

Mortality improvements in South Africa: insights from pensioner mortality

by Ronald Richman and Gary Velcich

Presented at the Actuarial Society of South Africa's 2020 Virtual Convention
6–8 October 2020

ABSTRACT

The study of mortality improvements in South Africa has been complicated by data limitations: from a population perspective, the data are incomplete and misreported whereas pooled data collected from insurance companies for industry studies often span short time periods and exhibit significant heterogeneity. Notwithstanding these issues, accurate quantification of mortality improvement is critical for actuarial valuations of life insurers and pension funds. In this study, we analyse a unique pensioner dataset, covering the years 2000–2019, through the dual lenses of traditional actuarial analysis and deep learning techniques. We report on aggregate mortality improvements, as well as how mortality improvements differ depending on gender and pension amount. We investigate whether the observed mortality improvements might be attributed to changes in the composition of the pensioner dataset by contrasting these results with those derived from a more complex model that allows for these changes over time. Finally, we show the estimated impact of the observed improvements on pension liabilities and calibrate models to provide uncertainty around the estimated mortality improvement rates, to benchmark the SAM longevity stress and provide a basis for IFRS 17 risk adjustment calculations.

KEYWORDS

Mortality, Lee–Carter model, neural networks, deep learning, IFRS 17, uncertainty quantification

CONTACT DETAILS

Ronald Richman, QED Actuaries and Consultants; Email: ronald.richman@qedact.com

Gary Velcich, Alexander Forbes; Email: velcichg@forbes.com

1. INTRODUCTION

Mortality improvement assumptions are basic inputs into actuarial models used in life insurance and pensions practice and the evaluation of damages, and stressed scenarios involving these assumptions form a significant part of the reserving and capital requirements for life insurance business written in South Africa (FSB, 2014b). Updated insurance accounting standards (IFRS 17), which are applicable both to risk and annuity business, will require insurers to make an allowance for the risk that assumptions on trends in mortality rates deviate from expectations (IASB, 2020). Nonetheless, there is little publicly available empirical information on mortality improvements in South Africa: at a population level this is probably due to the problems involved in analysing incomplete and misreported population exposure and death data whereas industry studies of mortality suffer from changes in the mix of business from different companies and employers, making these data difficult to aggregate. Thus, unless company specific data for performing an analysis is available, actuaries working in practice in these fields are forced to combine the limited empirical information available with expert judgement, informed by international studies, when setting mortality improvement assumptions. Moreover, despite the requirements within insurance practice imposed by the Solvency Assessment and Management (SAM) regulations to perform an Own Risk and Solvency Analysis (ORSA) (Prudential Authority, 2018b), reliance is often placed on standard parametrisations of the distribution of mortality improvements based on European data, which, if done, falls short of the SAM requirements.

Relatively few studies have addressed the issue of mortality improvements in South Africa. Focussing on insured lives or pensioners, the most recent report on pensioner mortality in South Africa (CSI, 2017), covering mortality in the years 2005–2010, provided graduated pensioner mortality tables, but did not address the issue of trends in pensioner mortality. Similarly, a recent study of annuitant mortality (CSI, 2012) was not able to estimate mortality improvements because the data available were too heterogeneous and an earlier study setting out standard tables for annuitant mortality by Dorrington & Tootla (2007) concluded that the data available for these tables did not allow for the estimation of mortality improvement rates. Thus the South African actuarial literature does not provide a strong empirical basis for setting mortality improvement assumptions.

At a population level, Richman (2017) studied the mortality of the oldest old, that is, those aged 75+, in the South African population, by first correcting the death data for incompleteness of reporting and then applying demographic methods to reconstruct the population based only on death data. Mortality improvement rates at a national level of 0.7% and 0.1% per annum over the period 1985–2011, for males and females respectively, were found. Considering this at the level of the population groups, it was found that mortality

rates were not improving for the African population group, whereas other population groups were experiencing mortality improvements.¹ These improvement rates present two main shortcomings: improvement rates are required at ages younger than 75 especially for annuity business, and improvement rates for the insured population may differ from improvement rates for the population nationally, because of the selective effect of underwriting and the better relative socio-economic status of those able to afford insurance. Other sources on trends in population mortality in South Africa are also reviewed in Richman (2017), however, the results presented in these sources are also not ideally suited for use by actuaries.

In this study, we aim to provide empirical information on mortality improvements in South Africa that occur after retirement, taking a step towards addressing some of the shortcomings of the information currently available. This analysis is based on a large pensioner dataset provided by a pension administration company, which contains exposure and death information covering the period 2000–2019. This is considered to be a long enough period of observation to estimate trends in mortality. The data contain other descriptive information from which we can derive more detailed mortality improvement information based, primarily, on proxies for socio-economic level. To the extent that practising actuaries believe that the experience of this pensioner portfolio is representative of the mortality experience of the portfolios for which they are responsible, then the mortality improvement rates and uncertainties quantified in this study should enable a more rigorous process of assumption setting for both pricing and reserving and allow better compliance with regulation.

The problem of deriving mortality rates and improvements is tackled in this study using two broad categories of techniques, referred to as: “traditional” techniques and “deep learning” techniques. The former includes standard actuarial and demographic techniques, such as the analysis of age standardised mortality rates and the application of the Lee–Carter model (Lee & Carter, 1992). The latter refers to the modern implementation of deep neural networks, which have been shown in recent studies to have the potential to outperform traditional mortality forecasting methods (Perla et al., 2020; Richman & Wüthrich, 2019). Traditional techniques are applied to derive mortality rates for the overall pensioner portfolio, for both genders, as well as at a more granular level of disaggregation by pension amount. The deep learning techniques are used to compare forecasts to those of the Lee–Carter model, as well as to build a mortality forecasting model using all the descriptive covariates in the pensioner dataset.

The rest of the paper is organised as follows. Section 2 reviews the data and methods used in this study and provides notation and definitions. In section 3 an exploratory analysis of the data is provided. Section 4 focuses on modelling and forecasting mortality in each quartile of the pension amount distribution using the Lee–Carter model and a simple neural network. In section 5 a more complex neural network is applied to these data and key results are extracted. Uncertainty around the mortality improvement rates is covered in section 6,

¹ See Moultrie & Dorrington (2012) for a discussion of the use of race in demographic analysis in South Africa.

and our findings and conclusions are discussed in section 7. The appendices to the paper provide code for the neural networks and extra figures relating to the results of the study.

2. DATA, DEFINITIONS AND METHODS

2.1 Data source

The data used in this study consists of exposure and death information provided by one of the largest retirement fund administrators in South Africa, for the period from 1 January 2000 to 31 December 2019. The underlying pensioner data are maintained as part of the ongoing administration and pensioner payroll services provided by the administrator to both defined benefit and defined contribution pension funds. The data are limited to those pensioners that are paid in-fund, and accordingly excludes retirees who purchased an annuity outside of the fund (which would be typical of most defined contribution funds). One result of this is that the dataset shows an increase in the average age of pensioners over the period, given the reducing number of defined benefit funds in South Africa and the limited number of defined contribution funds that provide in-fund annuities.

The data were processed to derive central exposure to risk figures in each year for each life, using age last birthday as the age definition. The usual descriptive variables of age (last birthday) and gender are available for each life. In addition, several other variables are provided in the data relating to the size of the pension in payment, the industry in which the pensioner worked before retirement, whether retirement was early, whether the pensioner is married, the residential address of the pensioner and whether the pensioner is a member of a medical scheme. An exploratory analysis of the dataset is provided in the next section.

We utilise models of mortality rates to investigate the extent to which mortality improvements were observed in the dataset. The neural network models (see section 2.3.2) used in this study are highly flexible and can fit the data on which they are estimated almost exactly. As a result it is common practice to divide the data into two disjoint subsets: a training dataset on which the models are estimated, and a testing dataset which is used to ensure that the models generalise, in other words, predict out-of-sample data well. In the case of mortality forecasting, it must be shown that the models produce high quality forecasts in unseen time periods, thus, in this study the training data is defined as the data covering the period from 2000–2014 and the testing data cover the period 2015–2019. We select a random sample of 5% of the training dataset to track the expected out-of-sample performance of the neural network models as they are trained; this random sample from the training data is referred to as the validation set in what follows.

2.2 Definitions

This study focuses on modelling the observed force of mortality at age x at time t , $\mu_{x,t}$, and mortality improvement rates, defined as $\kappa_{x,t} = \frac{\mu_{x,t}}{\mu_{x,t-1}} - 1$. x is limited to the age range 50–99 and t spans the years 2000–2019. Values of $\kappa_{x,t}$ different from 0, which imply either

mortality improvement or worsening, may, on the one hand, result from changes over time in the composition of the population under investigation. For example, it is well known that pensioners with a higher absolute pension amount experience lower mortality. If, over time, pensioners with a higher pension come to comprise a greater proportion of the population, then what appears to be mortality improvement might be observed, but in reality, these improvements would be attributable to the changes in the population under investigation. On the other hand, there may be improvements in mortality that are driven by factors that do not depend on the composition of the portfolio, for example, improvements in medical care and technology. This latter type of mortality improvement, that does not depend on changes in the composition of the population, is the primary concern of the study. The models we employ accordingly control for the composition of the population in as detailed a manner as possible. These models utilise the extensive descriptive information that is available to segment the pensioner portfolio used in this study into sub-populations. Thus, we extend the notation defined above to $\mu_{x,t}^p$ and $\kappa_{x,t}^p$, where p represents one of P sub-populations. The sub-populations under consideration are defined in more detail in the sections that follow.

2.3 Methods

Observed mortality rates, especially those of smaller populations, or of insured and pensioner populations, are often quite variable. This can obscure underlying patterns of mortality rates, or make it difficult to estimate mortality improvement rates accurately. Thus, instead of analysing observed mortality and improvement rates directly, the study focuses on estimates of mortality rates, $\hat{\mu}_{x,t}^p$, derived using suitable regression models. The estimated

rates of mortality improvement, $\hat{\kappa}_{x,t}^p = \frac{\hat{\mu}_{x,t}^p}{\hat{\mu}_{x,t-1}^p} - 1$, can then be derived using these estimates.

Two main regression techniques are employed in this study to model mortality rates: the Lee–Carter (LC) model of Lee & Carter (1992), and deep neural network models similar to those used in Richman & Wüthrich (2019). Age standardised mortality rates (ASMRs) are also considered in exploring the data.

As noted above, the neural network techniques applied in this study are highly flexible, thus, there is a risk that spurious estimates of $\hat{\kappa}_{x,t}^p$ could (theoretically) be derived if the neural network models over-fit the training data. To mitigate this risk, conclusions are drawn only about mortality improvements based on neural networks models that are shown to generalise well to unseen data.

2.3.1 TRADITIONAL MODELS

The exploratory analysis in section 3 estimates ASMRs to explore whether there is evidence of mortality improvements in the data. The average population at each age within

the dataset is defined as $N_x^p = \frac{\sum_{t=1}^T N_{x,t}^p}{T}$, where N_x^p is the average number of exposure years

at age x , taken as the average over the number of exposure years at age x in year t , $N_{x,t}^p$. Then, the ASMR for population p at time t is defined as:

$$ASMR_t^p = \frac{\sum_{x=50}^{99} \mu_{x,t}^p N_x^p}{\sum_{x=50}^{99} N_x^p}.$$

The LC model estimates (log) mortality rates as (note that the population notation has been dropped for clarity and we do not include an error term since we work with estimated rates):

$$\log(\hat{\mu}_{x,t}^{LC}) = a_x + b_x k_t,$$

where a_x represents the average mortality at age x in the time period under consideration, b_x is the rate of change of mortality at age x and k_t is the value of a time index at time t , applicable to all ages, which determines the extent to which mortality rates have changed over time. The LC model is quite parsimonious, requiring only $2X + T$ parameters to fit XT observed mortality rates. However these parameters cannot be inferred from the data using linear regression techniques, since the quantities are not directly observed. Rather they must first be estimated using Principal Component Analysis (PCA) or the model must be fitted using non-linear regression. To produce forecasts, the LC model is first fit to a dataset, and then the k_t parameters are extrapolated, usually using time series methods. In this study, the LC model is fit to the observed mortality rates in the period 2000–2014 using PCA and then is extrapolated to the years 2015–2019 using a random walk with drift.

2.3.2 Neural networks

As noted before, the study is concerned with modelling mortality in several sub-populations, mainly to ensure that estimated mortality improvements are not being caused by changes in the composition of the portfolio. The LC model can be applied without modification to the mortality experience of each sub-population. However this may produce poorer quality results due to lower volumes of data leading to forecasts of poor accuracy. This problem is exacerbated by the goal of the modelling in this study, which is to control the composition of the population in as detailed a manner as possible. This could potentially lead to sub-populations with relatively few deaths or low exposures. Instead of the standard LC model, multi-population extensions of that model could be considered, see, for example, Li & Lee (2005) and Kleinow (2015). These would however not solve the problem of low exposures (this is likely a problem that would be experienced with most traditional models of mortality) and these models have been shown in Richman & Wüthrich (2019) to perform quite poorly as measured by out-of-sample accuracy.

Thus, instead of traditional statistical models, we turn rather to deep neural networks. In the area of machine learning, deep neural networks have been successfully applied to model very high-dimensional tasks in computer vision and natural language processing (LeCun et al., 2015) and, in the area of actuarial science, have produced significantly more accurate mortality models than competing traditional models when applied to forecasting

population mortality (Richman & Wüthrich, 2019; Perla et al., 2020). For a full overview of the modern application of neural networks, the reader is referred to Goodfellow et al. (2016), and for a review in the context of actuarial science, to Richman (2018).

Neural networks are a class of highly flexible models that is distinguished from other machine learning techniques by the implementation of the paradigm of representation learning (Bengio et al., 2013). This is the automatic derivation of new sets of predictive variables from the data input into the model. The model then makes its predictions based on these new predictive variables. An example of representation learning in the wider machine learning literature is found in the field of computer vision: for the task of image classification, a specialised type of neural network processes image data into a vector of predictive variables which are highly correlated with the objects appearing within the image. The model then predicts the class of image on the basis of these variables, for example, the model may predict that the image is of a motor vehicle. Turning to a field closer to this study's application, an example of representation learning in actuarial science is the neural network used in Perla et al. (2020), which takes as its input a matrix of mortality rates spanning the age range 0–99 and covering 10 years of experience. The matrix is then converted by a neural network into new predictive variables that are used to forecast mortality rates using a Generalised Linear Model (GLM).

To define neural networks formally, the covariate information, including the variables defining the mortality rate being estimated (age and year) as well as the other variables used to define the sub-population p (such as gender, pension amount, medical aid status etc.), is organised into a matrix X which contains the values these variables take. This matrix of predictor variables is indexed by $i \in \{1 \dots I\}$ where I is the number of records in the data, and the columns are indexed by $j \in \{1 \dots J\}$, where J is the number of predictive variables.

