

Research Article

Deep Learning Incorporated Bühlmann Credibility in the Modified Lee–Carter Mortality Model

Joab Onyango Odhiambo 

Department of Mathematics, Meru University of Science and Technology, Meru, Kenya

Correspondence should be addressed to Joab Onyango Odhiambo; omishjabali@gmail.com

Received 11 October 2022; Revised 2 March 2023; Accepted 10 March 2023; Published 22 April 2023

Academic Editor: Jian Shi

Copyright © 2023 Joab Onyango Odhiambo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This research study proposes the inclusion of randomness or an error term in the modified Lee–Carter model, which improves the traditional Lee–Carter model for modeling and forecasting mortality risk for years in the actuarial science field. While the modified Lee–Carter model points out some of its common shortcomings, it has no distributional assumption that has been placed on the error/disturbance term. Incorporating a Gaussian distributional assumption on the error term is proposed, and then, the deep learning technique is used to obtain the parameter estimates. It is a departure from the traditional singular value decomposition estimation technique for estimating the parameters in the model. Finally, the Bühlmann credibility approach is incorporated into the model to determine its forecasting precision compared to the classical Lee–Carter model before applied in actuarial valuations.

1. Introduction

The Lee–Carter model has become the benchmark stochastic mortality model applied in many fields when modeling mortality rates, hedging longevity risk, forecasting, and predicting mortality risk. Many actuaries, statisticians, and demographers use the model in stochastic risk modeling from various actuarial applications. The conventional [1] model is defined as follows:

$$\ln(m(x, t)) = \alpha_x + \beta_x \kappa_t + e(x, t), \quad (1)$$

where $m(x, t)$ and t are the systematic mortality rate and time, respectively, having two constraints of $\sum_{x=1}^N \beta_x = 1$ and $\sum_{t=1}^T \kappa_t = 0$.

The random errors $e(x, t) \sim N(0, \sigma_t^2)$ and κ_t 's known as mortality indices are observed during the period. The stated two constraints ensure that the LC model is identifiable in its application. It is not possible to note the values of κ_t 's during computation; hence, the concept of the singular value decomposition method can be used when estimating the values of the unknown quantities (α_x , β_x , and $\kappa_t \forall$ all values of $x = 1, 2, 3, \dots, N$ and $t = 1, 2, 3, \dots, T$).

More interestingly, some mortality modeling papers have misunderstood the model in its operations, for instance, denoting $m(x, t)$ which is commonly defined as central death rates for a given life aged exactly x for a given time period of t years shown by [2–5] as

$$\begin{cases} \ln(m(x, t)) = \alpha_x + \beta_x \kappa_t, \\ \sum_{x=1}^N \beta_x = 1, \\ \sum_{t=1}^T \kappa_t = 0. \end{cases} \quad (2)$$

While the model in equation (2) can be pretty confusing since it states that the general mortality rate $m(x, t)$ is the death rate from the randomness property of κ_t 's. This concept is captured in several papers, such as in [6], Nar-yongo [7], and [8], where [1] is treated as having no distribution on the random error $e(x, t)$ as shown in equation (1). Furthermore, this might be a huge problem as it states that $\ln(m(x, t)) = \alpha_x + \beta_x * \kappa_t$ is dependent completely when both are calculated through the similar random

variable k_t , which are central death rates that are dependent across all age groups with the assumption of $\ln(m(x, t))$ and $\ln(m(x, t - 1))$ being dependent. This assumption becomes important when determined by the similar mortality trend k_t since the rates are independent across all ages t [9–11] and [12].

Today, several extensions and applications have been made, making actuarial science literature rich in terms of mortality modeling techniques. Also, the freeware statistical R package has included many packages, namely, “*StMoMo*,” “*demography*,” “*actuaryr*,” and “*gnm*,” thus helping in forecasting the rates of future mortality and at the same time fitting a model of time series especially to the predicted rates of the mortality index, see [2, 4, 13–17]. Deep learning mortality modeling has been conducted in [18], thus making it one of novel methodologies to deal with data paucity in developing countries.

