Quantile-based Interpretable Neural Network Models: Mortality Forecasting and Actuarial Simulations

Yang Qiao[a] Jinggong Zhang[a] Wenjun Zhu[a] Chou-Wen Wang[b*]

[a]Division of Banking and Finance, Nanyang Technological University

[b]Department of Finance, National Sun Yat-Sen University

Accepted version by *Insurance: Mathematics and Economics*

Abstract

This paper introduces a class of quantile-based, interpretable neural network (NN) models for mortality prediction: the Lee-Carter neural network (LCNN) model, the Renshaw-Haberman neural network (RHNN) model, and the simple neural network (simpleNN) model that balances simplicity with predictive performance. These models preserve the linear interpretability of classic stochastic mortality models while harnessing the flexibility of NNs to capture complex nonlinear patterns in mortality data, thereby achieving enhanced predictive performance. Leveraging a composite loss function that integrates pinball, median anchoring, and quantile-crossing penalty terms, we estimate mortality quantiles and develop an efficient simulation scheme based on interpolation. This framework provides distributional insights essential for pricing, reserving and risk management. Extensive empirical analyses across multiple populations demonstrate that the proposed models consistently outperform traditional approaches in predictive accuracy. We further illustrate their practical utility through an application to longevity swap pricing.

Key Words: Interpretable Neural Network, Mortality Prediction, Longevity Risk, Stochastic Mortality Model, Quantile Regression, Monte Carlo Simulation, Longevity Swap

*Corresponding author.

1 Introduction

Population aging and longevity risk have become critical global challenges, posing significant societal and economic pressures on various stakeholders (Cairns et al., 2009; Blake et al., 2013). Predictive models for mortality improvement are therefore indispensable for understanding and managing longevity risk while addressing the broader implications of population aging. For example, life insurers and annuity providers require future mortality projections to adjust premiums and benefits; regulators mandate life insurers and pension funds to account for longevity risk to ensure compliance and solvency; governments need reliable mortality forecasts for reforms to retirement ages, contribution rates, and benefit structures to maintain the sustainability of social security systems (Zhou et al., 2014; Office of the Chief Actuary, 2024; Atance et al., 2025). Such mortality models provide critical tools to support effective planning and decision making by stakeholders such as governments, insurers, pension funds, and healthcare providers.

Since the early 1990s, stochastic mortality models have been popularly advocated and adopted to extrapolate past mortality trends by fitting statistical models. These models explicitly account for uncertainties in mortality projections and scenario analysis, while maintaining interpretable factors such as age, time, and cohort effects. This interpretability allows users to adapt the models to specific contexts. The Lee-Carter (LC) model (Lee and Carter, 1992) characterizes the logarithm of the central death rate as a linear function of age and a time trend for mortality rates. The Renshaw-Haberman (RH) model (Renshaw and Haberman, 2006) extends the LC model by incorporating a cohort effect. The Cairns-Blake-Dowd (CBD) model (Cairns et al., 2006) advances the literature by employing simple parametric forms with logistic regression, effectively addressing identifiability issues and improving predictive accuracy for older age groups. Cairns et al. (2009) conduct a comprehensive horserace analysis to evaluate the performance of different mortality models. Further developments by Haberman and Renshaw (2012, 2013) identify that increments in the logarithm of death rates follow a Gaussian distribution. Based on this finding, Mitchell et al. (2013) modify this type of models by choosing the increments of the logarithm of mortality rates as explanatory variables to address confounding issues with the time trend during the estimation process.

Another key direction of research of mortality modeling is the investigation of co-movements in mortality trends across different populations. Li and Lee (2005) identify a common factor in

the mortality rates of different population groups, suggesting that leveraging information from other populations enhances predictive accuracy. This approach has inspired subsequent studies on multi-population stochastic mortality models (Li and Hardy, 2011; Dowd et al., 2011; Cairns et al., 2011b; Jarner and Kryger, 2011; Zhou et al., 2013, 2014; Li et al., 2017; Li and Liu, 2020; Li et al., 2021). Additionally, dependency structures in mortality rates have been explored through methods such as copulas (Chen et al., 2015; Lin et al., 2015; Wang et al., 2015; Zhu et al., 2017; Chen et al., 2017; Zhou, 2019; Zhou and Ji, 2021).

In recent years, the rapid advancements in neural networks (NNs) and machine learning (ML) have been driven by both the breakthroughs in algorithms and the increasing computational power of modern hardware, enabling unprecedented performance across diverse areas including actuarial science (e.g., Gabrielli and Wüthrich, 2018; Noll et al., 2018; Kuo, 2019; Delong et al., 2022; Noll et al., 2018; Schelldorfer and Wuthrich, 2019; Gao and Wüthrich, 2018; Gao et al., 2019, 2022; Chen et al., 2024; Nigri et al., 2021; Levantesi et al., 2022; Zhu, 2024). Specifically, there is growing interest in applying ML and NNs to mortality and longevity forecasting, as these methods effectively handle high-dimensional inputs and intricate non-linear relationships. Hainaut (2018) incorporate an autoencoder into the LC model framework, demonstrating improved performance over the original LC model. Richman and Wüthrich (2021) extend the multi-population LC model using NNs, and Miyata and Matsuyama (2022) employ a variational autoencoder for Bayesian estimation.

