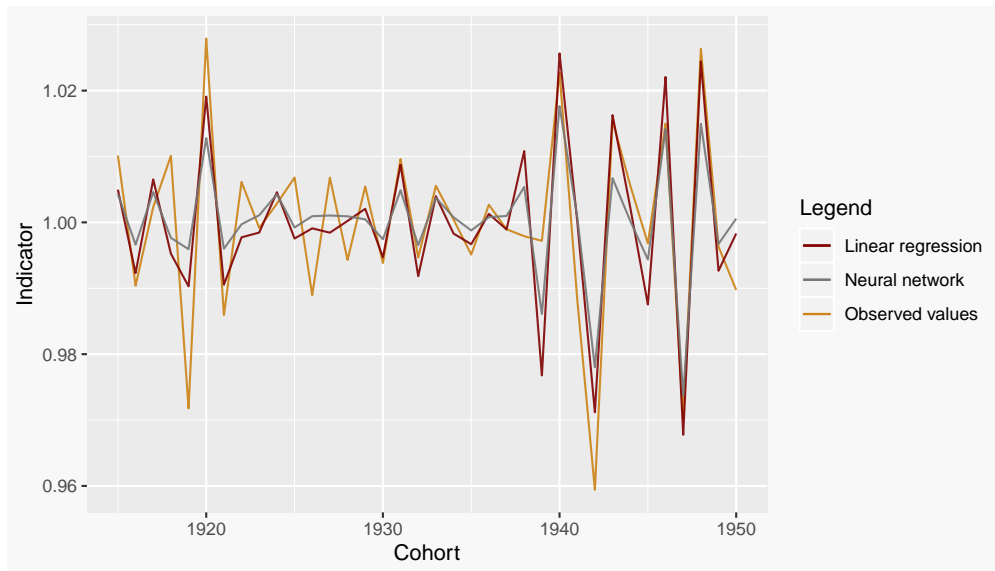


Figure 38: Spain: Predicted correction indicator (averaged by age within each cohort)

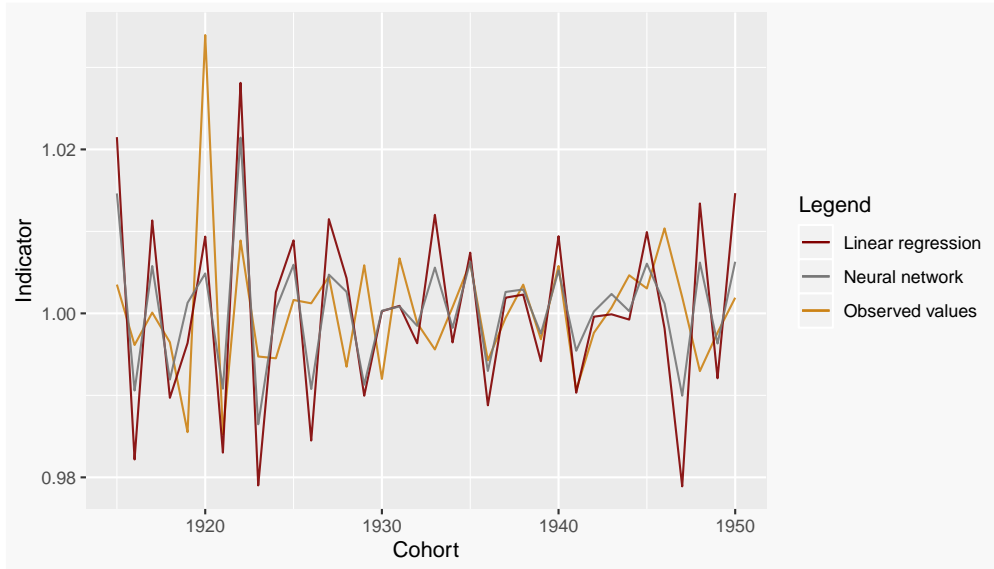


Figure 39: Denmark: Predicted correction indicator (averaged by age within each cohort)