In this research study, the concept of placing a distributional assumption on the error or disturbance term is proposed. It is suggested that a normal distributional assumption on the error term be placed, and then, the deep neural network techniques be used to train and obtain the parameter estimates. Besides, the concept of the Bühlmann credibility approach is introduced to get the mortality forecast. Using the credibility technique, an approach proposed in [19] is found to enhance the forecasting model precision.

2. Modeling of the Modified Lee–Carter Model

From equation (1), it is proposed to define the distribution of the random error term, $e(x, t)$, to be a normal density with a mean and variance of θ and σ^2 , respectively, at every age x . Since there exists a constraint on all the unobserved random mortality index of k_t directly, this ensures that the proposed model used in modeling is identifiable. The next step is to estimate the parameters in an environment of limited data availability like the Kenyan population.

2.1. Deep Neural Network Estimation

2.1.1. Model Parameters. It is suggested that the deep neural network technique be used to train and obtain the parameter estimates of the model. The mean-squared error is a loss function used in deep neural networks given by

$$\text{Loss}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y - \hat{y})^2, \quad (3)$$

where y and \hat{y} are actual and predicted values, respectively, from the deep neural network in n observations.

With one input layer having two nodes and a hidden layer of three nodes, its output node will have a single node [20]. After training the data, the imputed data in the input layer will have nodes within the hidden layer taking the corresponding values from its input layer before multiplying by a weight w_j and adding a corresponding bias of b_{i1} , with 1 representing the layer and $i \in \{1, 2\}$ representing its corresponding node [21].

From the above two-tier artificial neural network, we want to train the parameters in equation (1) to estimate its parameters. The ANN architecture can be presented as;

The above backpropagation and feedforward processes will be repeated many times until the errors are negligible when estimating the parameters from the ANN architecture shown in Figure 1. Using python software, we code and train to estimate the parameters for estimations.

Deep neural networks often capitalize on the ANN component. They work so well by improving a model since each node in the hidden layer makes both associations and grades the importance of the input in determining the nature of the output. Therefore, the deep network has multiple hidden layers. In addition, “deep” refers to the layers of the model being multiple layers deep.

2.1.2. Activation Functions

Definition 1. Let $R(y)$ be the Rectified Linear Activation Unit (ReLU) activation function given by

$$R(y) = \begin{cases} 0 & \text{for } y < 0, \\ y & \text{for } y \geq 0, \end{cases} \quad (4)$$

where $R'(y) = \begin{cases} 0 & \text{for } y < 0 \\ 1 & \text{for } y \geq 0 \end{cases}$, with the values ranging from 0 to ∞ .

ReLU defines the output functions of deep neural networks.

Definition 2. Let $S(y)$ be the sigmoid of the activation function within a deep neural network defined by

$$S(y) = 1/1 - e^{-y} \quad \text{for } S'(y) = \varphi(y)(1 - \varphi(y)), \quad \text{with values ranging from 0 to 1.}$$

This sigmoid defines the output functions of deep neural networks.

These two activation functions are used to provide nonlinearity without which deep neural networks cannot model nonlinear relationships of the systematic mortality risk.

2.1.3. How the above Activation Function Works. An activation function is a mathematical function that is applied to the output of a neuron in a neural network. Its purpose is to introduce nonlinearity into the network, allowing it to learn more complex patterns. Two activation functions will be used, namely, the sigmoid function and the ReLU (rectified linear unit) function. Each of these functions has a slightly different form and characteristics, but all of them serve to squash the output of a neuron into a range of values, which is more useful for the learning process.

In our case, the sigmoid function maps any input value to a value between 0 and 1, making it useful for modeling probabilities. The ReLU function maps any negative input value to 0 and any positive input value to itself, making it useful for introducing nonlinearity into the network without introducing additional complexity.

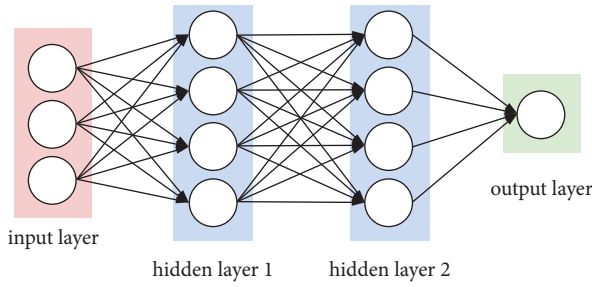


FIGURE 1: Parameter estimation under ANN architecture.