Architectural innovations have expanded the scope of NN-based mortality models. Early recurrent approaches by Nigri et al. (2019) integrat LSTM networks within the Lee-Carter framework, while Richman and Wuthrich (2019) demonstrate their predictive superiority through empirical validation. Data-efficient adaptations by Lindholm and Palmborg (2022) enable LSTM use in populations with limited data, and Nigri et al. (2021) extend beyond mortality rates to forecast life expectancy and lifespan disparity. Comparative studies have further evaluated architectures. Perla et al. (2021) use shallow CNNs to reveal geographic clustering patterns. Chen and Khaliq (2022) identify GRUs as consistently strong performers across U.S. regions. Zhang et al. (2022) propose hybrid LSTM-CNN architectures and Bravo (2021) investigate RNN-LSTM applications for Portuguese mortality data.

Alternative data representations and novel spatial-structural approaches have further trans-

formed mortality modeling. Wang et al. (2021) introduce 2D mortality neighborhood prediction processed through CNNs, with Qiao et al. (2024) extending this framework through ensemble long-term forecasts. Hsiao et al. (2024) develop mortality improvement NNs with cohort and autoregressive effects, avoiding external time-series models. Multi-population methods have also gained prominence, with Richman and Wüthrich (2021) using embedding layers to learn optimal feature representations, Scognamiglio (2022) developing NN calibration for LC across populations, and Dong et al. (2020) applying tensor decompositions for parsimonious representations. Levantesi et al. (2022) combine clustering with multivariate LSTMs to strengthen cross-population forecasts.

Finally, several studies focused on uncertainty quantification. Marino et al. (2023) integrate bootstrap prediction intervals into LSTM frameworks. Schnürch and Korn (2022) develop CNN-based model with interval forecasts through bootstrapping. Corsaro et al. (2024) propose a quantile mortality framework using NNs to model multiple populations, offering distributional forecasts. Richman (2022) highlight the broader challenges of stability, uncertainty, and explainability in actuarial deep learning.

Despite these advances, NN-based extensions to classic stochastic mortality models face two major obstacles that remain inadequately addressed in the literature, limiting their practical applications. First, NNs are often regarded as "black-box" systems, making it challenging for users to understand how predictions are generated, thereby undermining confidence and adoption. Unlike classic models such as LC, CBD and RH, NN-based mortality models often lack transparency in representing key effects, which is critical in actuarial contexts. Actuaries require interpretable models to communicate results to various stakeholders effectively, and identify and manage risks in a prudent and defensible manner. Balancing interpretability and predictive performance is, therefore, essential (Vaughan et al., 2018; Agarwal et al., 2021; Richman, 2022; Hsiao et al., 2024). Second, most of the existing literature on NN-based models focuses on improving point predictions, while only a limited number of studies address distributional forecasting (with a few recent exceptions such as (Santolino, 2020; Corsaro et al., 2024)), which strongly restricts their applicability in scenarios that demand a comprehensive understanding of uncertainty for informed decision-making. Typical actuarial tasks, such as pricing and reserving, fundamentally require distributional information about the model outputs instead of merely

point estimates (e.g., Cairns et al., 2011a). As highlighted by Richman (2022), understanding and quantifying model uncertainties plays a vital role in actuaries' decision making process.

This paper aims to address the aforementioned challenges by proposing quantile-based interpretable NN-based models for mortality improvement and exploring their practical applications in longevity-linked derivative pricing.

More specifically, we develop the Lee-Carter Neural Network (LCNN) and Renshaw-Haberman Neural Network (RHNN) models by incorporating embedding and fully connected layers in extending the factor representations in the original LC and RH models. In addition, we introduce a simple Neural Network (simpleNN) model with streamlined age and cohort effects compared to LCNN and RHNN. These models maintain interpretable linear structures with transparent parameters, and meanwhile, thanks to their NN components, achieve improved predictive accuracy. They also simplify the modeling process by removing the need for complex underlying models (e.g., time series or copula models), thereby reducing potential model risk.

To address the second challenge and provide distributional information of predictions, we employ a multi-pinball loss function during NN training to estimate quantiles of predicted mortality rates. This allows us to develop a simulation scheme that interpolates among these quantiles to generate plausible mortality paths. The pinball loss function, originally introduced for quantile regression (Koenker and Hallock, 2001; Koenker, 2005) and has received attention among actuaries for handling the asymmetries in the distribution of residual and generating distributional insights for mortality simulations (Laporta et al., 2024; Fissler et al., 2023; Richman, 2022; Zhang, 2024). We illustrate its effectiveness through a derivative pricing example, where simulated paths are used to perform the Wang transform to obtain risk-neutral probabilities.

It is important to highlight that our approach differs from recent neural network mortality models in several key respects. Hsiao et al. (2024) extend Mitchell et al. (2013) by incorporating cohort and autoregressive effects, thereby improving short-run dynamics through dependence on lagged mortality rates. While this innovation enhances point forecast accuracy, it remains essentially a time-series specification and does not address the need for distributional forecasting. More importantly, whereas Hsiao et al. (2024) emphasize country-specific parameterizations to provide detailed population-level insights, our models demonstrate that a common factor model - where populations share b_x and k_t - achieves comparable accuracy with substantially greater

parsimony. This design requires far fewer parameters while effectively capturing the universal structure of mortality improvements across populations.