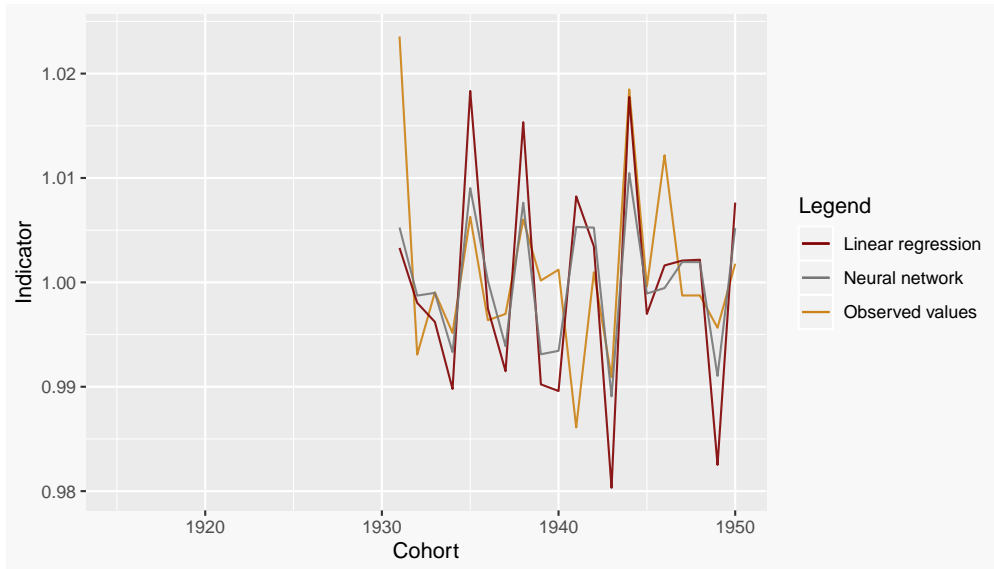


Figure 40: Norway: Predicted correction indicator (averaged by age within each cohort)

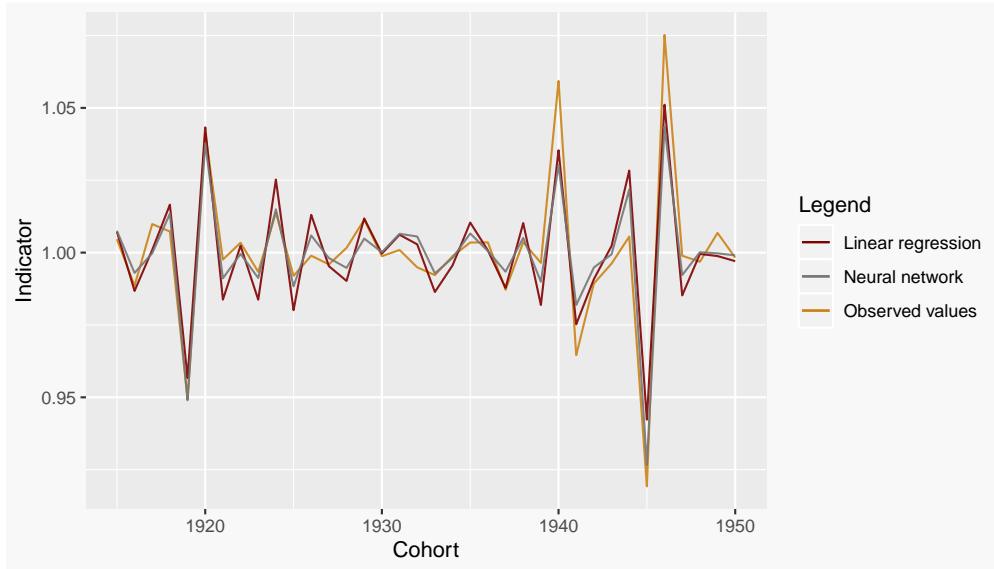


Figure 41: Finland: Predicted correction indicator (averaged by age within each cohort)