In summary, the activation function is an important part of the architecture of a neural network, and it plays a key role in the network's ability to learn and generalize to new data. The output data are the force of mortality, $\ln(m(x, t))$, calculated from the input data of survival probabilities of ${}_t p_x$ derived from the life table values for artificial neural networks.

2.2. Incorporation of the Bühlmann Credibility Approach

2.2.1. Mathematical Preliminaries. From a classical approach, the latest prediction U is defined as

$$U = Z(\bar{X}) + (1 - Z)\theta, \quad (5)$$

where \bar{X} is defined as the data sample mean, whereas θ is its general prior mean $E(X)$.

The Bühlmann credibility estimate or factor Z depends on the size of the sample n and the ratio of the expected present value of variance (EPV), and the variance of hypothetical mean (VHM) ratio is defined by K , which can be used in mortality modeling, see [22–24].

To be more precise, the value of Z varies with both the values of n and K in a way that Z increases as the number of sample size n increases of the data. The best estimate for the future value of $\hat{U}_{n+1} = Z(X + (1 - Z)\theta)$, where $Z = n/(n + k)$.

To apply the above concept of the Bühlmann Credibility approach to the model, $\ln(m(x, t))$ is used to model mortality. Figure 2 shows the curve of the expected value of $\ln(m(x, t))$ against the time of Kenyan data for both males and females at different ages.

To eliminate the downward trend, let $Y(x, t) = \ln(m(x, t)) - \ln(m(x, t - 1))$ for the values of $x = 1, 2, 3, \dots, n$ and $t = 1, 2, 3, \dots, m$ during forecasting. The Bühlmann credibility approach is used to compare the values with those of conventional [1], and it is assumed that $\hat{Y}(x_1, t)$ and $\hat{Y}(x_2, t), \hat{Y}(x_3, t), \dots, \hat{Y}(x_{n-1}, t), \hat{Y}(x_n, t)$ are all independent. Secondly, for values of $x = 1, 2, 3, \dots, n$, the distribution of $Y(x, t) = \ln(m(x, t)) - \ln(m(x, t - 1))$ and \forall values of x and t for $\hat{Y}(x, t)$ depends on the risk parameter $X = x$, and finally, $Y(x_1, t)/X = x, \dots, Y(x_n, t)/X = x$ are independent for each of $x = 1, 2, 3, \dots, n$.

Now, the Bühlmann Credibility estimate $\hat{Y}(x, t)$ is applied as follows:

$$Y(x, t) = Z \times Yx + (1 - Z)\hat{\theta}, \quad (6)$$

for values of $x = 1, 2, 3, \dots, n$ and the value of $Z = \{(n - 1)/(n - 1) + K\}$ and K is the ratio of the expected present value of variance (EPV) and the variance of hypothetical mean (VHM). In terms of parametric estimations, $EPV = \hat{\theta} = 1/(m \sum_{i=1}^n \bar{Y}x)$ and $VHM = \hat{v} = \{1/(n - 2 \sum_{i=1}^n (Yx - \bar{Y}x)^2)\}$.

2.2.2. Incorporating the Bühlmann Credibility into the Modified LC Model. The classical paper assumption of the overall mortality trend is taken following a simple random walk with a drift ϑ for the prediction of mortality: $\kappa_t = \kappa_{t-1} + \vartheta + e(t)$, where trend errors ($e(t)$) of the time follow Gaussian and are independently and identically distributed (*i.i.d.*) such that $e(t)$ and $t = t_x + 1, \dots, t_m$, see [8]. The error terms in our proposal are *i.i.d.* of white noises satisfying the Martingale structure in the following equation:

$$\begin{bmatrix} e(x, t) \\ e(t) \end{bmatrix} / F_{t-1} \sim N \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_e^2 \end{bmatrix} \right\}, \quad (7)$$

where F_{t-1} provides the information about the process up to a time $t - 1$ and covariances of the two random errors are zero.