Corsaro et al. (2024) develop a quantile mortality framework based on neural networks, estimating quantiles separately within a multi-population Lee-Carter setting. Although this design extends distributional modeling, it predicts each quantile independently, with the neural network serving primarily as a fitting device to pool cross-country data. This structure may lead to potential coherence issues such as quantile crossing, which limits its applicability for actuarial simulations. By contrast, our framework resolves this limitation through a joint quantile prediction with a composite loss function, incorporating median anchoring and monotonicity penalties. This approach ensures internal consistency across quantiles, and yields coherent distributional forecasts directly applicable to simulation-based actuarial tasks, including longevity risk valuation and derivative pricing. Finally, neither Hsiao et al. (2024) nor Corsaro et al. (2024) extend their methodologies to the actuarial practice in longevity risk management. Our paper is the first to bridge this gap, demonstrating how neural network-based mortality models can be harnessed for the pricing and risk management of longevity derivatives.

This paper justifies its value within the rapidly expanding literature on machine learning in longevity analysis by addressing two persistent, unsolved barriers to actuarial adoption: the *black-box* nature of neural networks and the absence of coherent, distribution-free stochastic simulations. Specifically, our contribution is fourfold. First, we advance the literature on interpretable neural networks by proposing a class of models (LCNN, RHNN, and SimpleNN) that embed classical stochastic age–period–cohort structures within neural network architectures. The use of embedding layers with explicit identifiability constraints preserves transparent factor interpretations while enhancing predictive performance through representation learning. In particular, the proposed RHNN architecture resolves the well-known convergence and numerical instability issues of the RH model, while delivering consistently improved accuracy in multi-population settings. This design moves beyond *black-box* mortality prediction and establishes a principled bridge between traditional actuarial factor models and modern representation learning.

Second, we develop a distribution-coherent multi-quantile estimation framework tailored for actuarial applications. Unlike existing neural quantile approaches that estimate quantiles

independently, we introduce an enhanced estimation strategy based on a composite pinball loss function with median anchoring and monotonicity penalties. This ensures internal distributional coherence and directly addresses quantile crossing, a critical issue in quantile-based simulation frameworks.

Third, leveraging the coherent multi-quantile design, we construct a flexible and distribution-free stochastic simulation framework. Dynamic factor representations are learned directly within the neural architecture, removing the need for separate ARIMA or copula layers and thereby reducing model layering and projection risk. The resulting framework offers an integrated approach to multi-population mortality forecasting without restrictive parametric assumptions.

Finally, to the best of our knowledge, this paper provides the first implementation of longevity-linked security pricing based on NN-driven stochastic mortality models. This work extends neural network methodologies beyond prediction to practical longevity risk management by integrating data-driven mortality forecasting with market-consistent valuation. The approach is readily adaptable to broader actuarial applications, including reserve planning and portfolio hedging.

The remainder of the article is organized as follows. Section 2 outlines the specifications of the proposed framework. Section 3 presents the empirical analysis including the dataset, evaluation framework, and modeling results compared with traditional models. Section 4 investigates an empirical case study on longevity-linked derivative pricing. Section 5 concludes.

2 Methodology

2.1 Stochastic mortality models

We begin with traditional stochastic mortality models and their notations. Let $\mathcal{X} = \{x_0, ..., x_\omega\}$ be the set of ages considered by the mortality study, and $\mathcal{T} = \{t_0, ..., t_n\}$ be the set of calendar years. As shown by Cairns et al. (2009), a large class of stochastic mortality models can be represented by the following form:

$$\varphi\left(m_{x,t}^{r,g}\right) = \sum_{v} \beta_{x,v}^{r,g} \cdot \kappa_{t,v}^{r,g} \cdot \gamma_{t-x,v}^{r,g}, \quad v = 1, 2, ..., V, \tag{1}$$

where $m_{x,t}^{r,g}$ represents the mortality rate at age $x \in \mathcal{X}$, time point $t \in \mathcal{T}$, region (or group) $r \in \mathcal{R}$, and gender $g \in \mathcal{G} = \{\text{male}, \text{female}\}$; $\beta_{x,v}^{r,g}$ represents the age effect parameter; $\kappa_{t,v}^{r,g}$ represents the time effect parameter; $\gamma_{t-x,v}^{r,g}$ represents the cohort effect parameter; V is the number of terms in the model; and $\varphi(\cdot)$ represents the transformation function.

In this paper we compress the notation of $m_{x,t}^{r,g}$ to $m_{x,t}$ (and similarly for other notations) whenever it does not cause any ambiguity. Although group and gender are not explicitly mentioned, they are still included in the multi-population model setting.

A well-known special case of (1) is the Lee-Carter (LC) model (Lee and Carter, 1992). When the transformation function $\varphi(\cdot)$ is chosen as then logarithm function, $V = 2$, and cohort-related parameters are omitted, equation (1) simplifies to the LC model,

$$\log(m_{x,t}^{LC}) = a_x^{LC} + b_x^{LC} k_t^{LC}. \tag{2}$$

This model structure captures age and time effects, where coefficient a_x^{LC} contains the average age effect parameter for a specific age x, b_x^{LC} is an age effect parameter, k_t^{LC} represents the time effect parameter at time t. Similarly, when the transformation function $\varphi(\cdot)$ is a logarithm function and $V = 3$, equation (1) becomes the Renshaw-Haberman (RH) model (Renshaw and Haberman, 2006), expressed as:

$$\log\left(m_{x,t}^{RH}\right) = a_x^{RH} + b_x^{RH} k_t^{RH} + \eta_x^{RH} \gamma_{t-x}^{RH}, \tag{3}$$

where a_x^{RH}, b_x^{RH} and k_t^{RH} retain the same interpretations as in the LC model; γ_{t-x}^{RH} represents the cohort effect parameter; η_x^{RH} represents the sensitivity of a specific age x to γ_{t-x}^{RH}.