A neural network (formally, a fully-connected neural network²) provides a non-parametric estimate of a function f that maps from the predictor variables X to the outcome of interest, this being mortality rates at age x in time t for sub-population p , as follows:

$$\log(\hat{\mu}_{x,t}^{NN}) = f(X).$$

The function f is composed of a series of non-linear transformations of the input data X . Focusing on one row of the predictor variables, $\mathbf{x}_p$, the first of these transformations takes the predictor variables as input and returns a new vector of variables, $\mathbf{z}^1$, by applying the following function:³

$$\mathbf{z}^1 = \sigma_0(\mathbf{a}_0 + B_0' \mathbf{x}_i), \quad (1)$$

where $\mathbf{a}_0$ is a vector of intercept parameters, B_0 is a matrix of regression parameters and σ_0

2 See Richman (2018) for an overview of other types of neural network that are not fully connected or contain recurrent connections.

3 Note that vectors are shown in bold and matrices are represented using capital letters.

is a non-linear function, referred to as an “activation” function and here set equal to the hyperbolic tangent function $\sigma_0 = \frac{e^x - e^{(-x)}}{e^x + e^{(-x)}}$, though there are other possible choices. The length of the vector of new variables, $\mathbf{z}^1$, is chosen by the modeller and is referred to as the width of the neural network. The rest of the neural network is comprised of several “layers” of functions that are similar to equation (1), except that these functions take the variables derived at the previous step as their input. For example, the n th function is defined as:

$$\mathbf{z}^n = \sigma_{n-1}(\mathbf{a}_{n-1} + \mathbf{B}'_{n-1}\mathbf{z}^{n-1}) \quad (2)$$

Finally, the output of the neural network, here the estimated mortality rates, is calculated using the variables estimated using the last of the intermediate functions represented in equation (2):

$$\hat{\mu}_{x,t}^{p,NN} = \mathbf{z}^{n+1} = \sigma_n(\mathbf{a}_n + \mathbf{B}'_n\mathbf{z}^n) \quad (3)$$

where $\mathbf{z}^n$ can be summarised as the output of a function comprised of multiple layers of the form of equation (2):

$$\mathbf{z}^n(\mathbf{x}_i) = (\mathbf{z}^n \circ \dots \circ \mathbf{z}^1)(\mathbf{x}_i) \quad (4)$$

Equation (3) can be seen to be nothing more than a Generalised Linear Model (GLM), with link function σ^{-1} , estimating mortality rates on the basis of a set of predictor variables $\mathbf{z}^n$ that have been derived using a function of the form of equation (4). Returning to the earlier definitions, the neural network has performed representation learning by deriving a new set of variables, $\mathbf{z}^n$, from the input predictors X , that are used to estimate mortality rates. For more on the connection between neural networks and GLMs, see Wüthrich (2019).

2.3.3 EMBEDDING LAYERS

In the previous section, X was defined as the matrix of predictor variables that are used to estimate mortality rates. The dataset has a combination of continuous and categorical predictor variables. Examples of the former are age last birthday⁴ and the year in which mortality rates are estimated. Examples of the latter are gender and industry in which the pensioner worked. In statistical modelling, categorical predictor variables are usually modelled with indicator variables which are set to 1 when the category corresponding to that indicator variable is present, and 0 otherwise (dummy-coding and one-hot encoding are two ways in which this can be done). For categorical variables with multiple levels, the transformation to a series of indicator variables results in a large and sparse matrix consisting mainly of zeroes, which can create problems when fitting models.

4 Age can be modelled both as a continuous variable and also as a categorical variable, and the latter approach is what is done in this paper, as described in what follows.

An alternative coding scheme that has been used successfully in the machine learning community (De Brébisson et al., 2015; Guo & Berkhahn, 2016), and in recent actuarial studies (Richman, 2018; Kuo, 2019) is embeddings (see Bengio et al. (2003) for the original use of embeddings in natural language processing). An embedding maps categorical data to low dimensional numerical vectors, the parameters of which are estimated when the model in which the embeddings appear is fit. For example, consider a categorical variable with P different levels. Whereas traditional models would require P different indicator variables, an embedding will map each of the levels, p to a new vector $\mathbf{p}$ with dimension k that can be chosen to be much smaller than P .

In the application of neural networks in this study, all categorical variables are mapped to embedding layers with three dimensions, and so the original matrix of predictors X is replaced by an extended matrix, X^* , that consists of the same continuous variables as X and embeddings for the categorical variables. Equation (4) can accordingly be stated more precisely as:

$$\mathbf{z}^n(\mathbf{x}_i^*) = (\mathbf{z}^n \circ \dots \circ \mathbf{z}^1)(\mathbf{x}_i^*)$$

2.3.4 FITTING THE NEURAL NETWORKS

Two different neural networks are applied in this study: a simpler model described in section 4 and a more complex model described in section 5. Both of these networks have three intermediate layers of width 32 (in other words, the dimension of each of the intermediate layers of the network is 32) and embeddings for each categorical variable of dimension 3. These “hyper-parameters” were set subjectively, by considering the amount of data within the pensioners dataset and could be optimised further. Nonetheless the performance of the networks is deemed to be satisfactory with reference to the forecasting performance of the LC model. The neural networks are fitted by specifying the Poisson deviance loss function (see Wüthrich & Büser (2018) for more detail), in other words, the networks are fit on the assumption of a Poisson distributed outcome variable, which is usually a reasonable assumption when modelling death counts, see, for example, Currie (2016). The exposure is multiplied by the estimated mortality rates produced by the network to produce an expected number of deaths. The Poisson deviance loss function is used to measure the difference between the observed and the expected number of deaths, and the deviance is minimised by calibrating the parameters of the network.

To avoid the problem of over-fitting the training data, the neural networks are regularised using two commonly applied techniques from the machine learning literature: dropout (Srivastava et al., 2014) and batch normalisation (Ioffe & Szegedy, 2015). During the parametrisation of the network, dropout randomly sets part of the vector of intermediate variables in equation (2), to zero, which empirically is shown to produce networks that generalise well. Srivastava et al. (2014) explain that dropout prevents the intermediate variables of the network from jointly over-fitting the training data. Batch normalisation, on the other hand, calculates the z-scores of the intermediate variables (that is, subtracts the mean and divides by the standard deviation) and then rescales the z-scores using new parameters

that are estimated from the training data. Empirically, applying batch normalisation makes networks easier to optimise and also has a regularising effect; see, for example, Luo et al. (2019).

Other standard practices in fitting modern neural networks are also applied. All continuous predictor variables are mapped to the range $[0,1]$ before they are input to the network. Stochastic mini-batch gradient descent is utilised with batch sizes of 64 for the simpler neural network fit to the same data as the LC model, and 4096 for the more detailed model. The Adam optimiser of Kingma & Ba (2014) is used to calibrate the network, with the default parameters suggested in that study and a standard learning rate (which controls the magnitude of updates to the network parameters) of 0.001. The expected generalisation performance is tracked on a subset of the training data (called the validation set) and the learning rate is decreased by a factor of 0.9 when the loss function measured on the validation set plateaus. The model parameters producing the lowest validation error during the training epochs are saved and used as the final parameters for that training run.

Finally, it has been observed that the estimates produced by neural networks vary by training run. This problem has been addressed in detail in Richman & Wüthrich (2020), who suggest that the averaged estimate of multiple networks be used as a network aggregation or “nagging” estimator (in the spirit of the “bagging” estimator of Breiman (1996)). Here, we average the estimates from 20 neural network models fit to the same data, and consider the performance of this averaged estimate when reporting the performance of the networks in this paper.

The neural networks are fit in the R programming language using the Keras package (Allaire & Chollet, 2017). Code for defining the networks can be found in appendix A.

3. EXPLORATORY ANALYSIS

3.1 Overview of the dataset

The dataset is mainly comprised of the experience of pensioners in South African defined-benefit pension funds. The majority of these funds are closed to new entrants, thus, the population within these funds is expected to show an increasing average age over time. The number of pensioners is correspondingly expected to reduce, offset somewhat by the administrator’s take on of new business over the period. Table 1 summarises the aggregate experience.

The relative increase observed in the number of female pensioners over the period is due to the fact that, with joint-life pensions, male partners tend to predecease their female partner (as a result of both higher male mortality, and being on average older), and thus, in the case of joint life pensions, the surviving female partners being added to the population over time.

Following page: TABLE 1 Exposure, deaths, average age and overall mortality rates, for males and females separately, in the period 2000–2019

	Central Exposure	Deaths	Average Age	Overall Mortality Rate
Female				
2000	23,375	807	67.88	0.035
2001	26,672	988	68.49	0.037
2002	27,040	1,069	68.88	0.040
2003	28,496	1,058	69.21	0.037
2004	28,393	1,066	69.60	0.038
2005	28,844	1,157	70.02	0.040
2006	30,383	1,259	70.32	0.041
2007	32,346	1,350	70.63	0.042
2008	34,129	1,368	70.99	0.040
2009	33,642	1,399	71.42	0.042
2010	33,811	1,346	71.84	0.040
2011	34,795	1,568	72.32	0.045
2012	34,072	1,658	72.71	0.049
2013	34,276	1,691	73.09	0.049
2014	34,041	1,751	73.47	0.051
2015	37,755	2,142	73.56	0.057
2016	38,828	2,092	73.72	0.054
2017	37,659	2,140	74.09	0.057
2018	36,741	2,145	74.38	0.058
2019	33,548	2,018	74.72	0.060
Male				
2000	23,777	1,468	69.17	0.062
2001	25,223	1,529	69.74	0.061
2002	24,250	1,480	70.25	0.061
2003	24,604	1,483	70.63	0.060
2004	23,798	1,416	70.93	0.060
2005	26,727	1,617	71.20	0.061
2006	28,379	1,790	71.47	0.063
2007	28,017	1,752	71.95	0.063
2008	27,258	1,726	72.45	0.063
2009	25,683	1,629	72.92	0.063
2010	24,676	1,483	73.37	0.060
2011	24,278	1,562	73.83	0.064
2012	22,993	1,633	74.23	0.071
2013	22,374	1,613	74.54	0.072
2014	21,471	1,597	74.82	0.074
2015	23,118	1,688	74.66	0.073
2016	23,206	1,743	74.65	0.075
2017	21,909	1,695	74.88	0.077
2018	20,983	1,547	75.00	0.074
2019	18,558	1,469	75.34	0.079

3.2 Overall level of observed mortality

Mortality rates calculated on an amounts basis, where the number of deaths and exposure is weighted by pension amount, are generally observed to be lower than those calculated on a lives, or person-year, basis. This is because pensioners with higher pension amounts will tend to have better socio-economic conditions (leading to lower mortality) than those with lower pension amounts. Figure 1 shows that the raw mortality derived from the dataset, using all years of data, for males and females separately and on a lives and amounts basis, conforms to this expectation.

The figure also shows that rates for males are higher than the rates for females at almost every age. It can also be seen that the (log of) mortality rates are relatively level until age 60, after which they increase relatively linearly up to the highest ages of 95, after which the rates are too variable to discern a trend associated with age. The initial flat mortality rates are perhaps due to the effect of higher mortality of the early retirees who might be retiring due to ill health; this point is further addressed in section 3.4.

The results presented in this study are on a lives basis, but allow for the dependence of mortality on pension amount by creating a categorical variable indicating into which quartile of the pension payment distribution a pensioner falls. Due to the high inflationary environment in South Africa, the quartiles of the pension amounts distribution is expressed in 2019 monetary terms. These bands are shown in the following table.

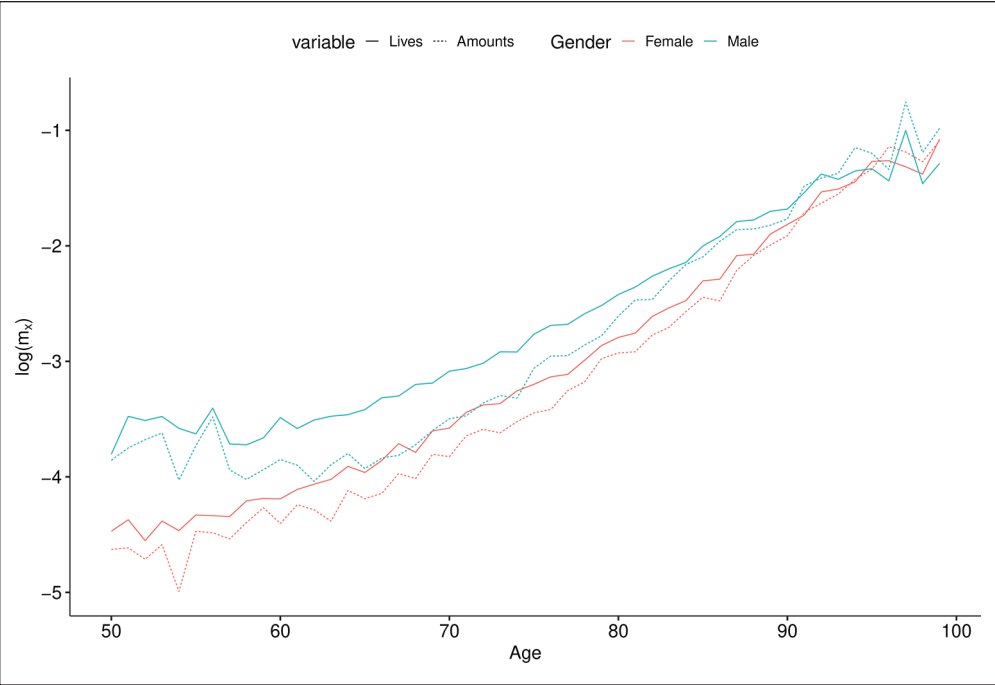


FIGURE 1 Observed mortality rates, males and females, 2000–2019, on a lives and amounts basis separately

TABLE 2 Ranges into which the pension amounts are allocated to derive quartile groups, expressed in 2019 monetary terms

Group number	Group range
1	[0, 1 501)
2	[1 501, 3 531)
3	[3 531, 7 818)
4	[7 818, ∞]

A minority of records do not have a recorded pension amount and are allocated to the first quartile group with a pension amount of 0. Since it is expected that the quartile into which the pension amount falls is to be strongly predictive of mortality rates, we also show the raw mortality rates calculated for each of these groups in figure 2. The raw mortality rates (on both a lives and amounts basis) are lower in each subsequent quartile group, that is, as pension amount increases. The figure shows that, as would be expected, the difference between the lives and amounts basis is significantly smaller than the grouped data shown in figure 1, except in the last quartile group (probably because of very large pension amounts in the highest quartile group).