Considering a random variable $W_{x,t}$ that denotes the central death rates between time t and $t - 1$ that means $W_{x,t} = \ln(m(x, t)) - \ln(m(x, t - 1))$,

$$W_{x,t} = \beta_x \times (k_t - k_{t-1}) + \Delta e(x, t), \quad (8)$$

$$W_{x,t} = \beta_x \times \vartheta + \beta_x \times e_t + \Delta e(x, t), \quad (9)$$

where the value of $x = x_1, \dots, x_m$ and $t = t_{x1} + 1, \dots, t_{xm}$ and $\Delta e(x, t) \sim N(0, 2 \times \sigma_x^2)$ and from equation (1), it is leading to $W_{x,t} \sim N(\beta_x \vartheta, \beta_x^2 \sigma_e^2 + 2\sigma_x^2)$, which follows the sum of independent variables of Gaussian distribution. From the conditional expectation as well as variance of $W_{x,t}$, it is easy to apply the Bühlmann credibility such that $\theta(x) = E[W_{x,t}/X] = \beta_x \vartheta$ and $\text{Var}[W_{x,t}/X] = \beta_x^2 \sigma_e^2 + 2\sigma_x^2$ in the same order. Since the expectation of value of the stated hypothetical mean, $\theta = E[\theta(x)] = E[E[W_{x,t}/X]] = \theta \times E[E[W_{x,t}/X]]$, and the estimated value of θ denoted as $\hat{\theta}$ is given by

$$\hat{\theta} = \frac{\hat{\vartheta}}{n} \sum_{x_1}^{x_n} \hat{\theta}_x = \frac{\hat{\vartheta}}{n}. \quad (10)$$

From equation (10), the variance process expected value denoted by $a = E[a(X)] = E[\beta_x^2 \sigma_e^2 + 2E[\sigma_x^2]]$ can be estimated as

$$\hat{a} = \frac{\hat{\sigma}^2}{n} \sum_{x_1}^{x_n} \hat{\theta}_x^2 + 2 \sum_{x_1}^{x_n} \frac{\hat{\sigma}^2}{m}, \quad (11)$$

while the hypothetical mean variance $c = \text{Var}[\theta(X)] = \vartheta^2 \cdot \text{Var}[\beta_x] = \vartheta^2 * E[\beta_x^2] - E[\beta_x]^2$. This can be estimated through

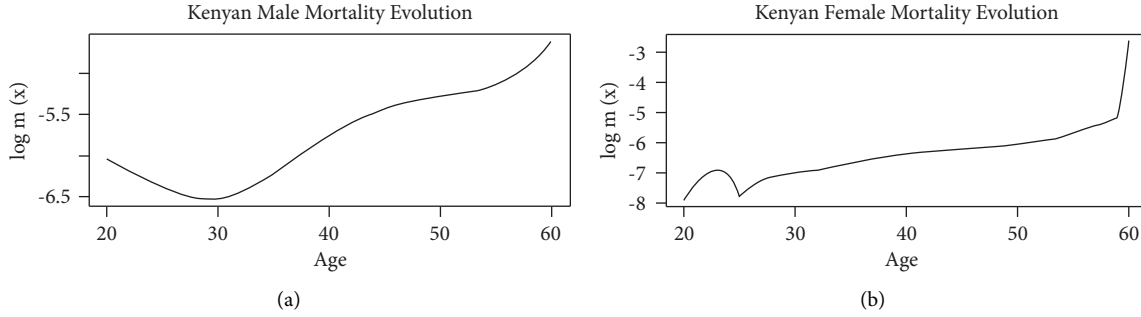


FIGURE 2: $\ln(m(x, t))$ against time t of Kenyan data for males and females (a, b).

$$\hat{a} = \hat{\vartheta}^2 \frac{\hat{\vartheta}}{n} \sum_{x_1}^{x_n} \theta_x^2 - \left(\frac{1}{n} \sum_{x_1}^{x_n} \theta_x \right)^2. \quad (12)$$

By writing the equation in the form of $Z(\bar{X}) + (1 - Z)\mu$, it is easy to estimate the value of $\hat{\mu}$ as $\hat{\vartheta}/n$, where Z is the credibility factor. More data availability from the Kenyan population means more emphasis is put on the Z values as the credibility estimate.

3. Data Analysis and Interpretation

To illustrate how this proposed model works, we apply it to Kenyan mortality data before making statistical inferences/deductions while noting the variations from the classical [1] model. To gain a firm conclusion, the population is studied under central death rates for males and females. The consolidated population is between 20 and 100 years after projecting it from 2020 to 2050 and using ten-year age lags. The lag time values of $M = 10$, as well as $T = 83$, are used.