2.2 Mortality improvement model

One key issue of the model formulation (1), as elaborated by Mitchell et al. (2013), is that the estimation process confounds with the time trend and thus the correlation of mortality rates among ages, genders and populations is likely to be overestimated. To overcome this problem, we follow Mitchell et al. (2013) and consider the mortality improvement version of model (1) . Specifically, the LC model (2) is modified to,

$$\log\left(m_{x,t+1}^{LCNN}\right) - \log\left(m_{x,t,n}^{LCNN,AVG}\right) = a_x^{LCNN} + b_x^{LCNN} k_t^{LCNN}, \tag{4}$$

where

$$\log\left(m_{x,t,n}^{LCNN,AVG}\right) := \frac{1}{n} \times \sum_{i=1}^{n} \log\left(m_{x,t-i+1}^{LCNN}\right), \tag{5}$$

representing the moving average of log mortality across the preceding n years. The left-hand side of (4) represents the difference form of mortality rates over time, which effectively makes the time series stationary and removes the confounding issue among different variables. The $\log\left(m_{x,t,n}^{LCNN,AVG}\right)$ represents the moving average over a period of n up to time t and mitigates the impact of temporary mortality shocks. When $n = 1$, it simplifies to $\log\left(m_{x,t}^{LCNN}\right)$, the original mortality improvement model proposed by Mitchell et al. (2013). Similarly, the improvement version of the RH Model (3) is given by,

$$\log\left(m_{x,t+1}^{RHNN}\right) - \log\left(m_{x,t,n}^{RHNN,AVG}\right) = a_x^{RHNN} + b_x^{RHNN} k_t^{RHNN} + \eta_x^{RHNN} \gamma_{t-x}^{RHNN}, \tag{6}$$

where

$$\log\left(m_{x,t,n}^{RHNN,AVG}\right) = \frac{1}{n} \times \sum_{i=1}^{n} \log\left(m_{x,t-i+1}^{RHNN}\right), \tag{7}$$

which is defined in the same way as Equation (5). Note that the moving averages in Equations (5) and (7) are estimated from historical data during training, but are computed recursively in the prediction phase, progressively incorporating forecasted values.

To incorporate a stochastic component into the model, the time effect k_t^{LCNN} in (4) and k_t^{RHNN} in (6), and the cohort effect γ_{t-x}^{RHNN} in (6), are typically further modeled by time series models such as ARIMA processes. For future mortality projection, one will extrapolate the time series models for the time and cohort dimensions while holding the age effect parameters unchanged. While the Model (1) provides interpretability through its factor structure, the time and cohort parameters (k_t and γ_{t-x}) must be estimated and subsequently modeled with time-series techniques for projection. This introduces substantial model risk, especially in a multi-population setting where the dependence can be complex across populations. In this paper, we incorporate embedding layers and fully connected layers of neural networks to extract factor representations. The resulting models preserve the interpretable linear structure of classical formulations while delivering improved predictive accuracy. Moreover, they streamline the modeling process by eliminating the need for separate time-series specifications, thereby mitigating potential model risk.

2.3 Neural networks and embedding layers

NNs are inspired by biological neural systems and use interconnected neurons in a layered structure. Each layer applies a linear transformation followed by a (possibly) non-linear activation function, forming a composite function. This structure allows NNs to approximate arbitrarily complex relationships, making them fundamental to deep learning and artificial intelligence. The most common architecture is the fully connected feed-forward neural network, where information flows in one direction and each neuron connects to all neurons in the subsequent layer. This network consists of three components: an input layer, a number of hidden layers, and an output layer. Let $\mathbf{x} = (x_1, x_2, ..., x_N)'$ denote the N-dimensional input features, and $\mathbf{h}^{(l)} = \left(h_1^{(l)}, h_2^{(l)}, ..., h_{n_l}^{(l)}\right)'$ represent the l-th layer with n_l neurons for $l = 1, 2, ..., L$. Also denote $\mathbf{y}$ as the output features, and then a fully connected neural network with L intermediate layers can be represented as follows,

$$\mathbf{h}^{(1)} := F^{(0)}(\mathbf{x}) = \sigma^{(0)}\left(\mathbf{b}^{(0)} + \mathbf{W}^{(0)}\mathbf{x}\right),$$

$$\mathbf{h}^{(l)} := F^{(l-1)}\left(\mathbf{h}^{(l-1)}\right) = \sigma^{(l-1)}\left(\mathbf{b}^{(l-1)} + \mathbf{W}^{(l-1)}\mathbf{h}^{(l-1)}\right), \quad l = 2, ..., L,$$

$$\hat{\mathbf{y}} := F^{(L)}\left(\mathbf{h}^{(L)}\right) = \sigma^{(L)}\left(\mathbf{b}^{(L)} + \mathbf{W}^{(L)}\mathbf{h}^{(L)}\right),$$

where $F^{(\cdot)}(\cdot)$ represents a composite transform from the previous layer to the next; $\mathbf{b}^{(j)}$ is a n_j-by-1 bias vector, and $\mathbf{W}^{(j)}$ is a n_j-by-n_{j-1} dimensional weight matrix for the linear transform; $\sigma^{(j)}(\cdot)$ is known as the activation function to perform the (possibly) non-linear transform; $\hat{\mathbf{y}}$ is the final output of the model, that is, the predicted value of $\mathbf{y}$. As shown by Richman (2022), neural networks can also be viewed as a more generalized form of Generalized Linear Models (GLMs), providing a familiar theoretical framework for actuarial applications.