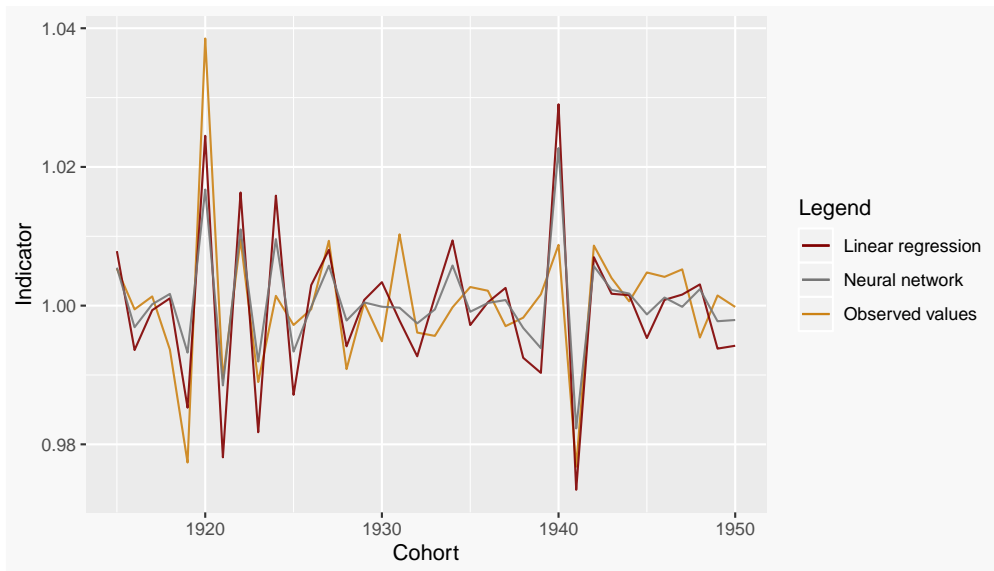


Figure 42: Sweden: Predicted correction indicator (averaged by age within each cohort)

Appendix B Parameters of the M1 model calibrated on raw and corrected mortality data

B.1 Calibration on West Germany mortality data

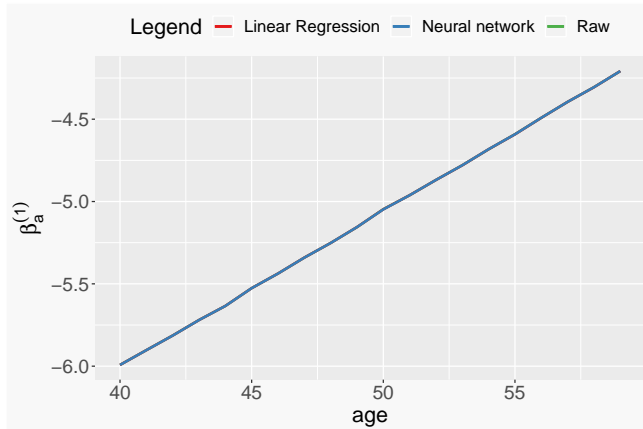


Figure 43: M1 model $\beta_a^{(1)}$ parameter — West Germany raw and retreated mortality data

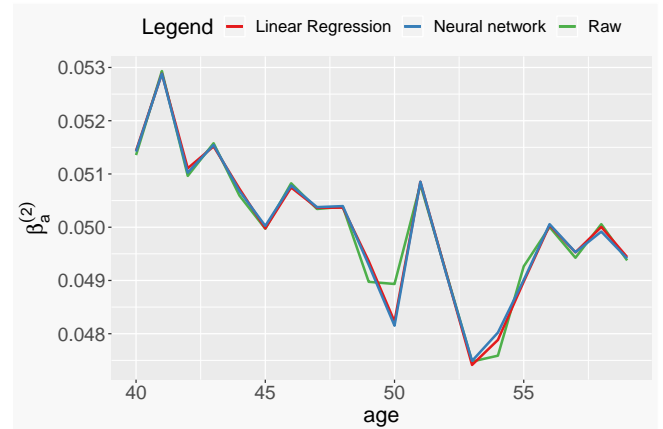


Figure 44: M1 model $\beta_a^{(2)}$ parameter — West Germany raw and retreated mortality data