The statistical R package known as “demography” is used on the model to obtain estimates for the values of α_x , β_x , and k_t 's after data training. The estimated values of α_x and β_x are tabulated in Tables 1 and 2, respectively, for both males and females and consolidated mortality rates.

The proposed inference is applied when fitting models to both males and females and combined rates of mortality. Moreover, the use of the “lm” package is applied when obtaining the proposed deep learning generated estimates before reporting the estimates for α_x 's, β_x 's, θ , and ϕ in Tables 1 and 2, respectively, for both males and females and consolidated mortality rates. Although the estimates for the value of ϕ obtained from the novel method is similar to those values derived from the traditional LC method, estimates for θ are different for both methodologies since the new method does not assume the value of $\sum_{t=1}^T k_t = 0$.

The proposed unit root test is also applied to both males and females and combined rates of mortality rates, where we will apply $\hat{\sigma}_e^2$ with the value of $L = 1/2\sqrt{T}$, \sqrt{T} , $2\sqrt{T}$ and the values of $\hat{\sigma}_e^2$, which is illustrated in equation (11). It is important to take a note that equation (11) is given as $k_0/T \rightarrow 0$ as $T \rightarrow \infty$.

All of the obtained estimates of variance, test statistics, and p values are recorded in Tables 3–6, respectively, for males, females, and consolidated mortality rates. According

to the illustrations, it is easy to note that these quantities are pretty good for L selection. Besides, novel root testing rejects the hypothesis of the unit root for females and consolidated mortality rates but refuses to reject the hypothesis of the unit root especially for male mortality rates.

Ultimately, we examine the soundness/robustness test of the stated deduction on the unit root hypothesis for the mortality index by rerunning the above unit root test for female, male, and consolidated population ages between 20 and 100 years.

4. Numerical Results and Forecasts

To examine the finite sample performance of the unit root test and estimators, the model is considered for the value of $M = 10$, $\hat{\alpha}_x$, $\hat{\beta}_x$, and θ estimates obtained from mortality rates as tabulated in Table 7 for the traditional model. The estimates are much higher than under deep learning methods.

To examine the finite sample performance of the unit root test and estimators, the model is considered for the value of $M = 10$, $\hat{\alpha}_x$, $\hat{\beta}_x$, and θ estimates obtained from mortality rates, as tabulated in Tables 1 and 2. The values of $e(t, x)$ are independent of (t) . The values of $\hat{\sigma}_e^2$ and σ_e^2 are tabulated in Tables 1 and 2. With a sample of 10,000 that are randomly drawn samples from the model, with a sample size of $T = 60$ and 100, the value of $\phi = 1$ is considered to yield the values.

Tables 1 and 2 show the estimates of parameters determined under deep learning techniques after fitting the model, and SE is the standard error of the model for both males and females and consolidated rates, respectively. In addition, Table 2 shows the consolidated rates of mortality for ages between 20 and 100 under deep learning, which are lower than those of the traditional method.

By computing the estimators for values of $\hat{\alpha}_x$'s, $\hat{\beta}_x$'s, θ , and ϕ within the settings illustrated above before reporting the mean as well as standard deviation of the above estimators, Tables 1 and 2 show that estimators for those values of $\hat{\alpha}_x$'s, $\hat{\beta}_x$'s, θ , and ϕ are all correct. The obtained unit root test has been investigated looking under the stated settings.

We also investigate $\hat{\sigma}_e^2$ with the values of $L = 1/2\sqrt{T}$, \sqrt{T} , $2\sqrt{T}$ and $\hat{\sigma}_e^2$ that has been denoted by the value of L in the equation Ω as well as the true value of σ_e^2

TABLE 1: Individual male and female (in brackets) systematic mortality rates for ages between 20 and 60.