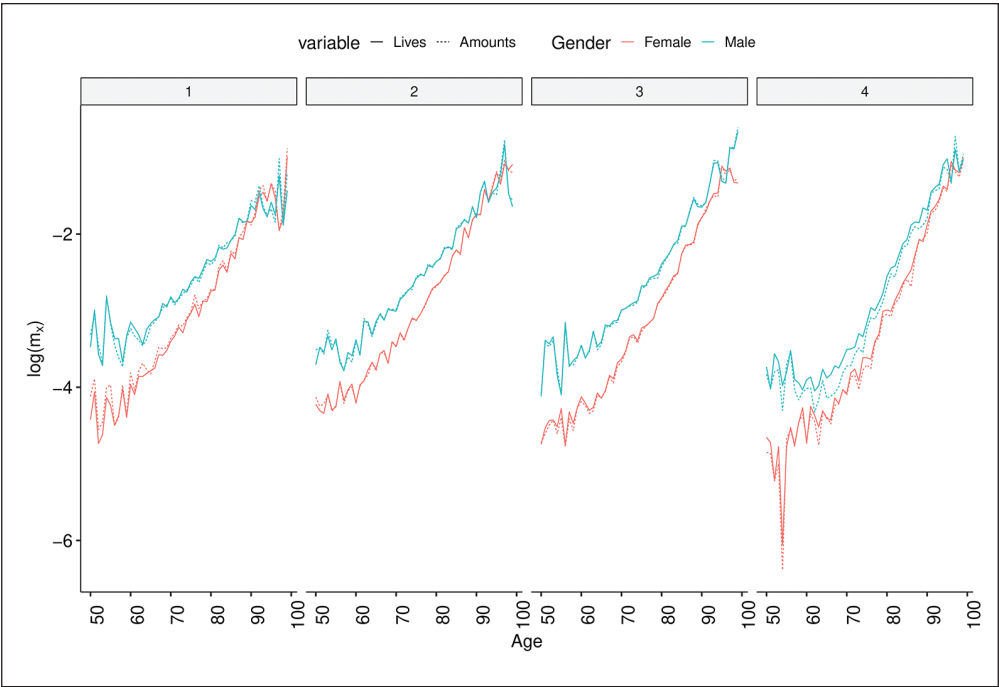


FIGURE 2 Observed mortality rates, males and females, 2000–2019, on a lives and amounts basis separately. Each quartile group is shown in a separate panel

3.3 Age standardised mortality rates

The calculation of age standardised mortality rates (ASMRs) is outlined in section 2.3.1. Figure 3 shows the average population used to calculate the ASMRs. It can be seen that, on average, the female population is larger than the male population, and that the majority of exposure is between the ages of 60 and 90, with notably less exposure at ages above and below this age range.

To calculate the ASMR, unrealistically large mortality rates arising from deaths occurring in groups with small exposure were excluded (specifically, those with $\hat{\mu}_x > 10$, which are not demographically reasonable⁵). The ASMRs for the aggregated dataset are shown in figure 4 and show evidence of varying levels of improvements in mortality over the period. The figure shows that the trend seems to have changed around 2010 for both males and females: before this year, there was a continuing trend of mortality improvement, whereas after 2010, mortality appears to have worsened and then remained at the same level. Interestingly, improvements in male life mortality are consistently more significant than those for female lives, in line with the findings in Richman (2017) for all population groups in South Africa up to the year 2011. The observation that mortality of males is falling more rapidly than that

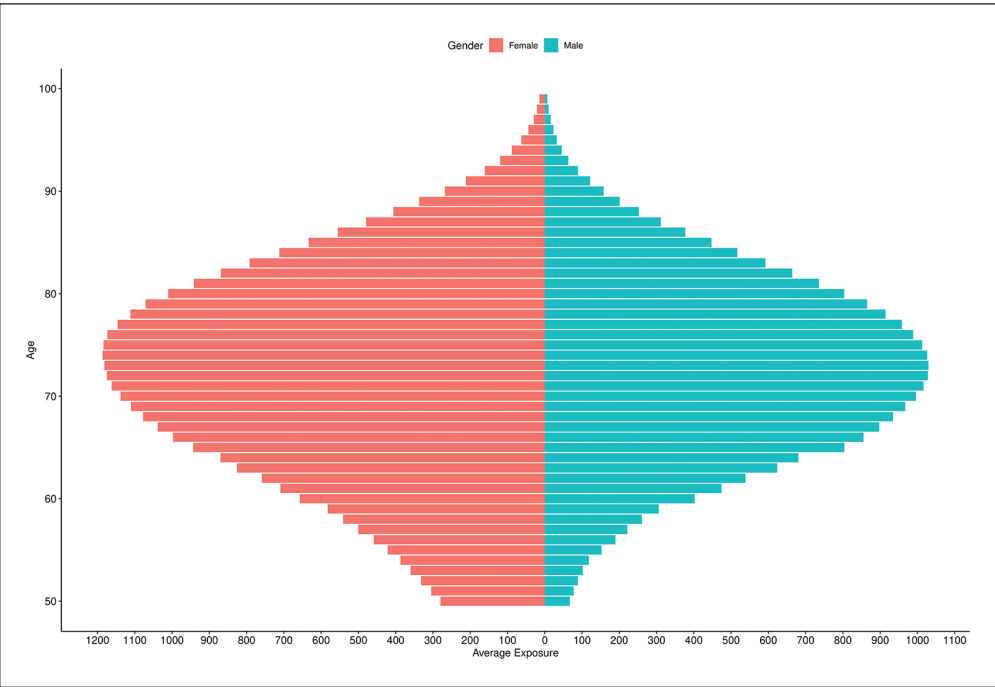


FIGURE 3 Average exposed to risk in each age over all years 2000–2019, males and females separately

5 Note that the force of mortality, μ_x , may exceed 1, whereas the annual probability of death, q_x , cannot exceed 1. Here, we exclude cases where $\hat{\mu}_x$ is large enough to cause distortions in the ASMRs.

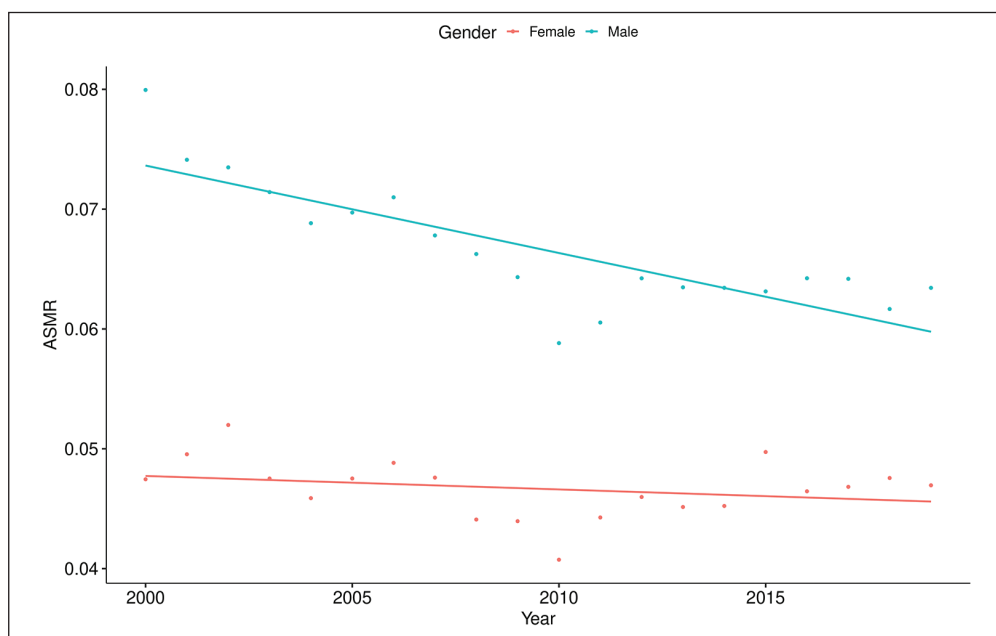


FIGURE 4 Age standardised mortality rates, males and females, 2000–2019 and linear regression trend lines

of females is similar to findings in some international studies which show similar or better levels of mortality improvement for male lives, such as the study for England and Wales in CMI (2020). The gender-specific levels of improvement are lower in South Africa than what was observed internationally for the period, particularly for female lives.

It is possible that the heterogeneity in the dataset (that is, changes in the proportions of pensioners) masks some of the underlying improvement. For example, figure 5, which only covers the experience of those retiring between ages 60 and 66 inclusive (and so excludes early retirements potentially occasioned by ill-health) shows steeper trend lines than figure 4. Considering ASMRs for each of the pension amount quartile groups (for retirements aged 60 to 66) in figure 6 shows that the most significant of the mortality improvements within the data are for males in the first quartile group, that is, those with the lowest pension amount, whereas the improvements for males in the other three groups are less pronounced. For females, mortality rates have improved much more slowly than for males in all quartile groups.

3.4 Early retirement and the possible effect of ill health retirement

The observed experience includes early retirees that may possibly have retired in ill-health. To understand if observed mortality is dependent on whether a pensioner retired early, and how long this effect might last, we estimated the observed mortality for four categories of pensioners: those retiring very early (before age 50) (these most likely represent

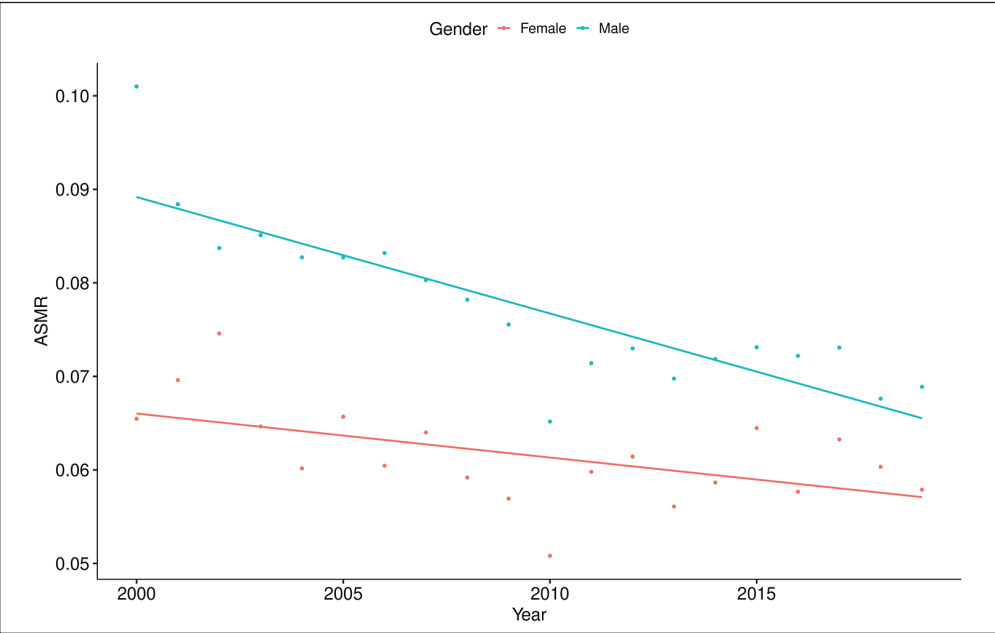


FIGURE 5 Age standardised mortality rates, normal retirements only, males and females, 2000–2019 and linear regression trend lines

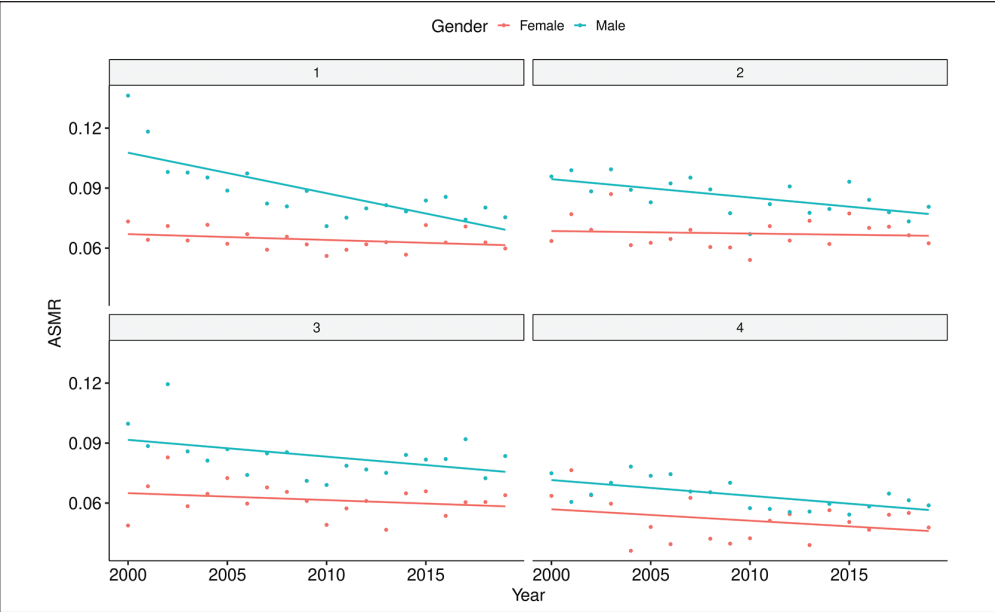


FIGURE 6 Age standardised mortality rates, normal retirements only, males and females in each of the four pension amount groups, 2000–2019 and linear regression trend lines. Each quartile group is shown in a separate panel.

be death-in-service pensions provided to a spouse, rather than the retirement of the main pensioner); those retiring early (50–59); those retiring within the age range 60 to 66 in which the normal retirement age for South African retirement funds is generally set; and those retiring after normal retirement age (67+). The average observed mortality in age bands spanning five years is shown in figure 7.