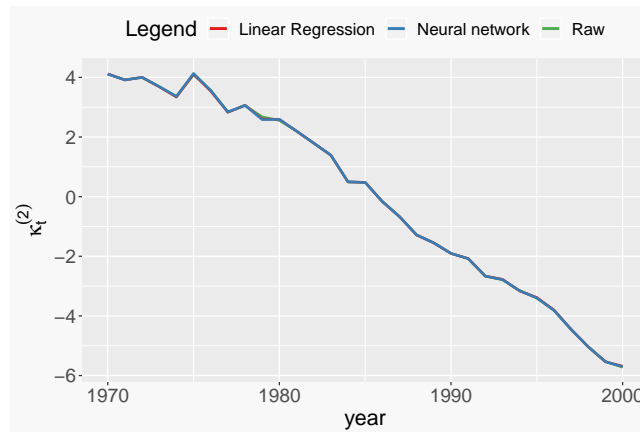


Figure 45: M1 model $\kappa_t^{(2)}$ parameter — West Germany raw and retreated mortality data

B.2 Calibration on Belarus mortality data

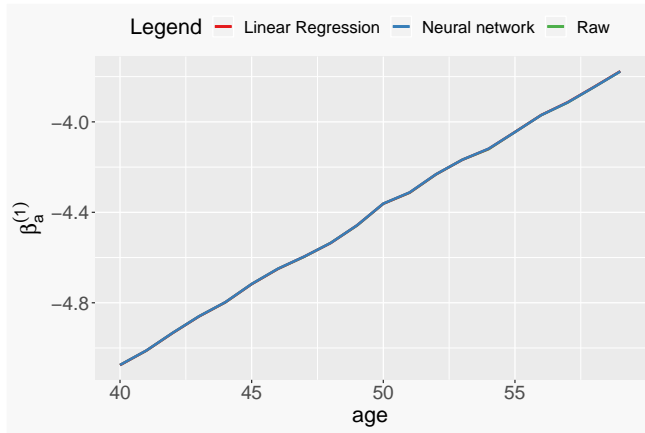


Figure 46: M1 model $\beta_a^{(1)}$ parameter — Belarus raw and retreated mortality data



Figure 47: M1 model $\beta_a^{(2)}$ parameter — Belarus raw and retreated mortality data

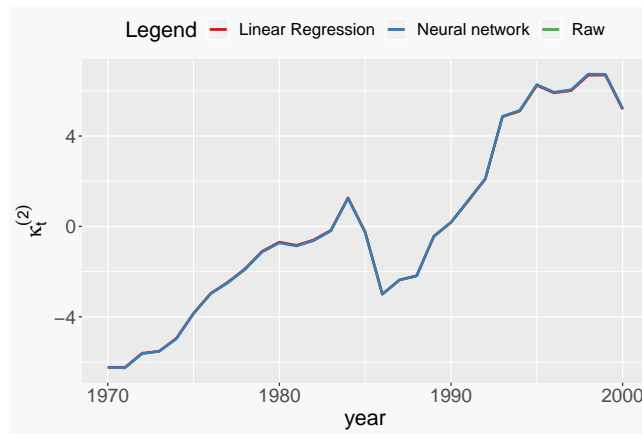


Figure 48: M1 model $\kappa_t^{(2)}$ parameter — Belarus raw and retreated mortality data

B.3 Calibration on Ukraine mortality data

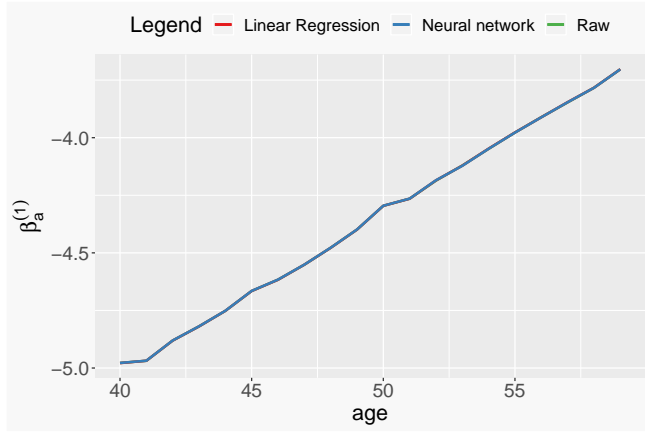


Figure 49: M1 model $\beta_a^{(1)}$ parameter — Ukraine raw and retreated mortality data