x	1	2	3	4	5	6	7	8	9	10
$\hat{\alpha}_x$	-7.26 (0.17)	-7.08 (0.07)	-6.89 (-0.04)	-6.25 (-0.35)	-6.05 (-0.64)	-5.85 (-0.53)	-5.65 (-0.43)	-5.25 (-0.79)	-5.05 (0.55)	-4.85 (0.99)
$\hat{\beta}_x$	0.08 (0.15)	0.08 (0.13)	0.10 (0.123)	0.11 (0.10)	0.11 (0.10)	0.12 (0.089)	0.125 (0.08)	0.105 (0.08)	0.101 (0.08)	0.19 (0.08)
	Estimates									
$\hat{\theta}$	-0.128 (-1.384)									
$\hat{\phi}$	0.895 (0.875)									
	SE									
	0.018 (0.0278)									
	0.006 (0.005)									

TABLE 2: Consolidated rates of mortality for ages between 20 and 60 under deep learning.

x	1	2	3	4	5	6	7	8	9	10
$\hat{\alpha}_x$	-7.225	-7.005	-6.214	-5.995	-5.835	-5.722	-5.458	-5.122	-4.828	-4.325
$\hat{\beta}_x$	0.123	0.123	0.123	0.123	0.123	0.123	0.123	0.123	0.123	0.123
	Estimates									SE
$\hat{\theta}$										-0.145
$\hat{\phi}$										0.875
										0.021
										0.007

TABLE 3: Male and female (in brackets) rates of mortality for ages between 20 and 60.

L	$\hat{\sigma}_e^2$	Test statistic	P value	Z
$[0.5\sqrt{T}]$	0.072 (0.043)	0.832 (0.683)	$2.819e^{-18}$ ($2.819e^{-18}$)	0.25
$[\sqrt{T}]$	0.059 (0.048)	0.906 (0.712)	$5.441e^{-15}$ ($5.116e^{-15}$)	0.50
$[2\sqrt{T}]$	0.042 (0.035)	0.962 (0.912)	$1.855e^{-20}$ ($1.752e^{-20}$)	0.75
Ω	0.050 (0.045)	0.997 (0.952)	$1.887e^{-21}$ ($1.355e^{-21}$)	0.95

TABLE 4: Consolidated rates of mortality for ages between 20 and 60 under deep learning vs. the traditional method (in brackets).

L	$\hat{\sigma}_e^2$	Test statistic	P value	Z
$[0.5\sqrt{T}]$	0.053 (0.065)	14.835 (15.752)	$4.825e^{-4}$ ($5.8325e^{-4}$)	0.25
$[\sqrt{T}]$	0.066 (0.075)	12.762 (13.225)	$1.133e^{-3}$ ($1.255e^{-4}$)	0.50
$[2\sqrt{T}]$	0.055 (0.065)	14.822 (15.650)	$3.852e^{-4}$ ($4.245e^{-4}$)	0.75
Ω	0.0560 (0.0625)	12.795 (13.242)	$1.214e^{-3}$ ($1.4525e^{-4}$)	0.95

TABLE 5: Male and female (in brackets) rates of mortality for ages between 20 and 100.

L	$\hat{\sigma}_e^2$	Test statistic	P value	Z
$[0.5\sqrt{T}]$	0.168 (0.245)	2.832 (2.774)	$6.196e^{-15}$ ($5.235e^{-15}$)	0.25
$[\sqrt{T}]$	0.199 (0.258)	5.752 (5.238)	$1.424e^{-12}$ ($1.352e^{-12}$)	0.50
$[2\sqrt{T}]$	0.175 (0.264)	5.245 (5.087)	$1.582e^{-14}$ ($1.242e^{-14}$)	0.75
Ω	0.186 (0.259)	1.985 (1.815)	$5.568e^{-14}$ ($5.345e^{-14}$)	0.95

TABLE 6: Consolidated rates of mortality for ages between 20 and 100 under deep learning vs. the traditional method (in brackets).