Embedding layers are a type of representation learning method used in NNs to handle categorical data such as words, phrases, or unique identifiers. In mortality modeling, as introduced by Richman (2022), categorical features can also be processed with embedding layers. These layers take integer inputs representing categorical indices and map them to dense, information-rich vectors in a continuous space. In other words, the function of an embedding layer is to map from a discrete space to an encoded vector, performing feature abstraction and encoding while preserving the inherent similarities among the inputs. For example, in natural language

processing, each word is assigned an encoded vector, expecting that the smaller the distance between these vectors, the closer their meanings are.

As part of an NN model, embedding layers can learn and adjust through input data, and thus more accurately reflect the relationships of categorical data for classification, regression, and prediction operations. Specifically, let $\Psi = \{\psi_1, ..., \psi_l\}$ be the possible integer indices of categories, with l being the number of different elements in the set. An embedding layer E_Ψ maps $\psi_j \in \Psi$ to a real-valued vector as shown below:

$$E_\Psi(\psi_j) = (E_{\Psi,1}(\psi_j), ..., E_{\Psi,d}(\psi_j))',$$

where $d < l$ is the dimension of vector after embedding.

Formally, an embedding layer is represented by a weight matrix $\mathbf{W}_E \in \mathbb{R}^{l\times d}$. For a categorical variable with index j, the embedding operation selects the j-th row of $\mathbf{W}_E$. These weights are optimized through backpropagation together with the other parameters of the network. In our mortality modeling framework, we set $d = 1$ for all categorical variables, including country, gender, age, year, and cohort. This choice facilitates direct interpretability: each weight corresponds to the effect of a specific category, preserving the interpretive structure of classical Lee-Carter parameters while minimizing the number of parameters to be estimated.

2.4 LCNN Model and RHNN model

In this section we propose the Lee-Carter Neural Network (LCNN) model and the Renshaw-Haberman Neural Network (RHNN) model by applying the modeling techniques in Section 2.3 to mortality improvement models (4) and (6), respectively. The model architectures of these two NN-based models are illustrated by schematic diagrams in Figures 1 and 2.

In the LCNN model, the first sub-network primarily handles the computation of a_x^{LCNN} in equation (4), which involves an embedding layer, a concatenating layer and a fully connected layer. Since a_x^{LCNN} is related to the group r, gender g, and age x, as in the approach of Richman (2022) and Hsiao et al. (2024), they are treated as categorical input features in this model. Let ϱ_R, ϱ_G, and ϱ_X denote the vector dimensions of embedding layers for group r, gender g, and age x, respectively. Therefore, the embedding layers $E_R^a(r)$, $E_G^a(g)$ and $E_X^a(x)$ are defined by:

$$E_R^a := E_R^a(r) = (E_{R,1}(r), ..., E_{R,\varrho_R}(r))',$$

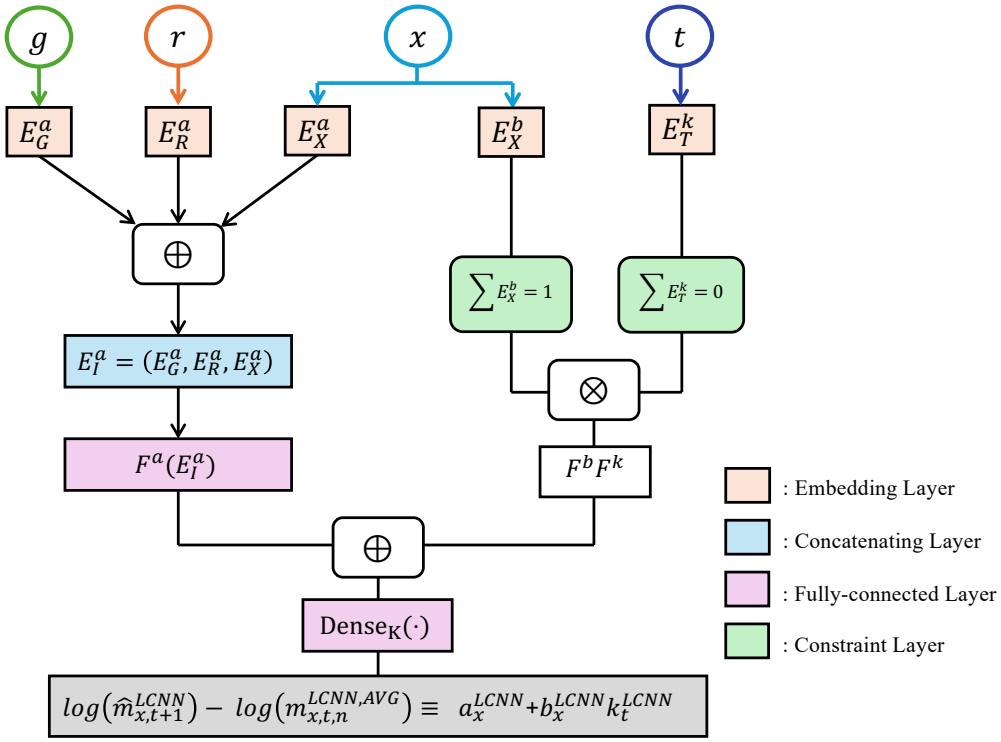


Figure 1: Schematic diagram of the architecture for the LCNN model

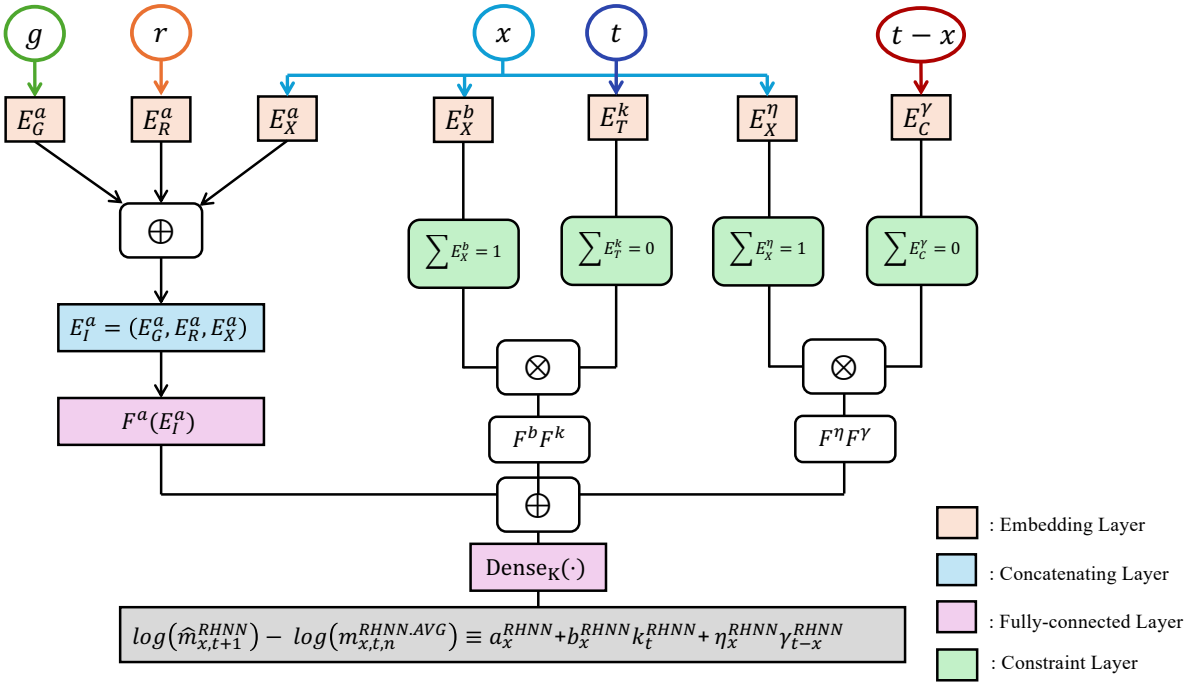


Figure 2: Schematic diagram of the architecture for the RHNN model

$$E_G^a := E_G^a(g) = (E_{G,1}(g), ..., E_{G,\varrho_G}(g))',$$

and

$$E_X^a := E_X^a(x) = (E_{R,1}(x), ..., E_{R,\varrho_X}(x))',$$

where embedding layers E_R^a, E_G^a, and E_X^a represent the feature vectors for group r, gender g, and age x, respectively. Next, $E_I^a := E_I^a(r, g, x) = (E_R^a, E_G^a, E_X^a)$ is obtained by concatenating the outputs of these three embedding layers. Finally, through the fully connected layer F^a with input feature E_I^a, we obtain

$$\hat{a}_x^{LCNN} := F^a(E_I^a) = \sigma^a\left(\omega_0^a + \omega_R^a E_R^a + \omega_G^a E_G^a + \omega_X^a E_X^a\right),$$

In this formulation, the age embedding for b_x^{LCNN} and the time embedding for k_t^{LCNN} are given by

$$\hat{b}_x^{LCNN} \propto E_X^b, \qquad \hat{k}_t^{LCNN} \propto E_T^k.$$

The RHNN model given by Equation (6) can be structured and estimated in a similar manner, using the architecture illustrated in Figure 2. Compared to LCNN, RHNN has two additional terms, where η_x^{RHNN} represents age embedding and γ_{t-x}^{RHNN} represents cohort embedding, given by

$$\hat{\eta}_x^{RHNN} \propto E_X^\eta, \qquad \hat{\gamma}_{t-x}^{RHNN} \propto E_C^\gamma.$$

The factors introduced in the LCNN and RHNN models are set proportional to the embedding outputs, which undergo linear transformations within the neural networks.

Identifiability constraint To model ensure identifiability, we use the same constraint as in traditional LC and RH models. Specifically, in the LCNN, we apply two constraints: $\sum_x b_x = 1$ and $\sum_t k_t = 0$; while in the RHNN, we extend these requirements by adding two more, so that all four constraints are enforced: $\sum_x b_x = 1$, $\sum_t k_t = 0$, $\sum_x \eta_x = 1$, and $\sum_{t-x} \gamma_{t-x} = 0$. At each training step, these constraints are maintained through weight normalization, which uniformly redistributes adjustments across the relevant parameters after every gradient update.