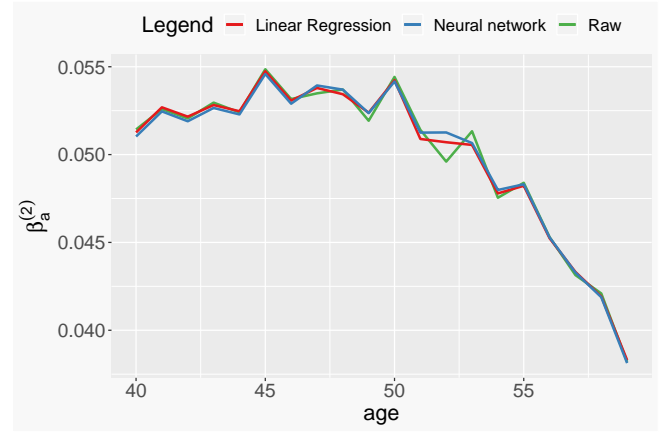


Figure 50: M1 model $\beta_a^{(2)}$ parameter — Ukraine raw and retreated mortality data

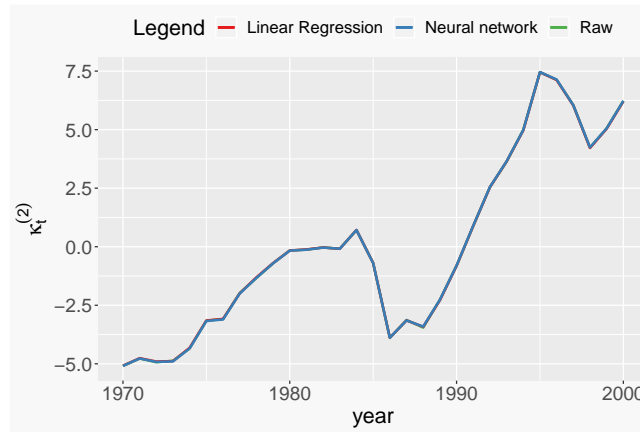


Figure 51: M1 model $\kappa_t^{(2)}$ parameter — Ukraine raw and retreated mortality data

B.4 Calibration on Poland mortality data

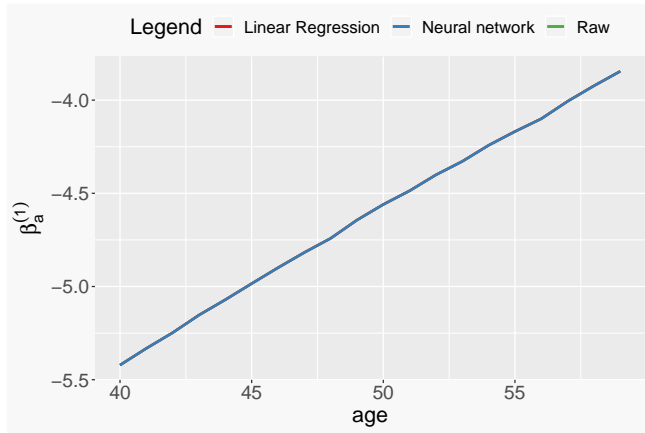


Figure 52: M1 model $\beta_a^{(1)}$ parameter — Poland raw and retreated mortality data

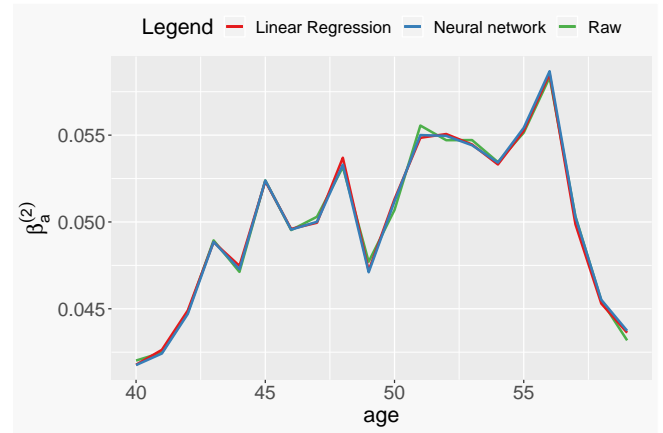


Figure 53: M1 model $\beta_a^{(2)}$ parameter — Poland raw and retreated mortality data