L	$\hat{\sigma}_e^2$	Test statistic	P value	Z
$[0.5\sqrt{T}]$	0.194 (0.205)	12.882 (13.005)	$4.892e^{-4}$ ($5.125e^{-4}$)	0.25
$[\sqrt{T}]$	0.259 (0.265)	11.072 (12.125)	$1.481e^{-3}$ ($1.945e^{-4}$)	0.50
$[2\sqrt{T}]$	0.240 (0.255)	12.279 (12.750)	$9.53 e^{-4}$ ($10.725e^{-4}$)	0.75
Ω	0.235 (0.253)	12.985 (13.245)	$8.817e^{-4}$ ($9.455e^{-4}$)	0.95

TABLE 7: Consolidated rates of mortality for ages between 20 and 60 under the traditional method.

x	1	2	3	4	5	6	7	8	9	10
$\hat{\alpha}_x$	-8.355	-7.45	-6.45	-6.55	-5.996	-5.85	-5.80	-5.62	-5.20	-4.55
$\hat{\beta}_x$	0.15	0.14	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
	Estimates									SE
$\hat{\theta}$										-0.355
$\hat{\phi}$										0.985
										0.065
										0.012

being denoted in equation as Ω^* . The variance values of estimators and empirical sizes of the novel unit root test illustrated in the lower panel are tabulated in Tables 3–6 showing the larger size than that of the nominal level and the selection of limit L with an effect on the test. In addition, the size becomes more precise as the value of T becomes large.

Table 3 shows the variance estimates, p values, Z values, and test statistics for $L = 1/2\sqrt{T}$, \sqrt{T} , and $2\sqrt{T}$, where the value of $\Omega = \hat{\sigma}_e^2$.

Table 4 shows the variance estimates, p values, Z values, and test statistics for $L = 1/2\sqrt{T}$, \sqrt{T} , and $2\sqrt{T}$, where the value of $\Omega = \hat{\sigma}_e^2$.

In summary, Table 3 shows the consolidated rates of mortality for ages between 20 and 100 under deep learning, which are lower than those of the traditional method.

Table 5 shows the variance estimates, p values, Z values, and test statistics for $L = 1/2\sqrt{T}$, \sqrt{T} , and $2\sqrt{T}$, where the value of $\Omega = \hat{\sigma}_e^2$.

Table 6 shows the variance estimates, p values, Z values, and test statistics for $L = 1/2\sqrt{T}$, \sqrt{T} , and $2\sqrt{T}$, where the value of $\Omega = \hat{\sigma}_e^2$. In summary, Table 5 shows the consolidated rates of mortality for ages between 20 and 100 under deep learning, which are lower than those of the traditional method.

5. Conclusions and Recommendations

The above results have shown that using deep neural network techniques to estimate parameter values improved from the low levels of standard error (a measure of model volatility). Also, incorporating the Bühlmann credibility approach into the model has helped foster the model's accuracy.

This method is vital when forecasting mortality for future dates for higher levels of precision, while the use of data in the training of deep neural networks is essential when modeling mortality and forecasting because the choice of parameter estimation methods plays a vital role in showing the model's accuracy.

For policymakers, when modeling events of concern, such as the death rates of a population or infectious diseases, the choice of the parameter estimation method is critical to the model's accuracy. However, adequate data help improve the "quality" of prediction/forecasting under the incorporated Bühlmann credibility approach since data are the king in systematic mortality risk modeling and actuarial valuations.

Data Availability

Data are available on request.

Conflicts of Interest

The author declares that there are no conflicts of interest concerning the publication of this paper.