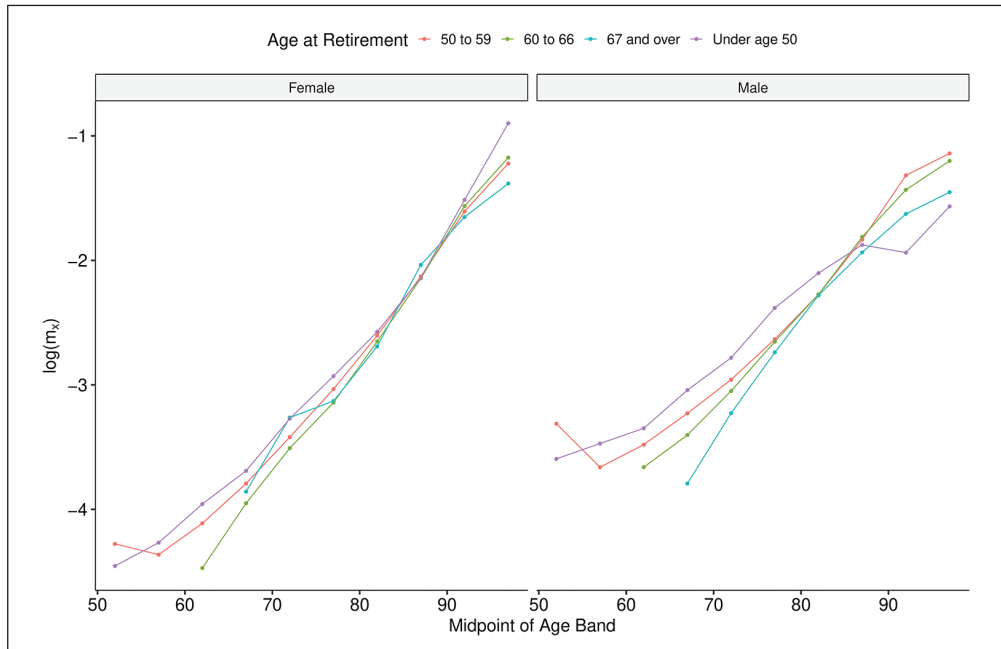


FIGURE 7 Average mortality rates, males and females in each of the four retirement age categories, 2000–2019

Both for males and females, the age at which retirement occurs strongly predicts the level of mortality up to the age band 80–84, with retirement before age 60 being predictive of higher mortality rates for both sexes. For males, those retiring at older than the normal ages (67+) and at the normal ages (60–66) have the lowest mortality, and for females, those retiring at the normal ages (60–66) have the lowest mortality rates. It is interesting that mortality remains higher for those retiring before age 60 for approximately 20 years. After age 80, the curves converge and then diverge, and there is probably too little data beyond this point to draw any significant conclusions based on this type of analysis.

3.5 Comparison of mortality to commonly used standard tables

South African pension fund actuaries and life office pricing actuaries have historically set their mortality assumptions with reference to published UK mortality tables, adjusting as appropriate for the characteristics of the population under consideration and for expected future improvements. Pension fund valuers might use the PA(90) table together with

gender-dependent age ratings and annual improvements in mortality from a fixed date. An example of this could be the use of PA(90) rated down by four years for the male life and two years for the female life, with improvements of 1.0% per annum from calendar year 2000.

Tables 3 and 4 below compare actual to expected mortality, on the basis of several different mortality tables. The period of observation (2000 to 2019) is relatively long and it can be argued that it would be misleading to simply compare the aggregate observed mortality over this whole period to rates taken from a standard table. We have therefore calculated the expected mortality for each calendar year to equal the standard table in year 2000, adjusted by an improvement factor of 1.0% for each year thereafter. The expected mortality rate is therefore $\mu_{x,t}$ where $\mu_{x,t} = \mu_x \times (1-\alpha)^t$, and α is equal to 0.01 (the 0.01 factor reflects a rough approximation for the aggregate observed mortality improvement over the period that is often used in actuarial assumptions in South Africa). The comparison is performed by dividing the observed number of deaths by the expected number of deaths over the period, which are derived by multiplying the exposure at each age by the probability of death implied by the mortality tables (and mortality improvement assumption, α). This calculation is performed for 10 year age bands as well as for all ages together, and for males and females separately. These results are presented in table 3.

TABLE 3 Observed mortality compared to standard table mortality, adjusted by 1% per annum since 2000, for males and females separately, in the period 2000–2019

	a(55)		PA(90)		PFA/PMA ⁶	
	Female	Male	Female	Male	Female	Male
50 to 59	2.5	3.0	3.3	2.7	2.9	5.7
60 to 69	1.6	1.5	1.8	1.4	2.7	3.1
70 to 79	1.2	1.1	1.3	1.1	1.8	1.9
80 to 89	1.0	1.0	1.1	1.1	1.4	1.4
90 to 99	1.2	1.0	1.2	1.1	1.4	1.2
all ages	1.2	1.1	1.3	1.2	1.7	1.7

Aggregate observed mortality is surprisingly heavy when compared to the (adjusted) standard tables, particularly so for the earlier ages, even after considering the effect of possible ill-health retirements. Observed mortality approaches the standard tables only once the possible ill-health retirements are removed and only the top quartile pension income band is considered, as shown in table 4.

6 PFA/PMA refers to the ‘S1’ series, that is S1PFA and S1PMA published in CMI Working Papers 34 and 35, 2008

TABLE 4 Observed mortality for normal health retirements, top quartile pension income, compared to standard table mortality adjusted by 1% per annum since 2000, for males and females separately, in the period 2000–2019. Note that results are not shown for ages 50–59 since these are assumed to be early retirements.

	a(55)		PA(90)		PFA/PMA	
	Female	Male	Female	Male	Female	Male
50 to 59	n/a	n/a	n/a	n/a	n/a	n/a
60 to 69	1.0	0.8	1.1	0.7	1.7	1.5
70 to 79	0.8	0.7	0.9	0.7	1.2	1.2
80 to 89	0.9	0.9	0.9	1.0	1.2	1.3
90 to 99	1.1	1.1	1.2	1.2	1.3	1.3
all ages	0.9	0.9	1.0	0.9	1.2	1.3

3.6 Seasonality in observed data

Observed mortality, and particularly for older population groupings, tends to rise in the winter months and fall off again in the summer months (Bull & Morton, 1975). A similar trend is observed in the dataset, with number of deaths peaking at 120% of the monthly average in July, and dropping to 90% for the summer months of November to February.

3.7 Summary

In this section, an exploratory analysis of the pensioner dataset was performed, mainly using summary statistics such as the ASMR, that reduce the mortality rates observed over a range of ages to a single figure. Based on the ASMRs shown in this section, it appears that some mortality improvements have occurred for the pensioners within this dataset, however, these occurred mainly before 2010, and vary quite significantly by quartile of pension amount received by the pensioner. A shortcoming of an analysis using ASMRs is that changes in mortality by age are obscured. To overcome this shortcoming, mortality improvement rates by age are derived using models in the next two sections.

4. LEE-CARTER MODEL AND SIMPLE NEURAL NETWORKS

In this section, we focus on modelling mortality by gender, age and quartile of the pension payment distribution. This restricted set of predictor variables is chosen so that (hopefully) enough exposure is available at each age for a traditional model, namely the LC model, to be applied to the eight combinations of gender and quartile group separately. The LC model is applied to the aggregated data by gender as well. A neural network is also calibrated to the same data, but in this case, all combinations of the predictor variables are modelled jointly, as described in section 2. Since the predictor variables are relatively simple in this case, the neural network fit is referred to in this section as the simple neural network (“Simple NN”). The Simple NN models were calibrated for 100 training epochs (where an epoch refers to a training run over all of the data).

The modelling is restricted to the age range 65–99 since at ages younger than 65, the mortality rates are too high for healthy early retirements, as also shown in the exploratory analysis in the previous section. Since the LC model is fit directly to the mortality rates estimated from the data, any unreasonably large rates generated when a death occurs with a small exposure are removed. In cases when the estimated mortality rates are zero, the LC model cannot be fit, since the logarithm of zero is infinite. In these cases, the mean mortality for that gender, age and year are imputed from the rest of the data and the observed rate is set equal to the imputed rate (instead of imputing rates for single ages, another solution would be to work with the data in age bands, but this would not allow for the derivation of mortality improvements rates for single ages). Imputation is required for approximately 3.4% of the dataset. In contrast the neural networks directly model the number of deaths (since the output of the network is an estimated mortality rate multiplied by the exposure), thus, there is no need to impute values for observed mortality rates of zero when fitting these models.

The performance of the models is measured by calculating the Mean Squared Error (MSE) of the model predictions against the observed mortality rates, for each subset of the data, quartile group, gender and age separately. The MSE metric is selected since this is consistent with the calibration of the LC model, as discussed in more detail in Richman & Wüthrich (2019), but nonetheless, other error metrics could be considered, for example, the Mean Absolute Percentage Error (MAPE), which is quite commonly used when assessing the quality of forecasts. These results are presented in table 5 and the test set results⁷ are shown in figure 8. The fitted parameters of the LC model are detailed in appendix B. Focusing on the testing data, both the LC and Simple NN models fit the aggregated data better than the data split by pension amount quartile. The LC model performs quite well on the aggregated data, however performance of the LC model is significantly worse on the data split by quartile group. The Simple NN model outperforms the LC model on the out-of-sample data in all of the ten different groups. In some instances, the improved performance is significant, whereas in other cases, the improvements are more marginal. On the training data, the LC model outperforms the Simple NN model, probably because the neural network is not trained to convergence on these data, but rather optimisation is halted once the validation error plateaus, as discussed in section 2. In other words, we do not select the neural network model with the lowest error on the training data, but rather select the model with the lowest error on the validation data.

The out-of-sample predictions for the LC and Simple NN models (that is, in the years 2015–2019) are compared against the observed data in figure 9. Both models appear to fit the data reasonably well, however, whereas the LC model predictions are sometimes quite variable over the age range, the Simple NN model produces predictions that are relatively smooth (except at the highest ages). Nonetheless, the LC model predictions are quite reasonable for the aggregated data, suggesting that not enough data exists in each quartile group for the attempt to fit separate LC models for each quartile group to be successful. In

7 The split of the data into a training and test set is explained in section 2.1.

TABLE 5 Mean squared error of the Lee–Carter and Simple NN models, on the training data (2000–2014) and the testing data (2015–2019), males and females separately, for each of the quartile groups and for all groups together

Quartile group	Gender	LC MSE	Simple NN MSE	Best performance
Test				
1	Female	0.00399	0.00346	Simple NN
1	Male	0.01278	0.00229	Simple NN
2	Female	0.01200	0.00971	Simple NN
2	Male	0.01388	0.00556	Simple NN
3	Female	0.00837	0.00353	Simple NN
3	Male	0.03704	0.01229	Simple NN
4	Female	0.00753	0.00595	Simple NN
4	Male	0.01802	0.01549	Simple NN
all	Female	0.00234	0.00144	Simple NN
all	Male	0.00266	0.00154	Simple NN
Train				
1	Female	0.01484	0.01931	LC
1	Male	0.05207	0.05386	LC
2	Female	0.01667	0.01738	LC
2	Male	0.02979	0.03526	LC
3	Female	0.02563	0.02173	Simple NN
3	Male	0.03970	0.04446	LC
4	Female	0.01184	0.01086	Simple NN
4	Male	0.04092	0.04304	LC
all	Female	0.00289	0.00486	LC
all	Male	0.00597	0.00837	LC

most cases, there is good agreement between the level of the LC and Simple NN predictions. At the highest ages, the Simple NN model produces mortality rates that appear to flatten with age, which should be corrected if graduated life tables were being produced from this analysis.

Mortality improvement rates are shown for the LC and Simple NN models in figure 15 in appendix B. Similar to the comment on mortality rates, the mortality improvement rates implied by the LC model for each of the quartile groups are highly variable with age in a manner that is not demographically reasonable while the mortality improvement rates for the aggregated data appear slightly better. On the other hand, the neural network mortality improvement rates show a far smoother progression with age for all the quartile groups and the aggregated data. Thus, we focus primarily on the mortality improvement rates from the neural network, which we show in figure 10 for the years 2010–2019.

Similar to the observations made in section 3, mortality improvement rates estimated

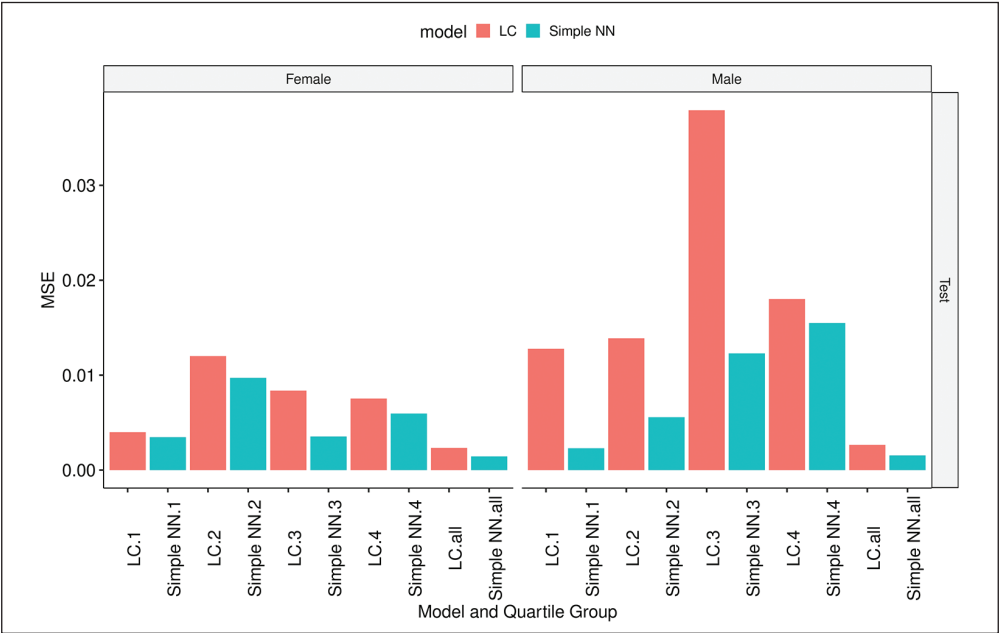


FIGURE 8 Mean squared error of the Lee–Carter and Simple NN models, on the testing data (2015–2019), males and females separately, for each of the quartile groups and for all groups together

from the Simple NN model are generally more significant for males than for females over most of the age range 65–90, after which the improvement rates are of a similar magnitude. According to the model, mortality improvements are quite strongly dependent on age. The usual simplifying assumption made by actuaries of a constant rate of mortality improvement for all ages does not appear appropriate, at least for these data, which show that the highest improvements are at the relatively younger ages in the age range, and that less mortality improvement (or, for females, even mortality worsening) has occurred in the middle of the age range. Using a constant assumption within actuarial cash flow modelling would likely produce cash flow projections with an incorrect timing, however, the significance of the error would depend on the purpose of the modelling exercise. The rate of improvement also depends on the quartile group into which the pension amount falls, with more significant improvements estimated in quartile groups one and two, compared to groups three and four. Finally, the estimated mortality improvement rates themselves are shown to be changing over time, with improvement rates generally becoming lower in magnitude as the years progress. A summary of the estimated mortality improvement rates is provided in the appendix in table 11.