Similar to the traditional LC and RH models, once the model parameters are estimated, parameter values representing time, age and cohort effects are directly observable, enabling users

to understand the rationale behind the model's predictions. Our choice of embedding layers with shallow feedforward structures naturally handles categorical variables while maintaining direct correspondence to classical model parameters. This architecture automatically learns factor interactions that traditional models cannot capture without manual specification, yielding improved predictive performance while preserving interpretability, as demonstrated in the empirical analysis.

2.5 SimpleNN model

To further explore the balance between model parsimony and predictive sophistication, we introduce a simpleNN model alongside the LCNN and RHNN frameworks. The simpleNN model is defined by,

$$\log\left(\hat{m}^{\text{simple}}_{x,t+1}\right) - \log\left(m^{\text{simple,AVG}}_{x,t,n}\right) = a^{\text{simple}}_{x} + w_1 \cdot t + w_2 \cdot (t-x), \tag{8}$$

where

$$\log\left(m^{simple,AVG}_{x,t,n}\right) = \frac{1}{n} \times \sum_{i=1}^{n} \log\left(m^{simple}_{x,t-i+1}\right), \tag{9}$$

where w_1 and w_2 are the learned weights in the neural network, with a^{simple}_x, time t, and cohort $t-x$, as input features. The model architecture is graphically illustrated by Figure 3. The simpleNN model maintains the general linear form, providing strong generalizability and offering even greater interpretability compared to the LCNN and RHNN models by directly inputting both time and cohort effects into the output layer without embedding. Additionally, it generally reduces the overall computational cost and the risk of over-fitting while retaining the flexibility to capture non-linear relationships through the first term $F^a(\cdot)$ in the model.

2.6 The loss function

The training process of NNs fits the model by iteratively updating weights through the back-propagation algorithm, i.e., calculating the gradient of the loss function and adjusting parameters to fit that gradient direction. The loss function, which quantifies the distance between the model predictions $\hat{y}$ and the true target values y during the training process, thus plays a pivotal

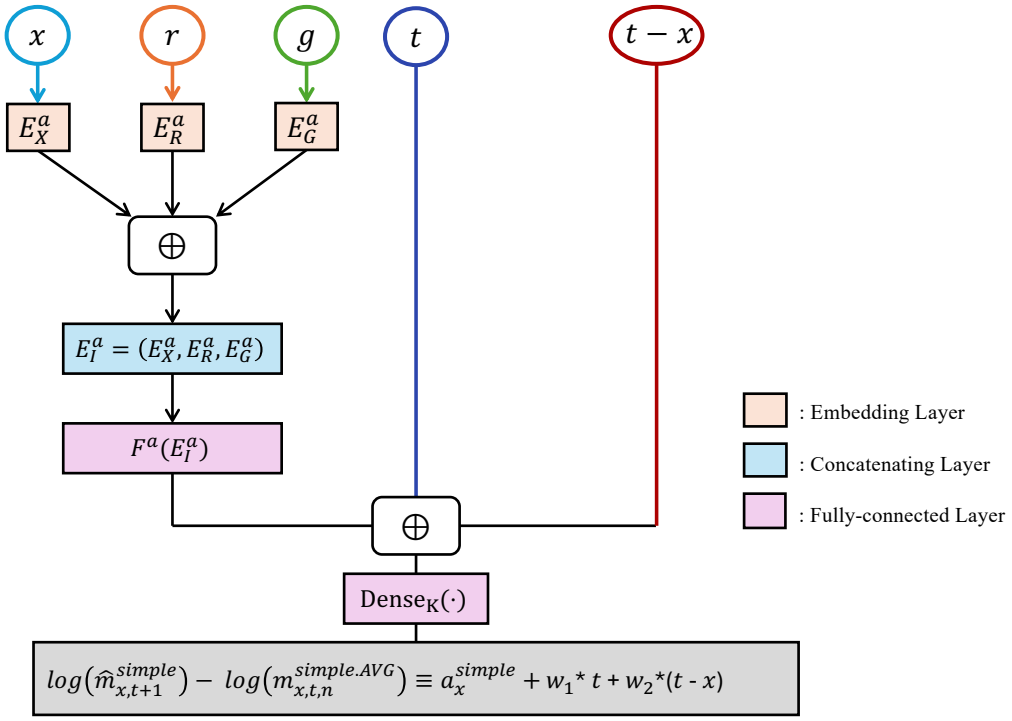


Figure 3: Schematic diagram of the architecture for the simpleNN model

role in guiding the optimization process and determining the final estimated parameter values.

In many applications, predictions for the mean of y are typically the primary focus and therefore the mean squared error (MSE) is typically adopted (e.g., Richman and Wüthrich, 2021). However, information about the entire distribution of y is often of high importance to actuaries to, for instance, price insurance products, calculate risk measures and capital reserves, construct hedging strategies, and perform scenario analysis and stress testing. In the traditional framework of stochastic mortality models, the theory of error analysis and simulation is well-established, allowing model uncertainties to be explicitly captured through methods such as time series models. In contrast, discussions on translating NN-based mortality predictions into actuarial decision-making processes remain limited.

To estimate the entire distribution of the true target values y, we borrow the idea from quantile regression and apply the pinball loss functions (Koenker and Hallock, 2001; Koenker and Bassett Jr, 1978) to the model fitting process. The pinball loss, also known as the quantile loss, can predict a specific quantile, by adjusting the loss calculation based on whether the

prediction underestimates or overestimates the actual value and penalizing underestimations and overestimations asymmetrically based on the target quantile. Therefore it is particularly useful to perform risk-sensitive tasks such as insurance rate making and regulatory reserve planning.