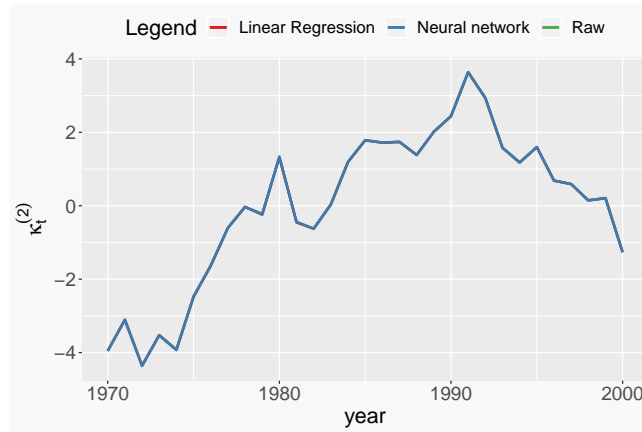


Figure 54: M1 model $\kappa_t^{(2)}$ parameter — Poland raw and retreated mortality data

Appendix C Discussion: Details of HMD version 6 corrected estimates

In the HMD mortality data correction with fertility data (Wilmoth et al., 2007, 2019), the exposures to risk are estimated as follows :

$$\begin{cases} E^{(5)}(a, t) = \frac{1}{2}(P(a, t+1) + P(a, t)) + \frac{1}{6}(D^L(a, t) - D^U(a, t)) \\ E^{(6)}(a, t) = (P(a, t+1) + P(a, t))(1 - m_{t-a}) + \frac{D^L(a, t) - D^U(a, t)}{2(1 - m_{t-a})}((1 - m_{t-a})^2 - \sigma_{t-a}^2) \end{cases} \quad (19)$$

with

$$\begin{cases} m_{t-a} = \mathbb{E}(B(t-a)) \\ \sigma_{t-a}^2 = \mathbb{V}(B(t-a)) \end{cases} \quad (20)$$

where $B(t-a)$ designates the random variable in $[0, 1]$ which represents the distribution of birthdays within annual cohort $t-a$ for a fixed country and a fixed gender. Thus, the following relationship is satisfied:

$$I(a, t) = \frac{(P(a, t+1) + P(a, t))(1 - m_{t-a}) + \frac{D^L(a, t) - D^U(a, t)}{2(1 - m_{t-a})}((1 - m_{t-a})^2 - \sigma_{t-a}^2)}{\frac{1}{2}(P(a, t+1) + P(a, t)) + \frac{1}{6}(D^L(a, t) - D^U(a, t))}. \quad (21)$$

By denoting

$$\begin{cases} P_\alpha(a, t) = P(a, t+1) + P(a, t) \\ D_\alpha(a, t) = D^L(a, t) - D^U(a, t) \\ \beta_{t-a}^2 = \frac{1}{2}((1 - m_{t-a})^2 - \sigma_{t-a}^2) \end{cases} \quad (22)$$

we get

$$I(a, t) = \frac{P_\alpha(a, t)(1 - m_{t-a}) + \frac{D_\alpha(a, t)}{(1 - m_{t-a})}\beta_{t-a}^2}{E^{(5)}(a, t)}, \quad (23)$$

i.e.

$$I(a, t) = A(a, t)(1 - m_{t-a}) + B(a, t)\frac{\beta_{t-a}^2}{1 - m_{t-a}}, \quad (24)$$

with

$$\begin{cases} A(a, t) = \frac{P_\alpha(a, t)}{\frac{1}{2}P_\alpha(a, t) + \frac{1}{6}D_\alpha(a, t)} = \frac{P_\alpha(a, t)}{E^{(5)}(a, t)} \\ B(a, t) = \frac{D_\alpha(a, t)}{\frac{1}{2}P_\alpha(a, t) + \frac{1}{6}D_\alpha(a, t)} = \frac{D_\alpha(a, t)}{E^{(5)}(a, t)} \end{cases} \quad (25)$$

In this formula, the variables m_{t-a} and β_{t-a}^2 only depend on the cohort $t-a$, but the variables $P_\alpha(a, t)$ and $D_\alpha(a, t)$ depends on a and t . This is why there is an age variability of the target variable I (see Figures 4 to 7).