Based on these findings, when setting assumptions for projected mortality improvement, it is advisable for actuaries to consider the following issues:

- whether different improvement rates by sex are appropriate;
- whether an allowance for age-specific mortality improvement rates is appropriate;

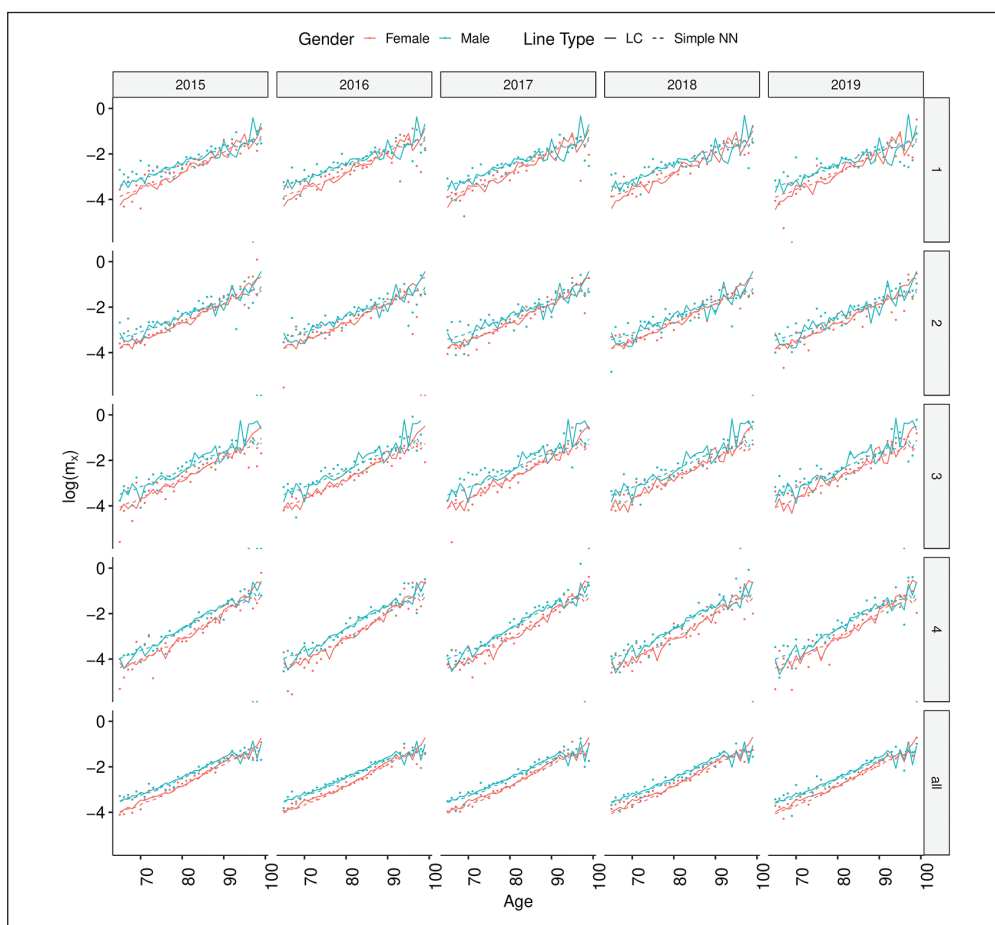


FIGURE 9 Actual (dots) and predicted (lines) mortality rates, 2015–2019, males and females shown in separately, produced by the LC and Simple NN models, for the quartile groups and all data together. Each of the years 2015–2019 are shown in a different panel.

- whether the mortality improvement rate assumptions should vary with time; and
- whether different mortality improvement rates should be set for pensioners with different socio-economic circumstances.

In this section, results of models fit to the data disaggregated by gender, age and quartile of pension amount have been presented. The study shows that each of these factors quite significantly predicts the extent of mortality improvements. However, it still might be the case that the mortality improvements in each group are only apparent due to changes in the underlying composition of the dataset over the years. For example, more early retirements of those in poor health occurring in later years might contribute to the trend of lower mortality improvements with time observed in this section. This issue is addressed next.

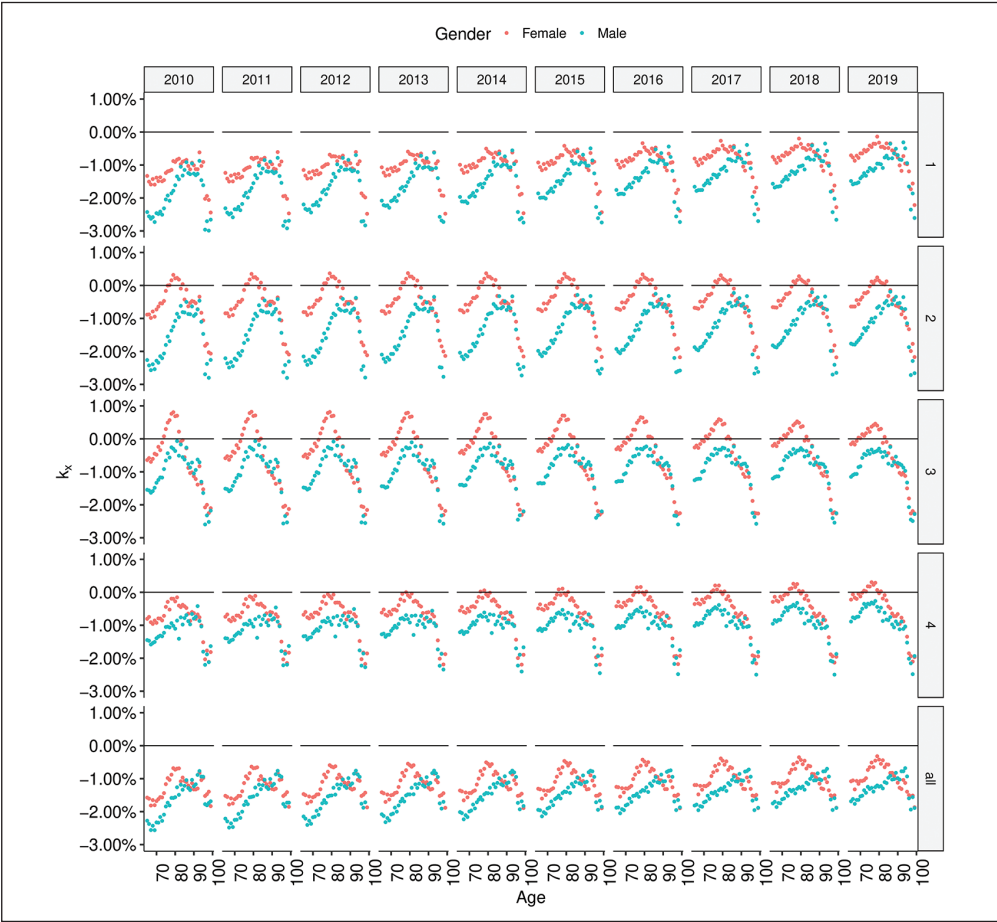


FIGURE 10 Implied mortality improvement rates, 2010–2019, males and females separately, produced by the Simple NN models, for the pension payment quartile groups and all data together. Each of the years 2015–2019 are shown in a different panel.

5. EXPLORING HETEROGENEITY USING A COMPLEX NEURAL NETWORK

To address the issue of possible changes in the composition of the pensioner dataset leading to incorrect inferences about mortality improvement, this section considers how a more complex neural network model (“Complex NN”) was fit to the pensioner dataset, and some of the results from this model are then discussed.

Aside from the variables used in the last section to model mortality rates (gender, age and quartile of pension amount), several other variables are available within the pensioner dataset, some of which were explored in section 3. These include: the age at which retirement occurred, the industry in which the pensioner worked, whether the pensioner has medical aid cover, where the pensioner lives, and whether the principal pensioner is in receipt of the pension, or a contingent life. Seemingly, among the most important of these is whether

retirement occurred early, due to presumed ill health, or not, as shown in figure 7, and these variables should therefore be included in a more complex model. On the other hand, some of these variables did not appear to have a significant impact on mortality, for example, the industry in which the pensioner worked did not appear to influence the ASMRs significantly (this analysis is not shown in the paper). Nonetheless, it is well known that machine and deep learning models can effectively use combination of variables weakly correlated with the outcome to improve predictions (see Efron (2020) for a recent discussion and example of this). Accordingly, we made no attempt to perform much variable selection; rather all of these variables just described are provided to a more complex neural network model and reliance is placed on the techniques applied to regularise the model to select only the most appropriate variables. Two exceptions to this are the place of residence of the pensioner and marital status, which are excluded from the model on the grounds of data limitations. The former variable is missing for a significant proportion of the pensioners and the latter variable is known to be updated infrequently by pension administrators, as it is only captured at retirement and subsequently in the event of the main life pensioner predeceasing his or her spouse.

The structure of the Complex NN model is very similar to the Simple NN model, except that the additional variables just mentioned are added to the model using embedding layers (which were defined in section 2.3.3). The models were calibrated over 50 training epochs. Note that we fit the Complex NN model for all ages up to and including age 99 available in the pensioner dataset, whereas we limited the fit of the Simple NN model to ages 65–99. Thus, we explicitly model the mortality rates of those who retire early. Since the Complex NN model calibrates the age embedding at ages younger than 60, and the model also calibrates embeddings for whether retirement occurs early or not, we are able to extrapolate mortality rates for normal health retirements occurring at ages 50+ using the Complex NN model. The predictive performance of the Complex NN model, compared to the Simple NN model, is shown in table 6 (the results of the LC model are not presented again, since the Simple NN model outperformed the LC model in every case in the testing data). The Complex NN model outperforms the Simple NN model except for the first quartile group of pension amount. The outperformance is sometimes significant, and other times produces results close to the Simple NN model. Thus one can conclude that adding the extra variables (discussed above) to the model has improved out-of-sample forecasting performance, and the model predictions are robust at the level of data aggregated by quartile group, age and sex.

We now investigate whether the same conclusions about mortality improvement that were made in the previous section can also be drawn from the more Complex NN model. One might approach this problem by aggregating the predictions produced by the Complex NN model on the dataset containing the detailed variables to the same level as in the previous section, in other words, to produce predictions split by gender, age and quartile group. However, since changes in the composition of the portfolio will still affect these results, the problem of spurious inferences induced by changing portfolio composition over time will not be solved. Therefore, a different approach is followed: predictions from the Complex NN model are produced for every combination of the variables used to calibrate the model, even

for those combinations that do not appear in the pensioner dataset. These predictions are produced by assigning each combination of the variables within the Complex NN model the same amount of exposure – a single person year – and then making predictions of mortality rates using the Complex NN model. By design, equal weight is then given to each of the variable combinations in the analysis. For example, whereas those retiring early form a smaller proportion of the observed pensioner dataset, in this analysis, equal weight would be given to those retiring early and those retiring at normal retirement ages. The predictions on this hypothetical dataset can then be aggregated to the desired level of granularity, for example, these could be summarised using only the age, gender and quartile group variables, and the resulting inferences will not suffer from the effect of changes in the population composition over time. Since we are primarily interested in mortality rates and mortality improvement rates for pensioners who retire normally, we focus on specific combinations of the variables in this hypothetical dataset in what follows.

TABLE 6 Mean squared error of the Simple and Complex NN models, on the training data (2000–2014) and the testing data (2015–2019), males and females separately, for each of the quartile groups and for all groups together

Quartile group	Gender	Simple NN MSE	Complex NN MSE	Best performance
Test				
1	Female	0.00346	0.00353	Simple NN
1	Male	0.00229	0.00297	Simple NN
2	Female	0.00971	0.00902	Complex NN
2	Male	0.00556	0.00487	Complex NN
3	Female	0.00353	0.00260	Complex NN
3	Male	0.01229	0.01166	Complex NN
4	Female	0.00595	0.00572	Complex NN
4	Male	0.01549	0.01448	Complex NN
all	Female	0.00144	0.00119	Complex NN
all	Male	0.00154	0.00152	Complex NN
Train				
	Female	0.01931	0.02097	Simple NN
1	Male	0.05386	0.05485	Simple NN
2	Female	0.01738	0.01744	Simple NN
2	Male	0.03526	0.03818	Simple NN
3	Female	0.02173	0.02221	Simple NN
3	Male	0.04446	0.04542	Simple NN
4	Female	0.01086	0.01080	Complex NN
4	Male	0.04304	0.04503	Simple NN
all	Female	0.00486	0.00563	Simple NN
all	Male	0.00837	0.01102	Simple NN

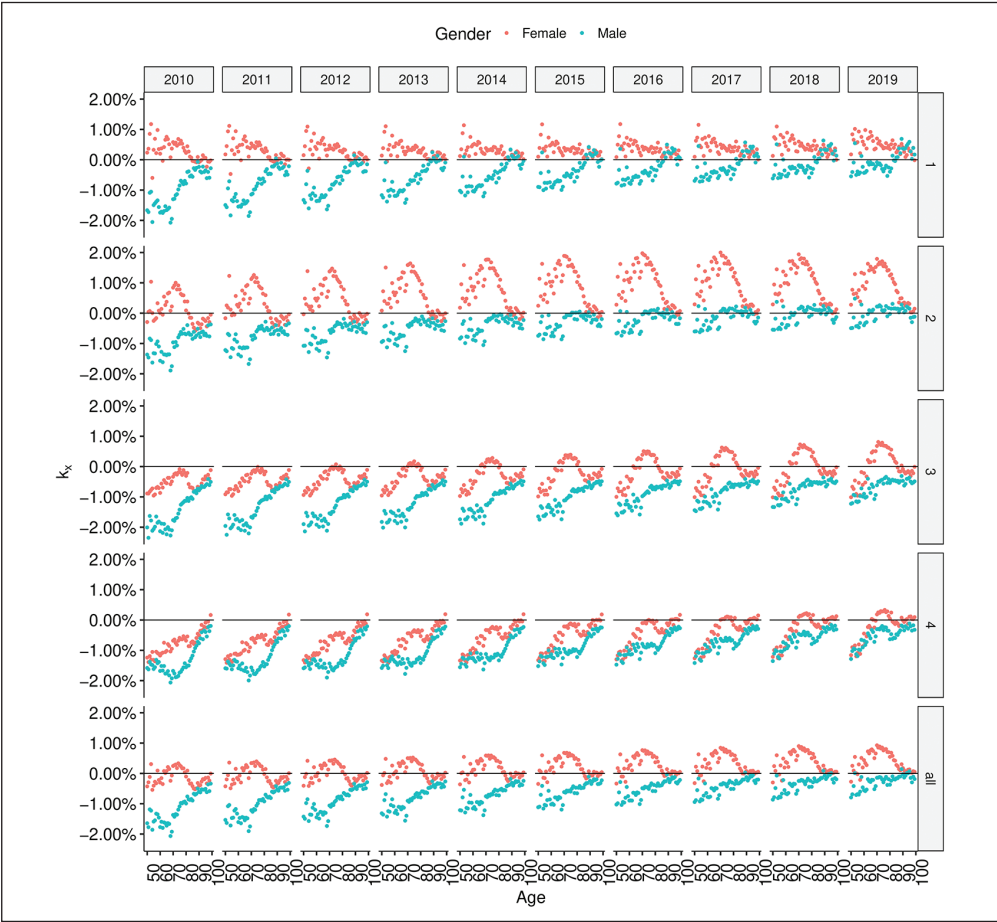


FIGURE 11 Implied mortality improvement rates, 2010–2019, males and females separately, produced by the Complex NN models, for the quartile groups and all data together. Each of the years 2015–2019 are shown in a different panel.