References

- [1] R. D. Lee and L. R. Carter, "Modeling and forecasting US mortality," *Journal of the American Statistical Association*, vol. 87, no. 419, pp. 659–671, 1992.
- [2] A. J. Cairns, D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, and M. Khalaf-Allah, "Mortality density forecasts: an analysis of six stochastic mortality models," *Insurance: Mathematics and Economics*, vol. 48, no. 3, pp. 355–367, 2011a.
- [3] K. Dowd, A. J. Cairns, D. Blake, G. D. Coughlan, D. Epstein, and M. Khalaf-Allah, "Evaluating the goodness of fit of stochastic mortality models," *Insurance: Mathematics and Economics*, vol. 47, no. 3, pp. 255–265, 2010.
- [4] J. Odhiambo, P. Ngare, and W. Patrick, "Bühlmann credibility approach to systematic mortality risk modeling for sub-Saharan Africa populations (Kenya)," *Research in Mathematics*, vol. 9, no. 1, Article ID 2023979, 2022.
- [5] V. Enchev, T. Kleinow, and A. J. G. Cairns, "Multi-population mortality models: fitting, forecasting and comparisons," *Scandinavian Actuarial Journal*, vol. 2017, pp. 319–342, 2017.
- [6] J. Li, H. Siu, "Pricing longevity risk with the parametric bootstrap: a maximum entropy approach," *Insurance: Mathematics and Economics*, vol. 47, no. 2, pp. 176–186, 2010.
- [7] R. Naryongo, J. Onyango, L. Njagi, and M. Nakirya, "Modeling of COVID-19 transmission under Markov chains in Uganda," *Journal of Applied Mathematics and Computation*, vol. 6, no. 1, 2022.
- [8] J. S. H. Li, W. S. Chan, R. Zhou, and R. Zhou, "Semicohherent multipopulation mortality modeling: the impact on longevity risk securitization," *Journal of Risk and Insurance*, vol. 84, no. 3, pp. 1025–1065, 2017.
- [9] S. H. Cox, Y. Lin, and T. Shi, "Pension risk management with funding and buyout options," *Insurance: Mathematics and Economics*, vol. 78, pp. 183–200, 2018.
- [10] K. Y. Kwok, M. C. Chiu, and H. Y. Wong, "Demand for longevity securities under relative performance concerns: stochastic differential games with cointegration," *Insurance: Mathematics and Economics*, vol. 71, pp. 353–366, 2016.
- [11] Q. Liu, C. Ling, and L. Peng, "Statistical inference for Lee-Carter mortality model and corresponding forecasts," *North American Actuarial Journal*, vol. 23, no. 3, pp. 335–363, 2019.
- [12] T. W. Wong, M. C. Chiu, and H. Y. Wong, "Managing mortality risk with longevity bonds when mortality rates are cointegrated," *Journal of Risk and Insurance*, vol. 84, no. 3, pp. 987–1023, 2017.
- [13] N. Brouhns, M. Denuit, and J. K. Vermunt, "A Poisson logbilinear regression approach to the construction of projected lifetables," *Insurance: Mathematics and Economics*, vol. 31, no. 3, pp. 373–393, 2002.
- [14] F. Giroso and G. King, *Understanding the Lee-Carter Mortality Forecasting Method* Institute for Quantitative Social Science, Cambridge, MA, USA, 2007.
- [15] N. Li and R. Lee, "Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method," *Demography*, vol. 42, no. 3, pp. 575–594, 2005.
- [16] E. Bisetti and C. A. Favero, "Measuring the impact of longevity risk on pension systems: the case of Italy," *North American Actuarial Journal*, vol. 18, no. 1, pp. 87–103, 2014.
- [17] J. Odhiambo, P. Weke, P. Ngare, R. Naryongo, and S. Sewe, "Poisson incorporated credibility regression modelling of systematic mortality risk for populations with finite data," *Mathematical Problems in Engineering*, vol. 2022, Article ID 1753542, 14 pages, 2022.
- [18] J. Odhiambo, P. Weke, and P. Ngare, "A deep learning integrated Cairns-Blake-Dowd (CBD) systematic mortality risk model," *Journal of Risk and Financial Management*, vol. 14, no. 6, p. 259, 2021.
- [19] C. C. L. Tsai, T. Lin, and T. Lin, "Incorporating the Bühlmann credibility into mortality models to improve forecasting performances," *Scandinavian Actuarial Journal*, vol. 2017, pp. 419–440, 2017.
- [20] M. Bayat, "Application of artificial neural networks for predicting tree survival and mortality in the Hyrcanian forest of Iran," *Computers and Electronics in Agriculture*, vol. 164, Article ID 104929, 2019.

- [21] S. Rocha, J. Silva Soares, Artificial neural networks: modeling tree survival and mortality in the Atlantic Forest biome in Brazil,” *Science of the Total Environment*, vol. 645, pp. 655–661, 2018.
- [22] M. C. Christiansen and E. Schinzinger, “A credibility approach for combining likelihoods of generalized linear models,” *ASTIN Bulletin*, vol. 46, no. 3, pp. 531–569, 2016.
- [23] O. Joab, *Stochastic Modelling of Systematic mortality Risk under Collateral Data and its Applications*, Ph.D. thesis, University of Nairobi, Nairobi, Kenya, 2022.
- [24] J.-B. Coulomb, Y. Salhi, and P.-E. Therond, “Credibility adjustment of the lee-carter longevity model for multiple populations,” HAL SHS, Lyon, France, 2020.