Results focusing on whether those retiring normally (i.e. principal lives retiring at either the normal retirement age or after) have experienced mortality improvements over time are shown in figure 11 and are summarised in table 7.

A slightly different set of conclusions emerge compared to the previous section. First, mortality appears to have worsened in some instances for females once the effects of changing portfolio composition have been controlled for, which is a surprising finding that needs more investigation, but is nonetheless not out of line with the findings at a population level in Richman (2017), who estimated a very low mortality improvement rate of 0.1% per annum for females in the years 1985–2011. The effect of mortality worsening is strongest in the two lower pension amount quartiles, whereas mortality improvements are estimated for the upper pension amount quartiles. For males, it is estimated that mortality improvements

TABLE 7 Mortality improvement rates⁸ estimated using the Complex NN model summarised in age bands spanning 10 years and year bands spanning four years, 2008–2019, males and females, and four quartile groups and aggregated data shown separately

Year band		1	2	3	4	all
Female						
50–59	2008–2011	0.4%	–0.0%	–0.8%	–1.2%	–0.2%
50–59	2012–2015	0.5%	0.6%	–0.8%	–1.2%	–0.0%
50–59	2016–2019	0.6%	0.8%	–0.7%	–1.0%	0.1%
60–69	2008–2011	0.4%	0.2%	–0.6%	–1.0%	–0.1%
60–69	2012–2015	0.4%	1.0%	–0.3%	–0.8%	0.2%
60–69	2016–2019	0.5%	1.4%	0.0%	–0.4%	0.5%
70–79	2008–2011	0.5%	0.6%	–0.2%	–0.7%	0.2%
70–79	2012–2015	0.4%	1.4%	0.1%	–0.4%	0.5%
70–79	2016–2019	0.4%	1.7%	0.6%	0.1%	0.8%
80–89	2008–2011	0.1%	–0.3%	–0.6%	–0.7%	–0.3%
80–89	2012–2015	0.2%	0.4%	–0.4%	–0.5%	–0.0%
80–89	2016–2019	0.3%	0.8%	0.0%	–0.2%	0.3%
90–99	2008–2011	–0.1%	–0.3%	–0.4%	–0.2%	–0.3%
90–99	2012–2015	0.1%	–0.0%	–0.3%	–0.1%	–0.1%
90–99	2016–2019	0.2%	0.2%	–0.2%	0.0%	0.1%
Male						
50–59	2008–2011	–1.6%	–1.3%	–1.9%	–1.5%	–1.6%
50–59	2012–2015	–0.9%	–0.7%	–1.6%	–1.4%	–1.1%
50–59	2016–2019	–0.4%	–0.3%	–1.2%	–1.1%	–0.6%
60–69	2008–2011	–1.8%	–1.6%	–2.1%	–1.8%	–1.8%
60–69	2012–2015	–1.1%	–0.8%	–1.6%	–1.4%	–1.2%
60–69	2016–2019	–0.4%	–0.3%	–1.0%	–0.8%	–0.6%
70–79	2008–2011	–1.0%	–0.9%	–1.5%	–1.7%	–1.2%
70–79	2012–2015	–0.6%	–0.2%	–1.1%	–1.3%	–0.7%
70–79	2016–2019	–0.3%	0.1%	–0.6%	–0.6%	–0.3%
80–89	2008–2011	–0.5%	–0.7%	–1.0%	–1.2%	–0.8%
80–89	2012–2015	–0.2%	–0.2%	–0.8%	–1.0%	–0.5%
80–89	2016–2019	0.0%	0.1%	–0.6%	–0.6%	–0.2%
90–99	2008–2011	–0.4%	–0.7%	–0.6%	–0.5%	–0.5%
90–99	2012–2015	–0.0%	–0.4%	–0.6%	–0.4%	–0.3%
90–99	2016–2019	0.3%	–0.1%	–0.5%	–0.3%	–0.2%

8 Note that negative percentages indicate an improvement (reduction) in mortality; positive percentages indicate a deterioration.

have occurred in all pension amount quartiles and at most ages, with larger improvements in the higher pension amount quartiles. Finally, it can be seen that there is a trend with time of decreasing mortality improvements across both sexes, at most ages and in all quartile groups.

Many other combinations of the variables could be explored, but to avoid an excessively long paper, these analyses are not reported here.

We conclude this section with a high level analysis of the effect of the year variable on mortality in the Complex NN model; in other words, we look to analyse what overall trend in mortality with time the model has estimated. A well-known method for interpreting “black-box” machine learning models is the partial dependence plot (PDP) of Friedman (2001), which is the marginal expectation of the effect of a variable on the output of a machine learning model, where all the other variables within the model are marginalised out. Zhao & Hastie (2019) show that the effect sizes in a PDP have a causal interpretation in some circumstances, which is a significant advantage of the PDP method compared to other techniques for interpreting machine learning models. Here, the PDP for each year is determined by calculating the average mortality rate for each year of the analysis, where the average is taken for every combination of the variables used in the Complex NN model except for year. These average mortality rates for each year are then divided by the average mortality rate in the year 2000, thus, the PDP shows how mortality rates have improved or worsened relative to mortality rates in the year 2000. The PDP plot is shown in figure 12, where it can be seen that the trend in mortality rates according to the Complex NN model follows a parabolic shape, falling until about 2012 and starting to rise from about 2013 to 2019. While this is an aggregate effect within the model, it has been shown, for example in table 7, that in recent years mortality improvements are smaller or mortality is worsening compared to earlier years. It can thus be concluded that continuing mortality improvement is not guaranteed in South Africa and, besides for building more quantitative understanding, the medical and social drivers of this should be investigated in more detail.

6. APPLICATION TO PENSIONS RESERVING AND INSURANCE SOLVENCY ASSESSMENT

In this section we discuss the effect of applying the mortality improvement rates in table 7 to pensions and insurance reserving.

6.1 Pension fund valuations

The Pension Funds Act (Republic of South Africa, 1956) requires defined benefit pension funds to undertake an actuarial valuation at least once every three years. The Financial Sector Conduct Authority is responsible for issuing directives and guidance, which, together with the standards of actuarial practice issued by the Actuarial Society, set the parameters within which the valuator performing the actuarial valuation works. This legal, regulatory and professional framework affords the valuator considerable latitude in exercising his or her professional judgement, particularly as it relates to setting the mortality basis and making allowance for future improvements. Although not strictly required by

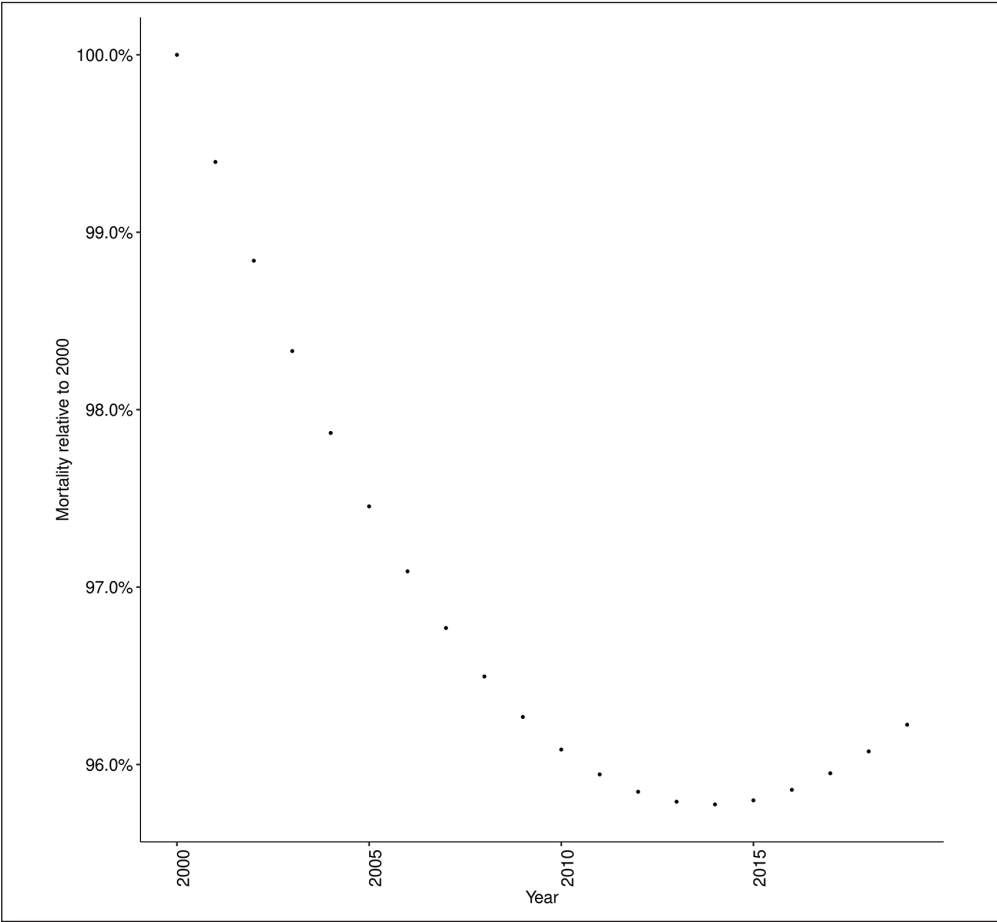


FIGURE 12 Partial dependence plot of the effect of year on mortality rates, comparing average mortality rates in each year to mortality rates in the year 2000

the framework, it has become standard practice for two sets of valuation calculations to be undertaken: a best estimate valuation and a solvency valuation. The latter will include a margin for the deterioration of future experience (in other words, for a scenario in which the valuation results are larger), including improvements in pensioner mortality over and above that assumed in the best estimate basis. As an example, the house-view actuarial basis of a large South African actuarial consultancy sets mortality improvements to be in the range of 0.5% to 1.0% per annum for best estimate bases, with the (more prudent) solvency bases between 1.0% to 2.0% per annum.

Pension fund liabilities are of course forward looking and long term in nature. The historical experience used to derive the mortality improvement rates in table 7 does not necessarily reflect what is expected to unfold over the entire period corresponding to the term of the liabilities. It is nevertheless instructive to observe what effect differing levels

of mortality improvement might have on pension fund liabilities, across gender, age bands and pension amount. Table 8 shows the reduction in the best estimate liabilities for a large defined benefit pension fund (whose best estimate basis assumes improvements of 1% per annum), by applying the mortality improvement rates in table 7 for the most recent period (2016–2019, for which it is noted above that the rates of improvement began to deteriorate). For the purpose of these calculations, we set the mortality improvement rate to zero where the Complex NN model indicated negative improvements over the period.

As is expected, the reduction in liabilities is greater for female lives and for the lower pension amounts, and in some cases, double-digit reductions in liabilities have been estimated.

TABLE 8 Reduction in pension fund best estimate liabilities, replacing the fund's provision for mortality improvements with rates estimated using the Complex NN model

Pension band	1	2	3	4	all
Female					
60–69	–10%	–10%	–10%	–6%	–8%
70–79	–11%	–11%	–11%	–11%	–11%
80–89	–11%	–11%	–11%	–9%	–10%
90–99	–8%	–7%	–4%	–6%	–5%
Male					
60–69	–5%	–7%	0%	–2%	–2%
70–79	–7%	–10%	–4%	–5%	–5%
80–89	–11%	–11%	–5%	–4%	–5%
90–99	–9%	–8%	–5%	–7%	–7%

6.2 Insurance solvency and accounting regulations

Modern insurance solvency regulations (SAM, in South Africa) and incoming accounting regulations (IFRS 17) require that the uncertainty around best estimate assumptions be quantified, for the calculation of capital requirements and risk margins on reserves. In SAM, capital is required to be held using the Value-at-Risk risk measure at the 99.5% percentile of the loss distribution and, in the case of mortality and longevity risks, the regulator (Prudential Authority, 2018a) has prescribed that these percentiles are a relative 15% increase and 10% decrease in best estimate assumptions mortality rates, respectively. In addition, for longevity risk an absolute increase in the mortality improvement assumption of 1% per annum over the best estimate assumption has been prescribed. For longevity risk, these assumptions were set “in the absence of any detailed information to calibrate the shock” (FSB, 2014a). Theoretically, insurers should test these assumptions against their own experience in the Own Risk and Solvency Analysis (ORSA) (Prudential Authority, 2018b).

On the other hand, the incoming IFRS 17 accounting standards require that, when accounting for the cashflows produced by insurance contracts, a so-called “risk adjustment”

be added to the best estimate cashflows “to reflect the compensation that the entity requires for bearing the uncertainty about the amount and timing of the cash flows that arises from non-financial risk” (IASB, 2020). Thus, insurers will need to calibrate risk adjustments to comply with the standard.

Here, the pensioner dataset is used to provide a simplified⁹ empirical view on the question of what these assumptions might be in South Africa. Since a practical manner of calculating uncertainty for neural network predictions is an open question (most neural network models require too much computation to apply bootstrapping), the output of a LC model calibrated to the combined experience for males and females is rather considered in the years 2000–2014 on the restricted age range 65–89. Although the LC model did not perform well on the data disaggregated by quartile group of the pension amount, on the aggregated data, the performance was not dramatically worse than the neural networks. To produce the results that follow, the LC model was calibrated as described in section 2 and the k_t parameters were forecast using a random walk with drift. The 75th, 90th, 95th and 99.5th percentiles (and the corresponding lower percentiles) of the stochastic simulations of the random walk were taken to produce eight stressed projections of k_t for the period 2015–2019 and corresponding mortality rates for the ages 65–89 in these years. Although SAM only requires stresses at the 99.5th percentile, this is not specified in IFRS 17, thus the lower percentiles of the loss distribution are also considered. Figure 13 shows these simulated values of k_t . Due to the slight trend in the mean value of k_t , the distribution of stresses is not symmetrical.

The first discussion relates to a single year projection of k_t (that is, for the year 2015) which was performed to align with the one year time horizon of SAM. Table 9 shows that the 99.5th stress on the average mortality rate in this age range is 9.4% above the mean in a mortality risk scenario and about 8.2% below the mean in a longevity risk scenario. Comparing these figures to the SAM prescribed mortality risk shock of 15%, it seems that, at face value, SAM is more conservative than these data indicate, whereas the 10% prescribed shock for longevity risk seems reasonable.

Nonetheless, it should be noted that the SAM prescribed shocks theoretically should cover the risks arising from the uncertain level (that is, the estimation error, which we refer to, in what follows, as “level” risk in line with the terminology in SAM) and trend in mortality rates, as well as volatility risk (which refers to the random variation in observed deaths arising from sampling error, see Prudential Authority (2018a) for details) whereas, the calculations here are based only on the trend risk over a single year and assume that the mortality rates which have been projected are free of estimation error. We do not attempt to estimate the level risk here, but provide estimates of volatility risk for illustrative purposes. These are calculated by performing a separate set of simulations based on the projected mortality

⁹ The stress results are only considered for both genders grouped together and a single stress for all ages is calculated, in line with the specification in SAM. Best practice would require that parameter error also is estimated, but we only consider the potential time series forecasting error here.

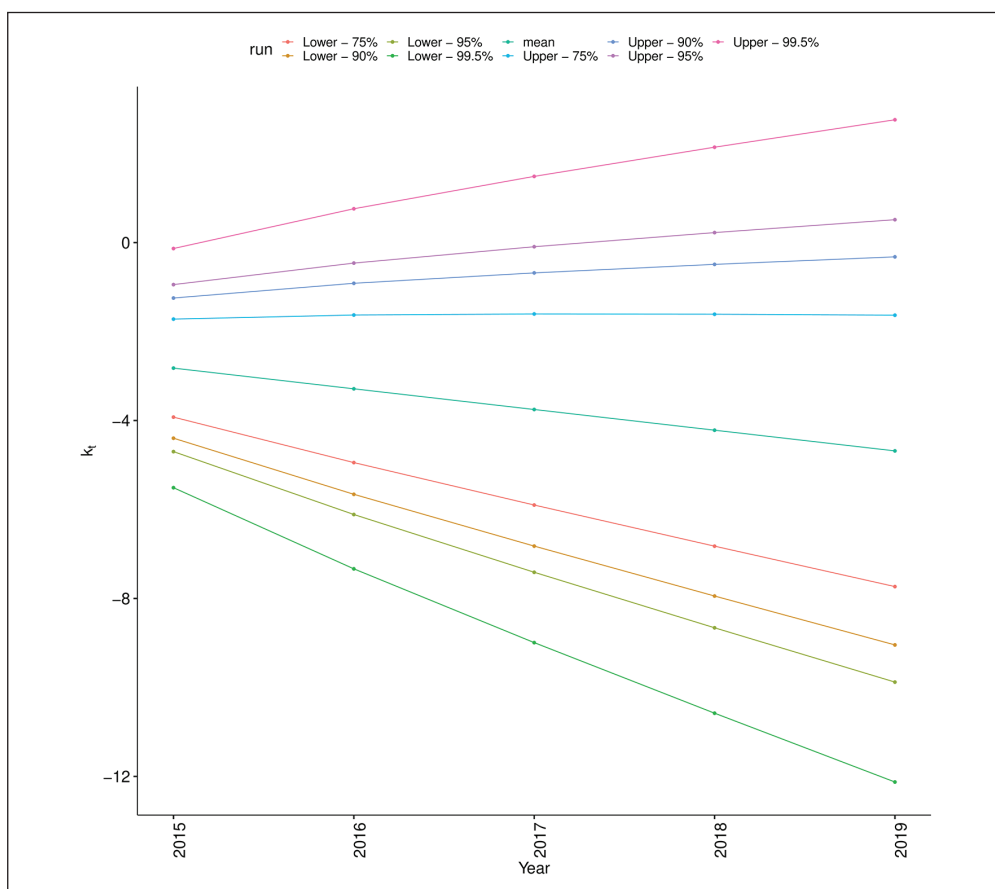


FIGURE 13 k_t in the years 2015–2019 based on a LC model fit to both sexes together, mean projection and 75th, 90th, 95th and 99.5th percentiles (and the corresponding lower percentiles)

TABLE 9 Mortality rate stresses calibrated for both sexes jointly, single year projection, mean and 75th, 90th, 95th and 99.5th percentiles

	Mortality rate	Stressed mortality rate	Relative stress	Absolute stress
Lower – 75%	0.065	0.063	96.5%	–0.002
Lower – 90%	0.065	0.062	95.1%	–0.003
Lower – 95%	0.065	0.061	94.2%	–0.004
Lower – 99.5%	0.065	0.060	91.8%	–0.005
Upper – 75%	0.065	0.067	103.7%	0.002
Upper – 90%	0.065	0.068	105.3%	0.003
Upper – 95%	0.065	0.069	106.3%	0.004
Upper – 99.5%	0.065	0.071	109.3%	0.006

rates in 2015. For each simulation, it is assumed that deaths at each age x are distributed according to a Poisson distribution, with the mean parameter of the distribution at each age (that is, λ_x) set equal to the expected deaths based on the LC projections for 2015. At the 99.5th percentile, the volatility risk is estimated to be around 5%. Thus, since our combined estimates of the stress for the trend and volatility components of the shock are around 14% and since the level component of mortality risk is also usually significant, it appears that the prescribed shock in SAM for longevity risk is significantly too low and that the prescribed shock for mortality risk is also too low.

Focusing now on mortality improvement rates, table 10 shows five year projections for mortality improvement rates in the same age range of 65–89. Here, the simulation results show that the prescribed SAM stress of an absolute increase of 1% over the best estimate mortality improvement rate is wholly inadequate, and would be inadequate even at the 75th percentile, at which point of the distribution it appears that an absolute increase of 1.8% is required. Also, whereas SAM does not require insurers to consider that mortality might worsen over time, the figures in table 10 show that adverse trends in mortality are significant. These results indicate that the SAM calibration for mortality and longevity risk should be subjected to a significantly more robust investigation, and that companies performing an ORSA should not rely on the SAM shocks without justification why these are believed to be appropriate.

TABLE 10 Mortality improvement rate stresses calibrated for both sexes jointly, five year projection, mean and 75th, 90th, 95th and 99.5th percentiles

	Improvement rate	Stressed improvement rate	Relative stress	Absolute stress
Lower – 75%	–0.019	–0.036	195.3%	–0.018
Lower – 90%	–0.019	–0.044	236.3%	–0.025
Lower – 95%	–0.019	–0.049	262.4%	–0.030
Lower – 99.5%	–0.019	–0.062	332.6%	–0.043
Upper – 75%	–0.019	–0.001	4.7%	0.018
Upper – 90%	–0.019	0.007	–36.3%	0.025
Upper – 95%	–0.019	0.012	–62.4%	0.030
Upper – 99.5%	–0.019	0.025	–132.6%	0.043

Unlike SAM, IFRS 17 does not specify exactly how the risk adjustment should be calculated, nor does it specify at what confidence level these should be estimated. For a detailed consideration of these points in the context of non-life insurance, the reader is referred to England et al. (2019). Here, the assumption is that the Value-at-Risk risk measure has been selected by the company to calculate the risk adjustment, and we provide examples of what assumptions might be appropriate at lower percentiles of the distributions than SAM, being the 75th, 90th and 95th percentiles of the mortality and mortality improvement distributions.

The main observation is that even at the 75th percentile of the mortality improvement distribution, a significant allowance for mortality worsening over time has been estimated, meaning to say that even if the best estimate mortality improvement assumption has been set at 0% per annum, a risk adjusted assumption at the 75th percentile would require mortality to worsen at 1.8% per annum. Thus, when setting assumptions for IFRS 17, insurers should consider appropriate trend assumptions that include the possibility of mortality worsening, and not rely on the SAM prescribed shocks, which do not consider this situation.

7. DISCUSSION AND CONCLUSION

In this study, a pensioner dataset has been used to provide empirical information on mortality improvements at post-retirement ages in South Africa. Whereas observed mortality improvements might be caused by changing proportions of lives with different mortality experience within a portfolio, this dataset contains enough information about each pensioner to control for this effect. This allows for the estimation of whether mortality has been improving due to the effect of factors that do not depend on the composition of the portfolio. The key findings are based on the application of a complex deep neural network model of mortality, which uses most of the variables within the dataset to model and forecast mortality rates, and which displays excellent out-of-sample performance, compared to the LC model. This model is used to predict mortality rates for a hypothetical portfolio that was constructed by allocating an equal number of pensioners to each sub-population defined by the variables used within this model, for example, the mortality of a constant number of male pensioners without access to medical aid is modelled in each year. In this way, we are able to ensure that the composition of the portfolio is held constant over time, and changes in mortality with time result only from the effect of external factors on mortality rates.

In addition to the neural network model, two simpler models were also utilised: the LC model fit to each pension amount quartile group separately, and a simple neural network fit to all quartile groups at once. Comparing the estimated mortality improvement rates from the simpler neural network, which indicate more significant rates of improvement, with the more complex model, which indicates less significant improvements and even significant worsening for some groups, the finding is that mistaken inferences about mortality improvement rates can easily be made if changes in portfolio composition are not allowed for.

The key findings in this study, summarised in table 7, indicate that within the two lower pension amount quartile groups, mortality for females is worsening on an annual basis, and that in recent years, female mortality in the two higher quartile groups has also started to worsen. Male mortality appears to be improving at most ages, but in recent years, has started to worsen at some ages in the two lower quartile groups. Overall, the model shows that mortality improvements are slowing down in South Africa for both genders and at all ages. There is also evidence of more significant mortality improvements in the higher pension amount quartile groups, which is similar to findings in other countries, see, for example, Lu et al. (2012) who draw similar conclusions based on data from the United Kingdom. Another finding is that mortality improvement rates are quite strongly dependent on age, and the

usual simplifying assumption within actuarial models of constant mortality improvement with age likely needs further consideration.

In this study, there is some quantitative evidence that mortality is worsening for females in recent years, and that mortality improvements for males are slowing down. Further research is needed to determine whether this is also the case in other insured populations, and whether similar trends in population mortality are occurring. This study does not attempt to provide reasons for these observations, and future research should aim to investigate what is causing these trends, and how they can be reversed.

The results indicate that when setting mortality improvement assumptions, actuaries should carefully consider the socio-economic circumstances of the lives comprising the portfolio under consideration, and that a simple assumption of mortality improvements for both genders and all ages likely does not have strong empirical support. Furthermore, there is a significant potential for mortality to worsen over time, which should be considered when setting assumptions.

This study has used both traditional methods and models, as well as more complex models that are recently being applied to problems of mortality estimation and forecasting. There is evidence that the LC model performs relatively well on aggregated data as measured by the MSE of the predictions on out-of-sample data and for these data the predictions made by the LC model are relatively smooth with age. However, once the data are disaggregated, and the LC model is fit for each quartile group and gender separately, the performance of the LC model degrades and predictions no longer appear reasonable. It might be the case that smoothed variations of the LC model, for example, those presented in Currie (2013), would perform better than the original LC model considered here. Since producing forecast uncertainty using the LC model is quite well studied, we have relied on this model to produce estimates of uncertainty that were compared to the SAM calibration.

The deep neural network models presented in this study both outperform the LC model as measured on the MSE of out-of-sample predictions and produce more demographically reasonable forecasts. Due to the flexibility of these models, it was possible for most of the variables in the pensioner dataset to be utilised, and for the changing composition of the portfolio to be controlled for when investigating mortality improvements due to factors external to the portfolio. On the other hand, more study of estimating uncertainty around the predictions of deep neural networks is required.

There is evidence that, at least based on the pensioner dataset, the calibration of the mortality and longevity stresses in SAM need to be revisited, since these appear to be too low. This recalibration should ideally be performed at an industry level, to pool as much data as possible. A similar effort could also be undertaken to provide an industry view of risk adjustments for IFRS 17.

Finally, this manuscript is being written as the effect of the COVID-19 pandemic on mortality in South Africa begins to be quantified (Bradshaw et al., 2020), which appears to be especially significant at the older ages. It is likely that the models used in this study would perform quite poorly if they were calibrated on data containing a significant increase in

mortality rates; thus, future research on mortality improvements in South Africa will need to consider the shorter and longer term impacts of COVID-19 and new quantitative methods will likely be required to fit the data from 2020 onwards.

ACKNOWLEDGEMENT

We are very grateful to Professor Rob Dorrington for alerting us to the problem of making incorrect inferences about mortality improvements in the case of heterogeneity within the population under investigation, and for his insightful review comments which helped improve the manuscript substantially. We also express our sincere gratitude to Lucia Viegas for her careful review of the manuscript which led to a substantially improved manuscript. We wish to thank the management and staff of the pensions administrator for providing us with the dataset used in this study.

REFERENCES

- Allaire, JJ & Chollet, F (2017). R Interface to 'Keras'. RStudio, Google.
- Bengio, Y, Courville, A & Vincent, P (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828
- Bengio, Y, Ducharme, R, Vincent, P & Jauvin, C (2003). A Neural Probabilistic Language Model. *Journal of Machine Learning Research*, 3(6), 1137–1155
- Bradshaw, D, Laubscher, R, Dorrington, R, Groenewald, P & Moultrie, T (2020). Report on Weekly Deaths in South Africa – 1 January–14 July 2020 (Week 28) (Tech. Rep.) : South African Medical Research Council and UCT Centre for Actuarial Research.
- Breiman, L (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140
- Bull, GM & Morton, J (1975). Relationships of temperature with death rates from all causes and from certain respiratory and arteriosclerotic diseases in different age groups. *Age and Ageing*, 4(4), 232–246
- CMI (2020). Mortality improvements and CMI_2019: frequently asked questions (FAQs) — Institute and Faculty of Actuaries
- CSI (2012). Annuitant Mortality 2001–2004 (Tech. Rep.). Cape Town: Continuous Statistical Investigation Committee, Actuarial Society of South Africa
- CSI (2017). Report on Pensioner Mortality 2005–2010 (Tech. Rep.). Cape Town: Continuous Statistical Investigation Committee, Actuarial Society of South Africa
- Currie, ID (2013). Smoothing constrained generalized linear models with an application to the Lee–Carter model. *Statistical Modelling*, 13(1), 69–93
- Currie, ID (2016). On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal*, 2016(4), 356–383
- De Brébisson, A, Simon, É, Auvolet, A, Vincent, P & Bengio, Y (2015). Artificial neural networks applied to taxi destination prediction. *arXiv*, *arXiv:1508*.
- Dorrington, RE & Tootla, S (2007). South African annuitant standard mortality tables 1996–2000 (SAIML98 and SAIFL98). *South African Actuarial Journal*, 7, 161–184
- Efron, B (2020). Prediction, Estimation, and Attribution. *Journal of the American Statistical Association*, 115(530), 636–655
- England, P, Verrall, RJ & Wüthrich, MV (2019). On the lifetime and one-year views of reserve risk, with application to IFRS 17 and Solvency II risk margins. *Insurance: Mathematics and Economics*, 85, 74–88
- Friedman, J (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232
- FSB(2014a). Position Paper 641 (v 4) Life SCR – Longevity Risk (Tech. Rep. No. February 2011). Pretoria: Financial Services Board
- FSB (2014b). Position Paper 64 – Life SCR Longevity Risk (Tech. Rep.). Pretoria: Financial Services Board
- Goodfellow, I, Bengio, Y & Courville, A (2016). *Deep Learning*. Cambridge, MA: MIT Press
- Guo, C & Berkahn, F (2016). Entity Embeddings of Categorical Variables. *arXiv*, *arXiv:1604*
- IASB (2020). IFRS – IFRS 17 Insurance Contracts. International Accounting Standards Board

- Ioffe, S & Szegedy, C (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. 32nd International Conference on Machine Learning, ICML 2015, 1, 448–456
- Kingma, DP & Ba, J (2014). Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*
- Kleinow, T (2015). A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics*, 63, 147–152
- Kuo, K (2019). Deeptriangle: A deep learning approach to loss reserving. *Risks*, 7(3)
- LeCun, Y, Bengio, Y & Hinton, G (2015). Deep Learning. *Nature*, 521(7553), 436
- Lee, RD & Carter, LR (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87(419), 659–671
- Li, N & Lee, R (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee–Carter method. *Demography*, 42(3), 575–594
- Lu, JLC, Wong, W & Bajekal, M (2012). Mortality improvement by socio-economic circumstances in England (1982 to 2006). *British Actuarial Journal*, 1–35
- Luo, P, Wang, X, Shao, W & Peng, Z (2019). Towards understanding regularization in batch normalization. In 7th International conference on learning representations, ICLR 2019. International Conference on Learning Representations, ICLR
- Moultrie, TA & Dorrington, RE (2012). Used for ill; used for good: a century of collecting data on race in South Africa. *Ethnic and Racial Studies*, 35(8), 1447–1465
- Perla, F, Richman, R, Scognamiglio, S & Wüthrich, MV (2020). Time-Series Forecasting of Mortality Rates using Deep Learning. *SSRN Electronic Journal*
- Prudential Authority (2018a). Prudential Standard FSI 4.2 Life Underwriting Risk Capital Requirement
- Prudential Authority (2018b). Prudential Standard GOI 3 Risk Management and Internal Controls for Insurers
- Republic of South Africa (1956). Pension Funds Act 24 of 1956
- Richman, R (2017). *Old age mortality in South Africa*. Cape Town: University of Cape Town.
- Richman, R (2018, Oct). AI in Actuarial Science. *SSRN Electronic Journal*
- Richman, R & Wüthrich, MV (2019). A neural network extension of the Lee–Carter model to multiple populations. *Annals of Actuarial Science*
- Richman, R & Wüthrich, MV (2020). Nagging Predictors. *SSRN Electronic Journal*, 1–25
- Srivastava, N, Hinton, G, Krizhevsky, A, Sutskever, I & Salakhutdinov, R (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1), 1929–1958
- Wüthrich, MV (2019). From Generalized Linear Models to Neural Networks, and Back. *SSRN Electronic Journal*
- Wüthrich, MV & Buser, C (2018). *Data Analytics for Non-Life Insurance Pricing* (Vol. 2018) (No. 17 June). Swiss Finance Institute Research Paper
- Zhao, Q & Hastie, T (2019). Causal Interpretations of Black-Box Models. *Journal of Business and Economic Statistics*

APPENDIX A

Keras Code

We briefly describe the Keras code used to fit the Simple NN model in this paper in Listing 1. Lines 3–11 are a for-loop that defines an input layer for each categorical variable, which has been recoded as an integer. Then, the input layer is fed into an embedding layer of dimension 3. Both the input and embedding layers are stored in a list object. Other inputs to the network (the year in which forecasts are to be made and the exposure) are fed into input layers on Lines 16–17. Dropout and batch normalisation are applied to the embedding layers on Lines 19–20. Lines 24–30 define some of the intermediate layers of the network and Lines 32–33 define a skip connection between the inputs of the network and the intermediate layers. On Line 37, the network transforms the last layer into a mortality rate prediction, using a sigmoid activation and predicted death counts are derived on Line 39 by multiplying the estimated mortality rate with the exposure for that group. The Poisson loss function is selected on Line 45 and the code to run the model fitting process appears on Lines 49–50.

Listing 1: Keras code to fit the simple neural network to the pensioner data

```

1  j = 0
2
3  for(i in categorical_vars){
4    j = j+1
5    name = paste0(i, "_input")
6
7    input_layers[[j]] = layer_input(shape = 1, name = i)
8
9    embed_layers[[j]] = input_layers[[j]] %>%
10      layer_embedding(input_dim = dims + 1, output_dim = 3, name = paste0(i, "embed"))
11  }
12
13
14  ### rest of model
15
16  input_layers[[j + 1]] = layer_input(shape = 1, name = "num_inputs")
17  input_layers[[j + 2]] = layer_input(shape = c(1), name = 'exposure')
18
19  features = embed_layers %>% layer_concatenate() %>%
20    layer_reshape(c(j * 3)) %>% layer_batch_normalization %>% layer_dropout(rate = 0.04)
21
22  features = list(features, input_layers[[j + 1]] ) %>% layer_concatenate()
```

```

23
24 middle = features %>%
25   layer_dense(units = 32, activation = "tanh") %>%
26   layer_batch_normalization() %>%
27   layer_dropout(rate = 0.01) %>%
28   layer_dense(units = 32, activation = "tanh") %>%
29   layer_batch_normalization() %>%
30   layer_dropout(rate = 0.01)
31
32 middle = list(middle, features ) %>%
33   layer_concatenate() %>%
34   layer_dense(units = 32, activation = "tanh") %>%
35   layer_batch_normalization() %>%
36   layer_dropout(rate = 0.01)%>%
37   layer_dense(units = 1 , activation = "sigmoid ")
38
39 output = list(middle, input_layers[[j + 2]]) %>% layer_multiply(name = 'output')
40
41 model = keras_model(inputs = c(input_layers), outputs = c(output))
42
43 adam = optimizer_adam(lr = 0.001)
44
45 model %>% keras::compile(optimizer = adam , loss = "poisson")
46
47 learn_rate = callback_reduce_lr_on_plateau(factor = 0.9,patience = 5,cooldown = 0, verbose = 1)
48
49 fit = fit(model, x = inputs, y = output, batch_size = 64, epochs = 100, callbacks = list(learn_rate),
50   validation_split = 0.05, verbose = 1)

```

APPENDIX B

Extra Results from Section 4

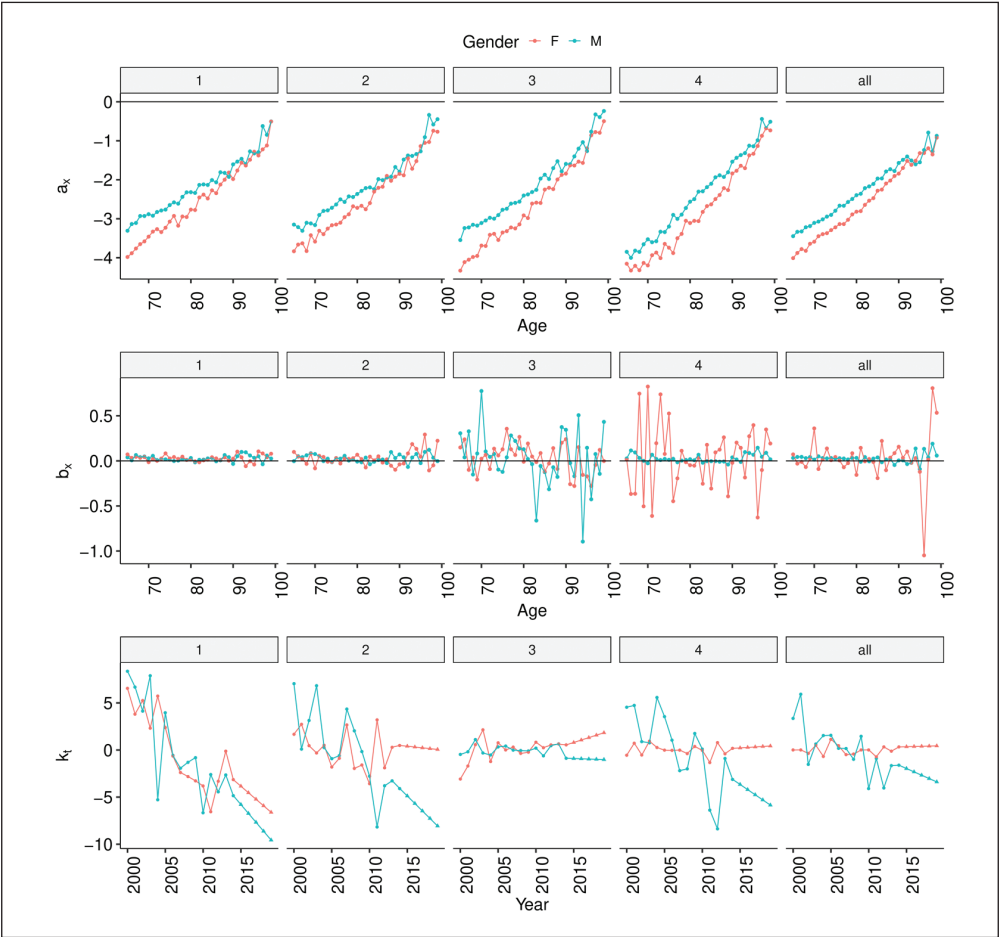


FIGURE 14 Parameters a_x , b_x and k_t of the LC model, fit on the training data (2000–2014) and extrapolated to the years 2015–2019, males and females shown separately, and quartile groups and overall data shown separately

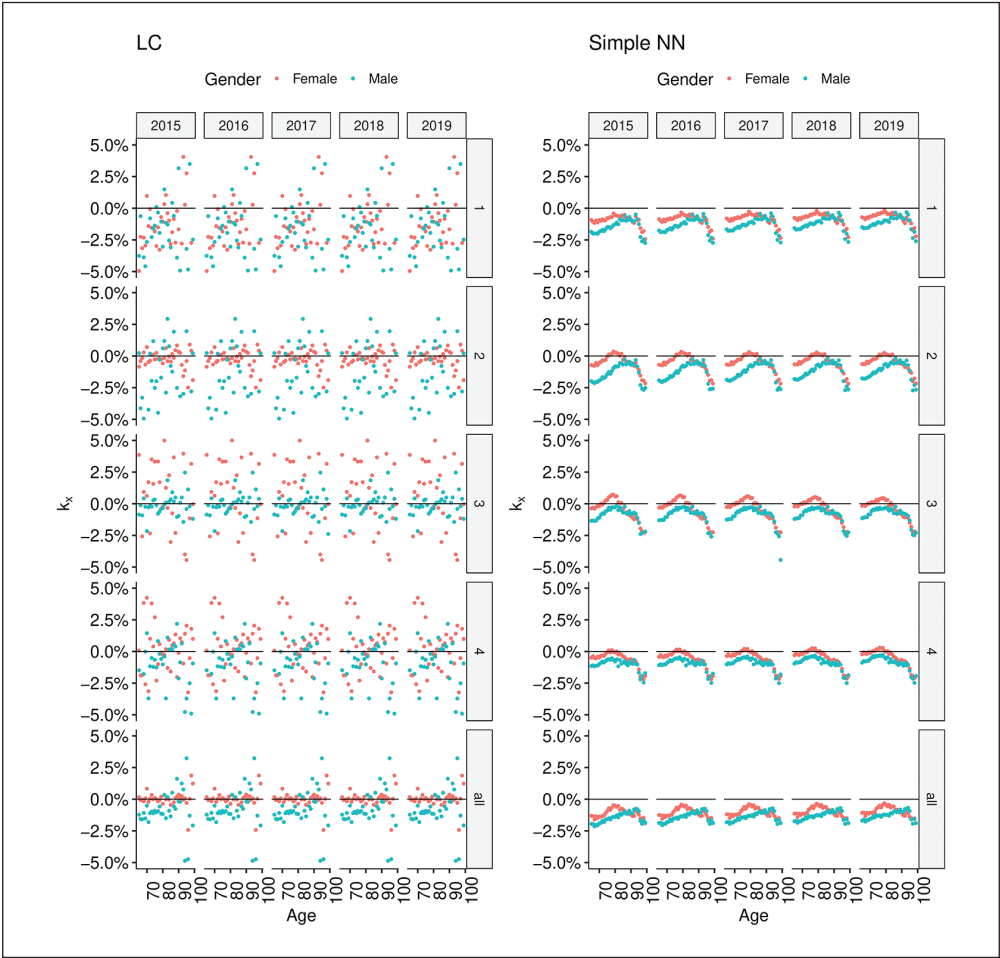


FIGURE 15 Implied mortality improvement rates, 2015–2019, males and females separately, produced by the LC and Simple NN models, for the quartile groups and all data together

TABLE 11 Mortality improvement rates estimated using the Simple NN model summarised in age bands spanning 10 years and year bands spanning four years, 2009–2019, males and females, and four quartile groups and aggregated data shown separately

	Year band	1	2	3	4	all
Female						
60–69	2008–2011	–1.5%	–0.9%	–0.6%	–0.9%	–1.7%
60–69	2012–2015	–1.2%	–0.8%	–0.4%	–0.6%	–1.5%
60–69	2016–2019	–0.9%	–0.6%	–0.2%	–0.3%	–1.2%
70–79	2008–2011	–1.3%	–0.3%	0.2%	–0.6%	–1.3%
70–79	2012–2015	–0.9%	–0.2%	0.3%	–0.3%	–1.1%
70–79	2016–2019	–0.6%	–0.1%	0.3%	0.0%	–0.8%
80–89	2008–2011	–1.0%	–0.2%	–0.2%	–0.5%	–1.0%
80–89	2012–2015	–0.7%	–0.1%	–0.2%	–0.4%	–0.9%
80–89	2016–2019	–0.5%	–0.1%	–0.2%	–0.4%	–0.7%
90–99	2008–2011	–1.4%	–1.2%	–1.6%	–1.3%	–1.3%
90–99	2012–2015	–1.4%	–1.2%	–1.6%	–1.3%	–1.3%
90–99	2016–2019	–1.3%	–1.2%	–1.5%	–1.3%	–1.4%
Male						
60–69	2008–2011	–2.6%	–2.5%	–1.6%	–1.5%	–2.5%
60–69	2012–2015	–2.2%	–2.2%	–1.4%	–1.2%	–2.2%
60–69	2016–2019	–1.7%	–1.9%	–1.1%	–0.9%	–1.8%
70–79	2008–2011	–2.2%	–1.9%	–0.8%	–1.2%	–2.0%
70–79	2012–2015	–1.8%	–1.6%	–0.7%	–0.9%	–1.7%
70–79	2016–2019	–1.3%	–1.3%	–0.5%	–0.5%	–1.5%
80–89	2008–2011	–1.3%	–0.8%	–0.5%	–0.9%	–1.3%
80–89	2012–2015	–1.0%	–0.7%	–0.5%	–0.9%	–1.2%
80–89	2016–2019	–0.8%	–0.6%	–0.5%	–0.8%	–1.1%
90–99	2008–2011	–2.0%	–1.6%	–1.6%	–1.5%	–1.3%
90–99	2012–2015	–1.7%	–1.5%	–1.7%	–1.5%	–1.3%
90–99	2016–2019	–1.4%	–1.4%	–1.5%	–1.5%	–1.2%