Mathematically, the pinball loss is defined as,

$$L(q, e_i) = \begin{cases} q \cdot |e_i| & \text{if } e_i \geq 0 \\ (1-q) \cdot |e_i| & \text{if } e_i < 0 \end{cases} \tag{10}$$

where q is the target quantile and $e_i = y_i - \hat{y}_i$ is the error for the i-th observation. With different specified values of q, we can obtain the estimate of each quantile of y.

Following the composite quantile regression framework (Zou and Yuan, 2008), we simultaneously estimate multiple quantiles using the sum of pinball losses across all K quantiles. However, in our experiments, we observed that: (1) median predictions can drift from the central tendency during joint optimization, and (2) quantile crossing frequently occurs, especially with a large number of quantiles (i.e., K is large), as neural networks have no built-in mechanism to preserve the natural ordering of quantiles. To address these challenges, we enhance the standard multi-quantile loss with two additional regularization terms:

$$L_{total} = L_{pinball} + L_{anchor} + L_{monotone}, \tag{11}$$

where

$$L_{pinball} = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} L(q_k, e_{i,k})$$

computes the average pinball loss across all K quantiles and N observations. Here q_k is the k-th quantile level, and $e_{i,k} = y_i - \hat{y}_i(q_k)$ is the error for the i-th observation y_i and the q_k-quantile predicted value.

The L_{anchor} term adds an L1 penalty specifically on the median prediction, providing additional stability to the central tendency estimation by leveraging the robustness properties of median regression (Koenker, 2005). The $L_{monotone}$ term prevents quantile crossing by penalizing violations of the natural ordering, inspired by the gradient penalty approach of Gupta et al.

(2019) but adapted for quantile monotonicity. Formally, these terms are defined as

$$L_{anchor} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_{i,0.5}| \tag{12}$$

$$L_{monotone} = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K-1} \max(0, \hat{y}_{i,q_k} - \hat{y}_{i,q_{k+1}}) \tag{13}$$

where $\hat{y}_{i,q_k}$ denotes the predicted value for sample i at quantile q_k.

Additionally, following Chernozhukov et al. (2010), we apply a post-prediction rearrangement step. Because our models generate all quantiles simultaneously, we enforce monotonicity by sorting the predicted values across quantiles for each age-time pair. This guarantees proper quantile ordering while preserving both computational efficiency and predictive performance. To the best of our knowledge, this work is among the first to introduce a composite loss framework that combines multi-quantile estimation with median anchoring and explicit monotonicity constraints in the context of neural network–based mortality forecasting. We believe this integrated approach highlights a novel direction for improving both the robustness and interpretability of mortality prediction models.

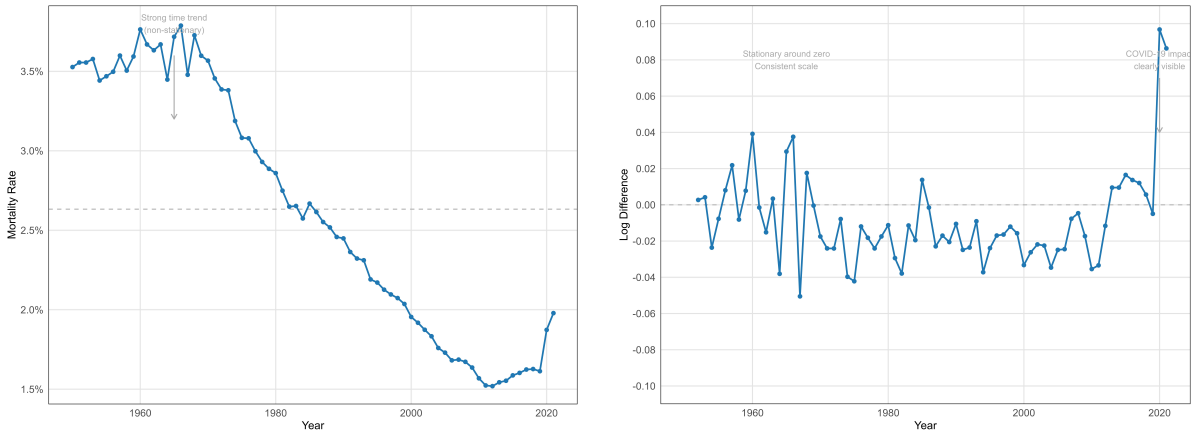
3 Empirical Analysis

3.1 Data and Evaluation Framework

Our study utilized mortality data of 24 countries/regions with complete records starting from 1950, obtained from the Human Mortality Database (HMD)[1]. This comprehensive dataset includes mortality rates for both male and female populations aged 55 to 90. Detailed data specifications for all populations are provided in Table 5 in Appendix A. Figure 4 illustrates the time trends of both raw mortality rates and their first order differences, which motivates us to consider the mortality improvement models introduced in Section 2.2.

In the model training process, temporal and demographic variables are normalized to facilitate efficient neural network optimization. The period 1950–1952 is employed to compute initial averages, with the year 1952 designated as the reference point ($t = 0$). Ages 55–90 are re-indexed on a scale from 0 to 35, and cohort values are standardized with respect to both

[1]Data source: www.mortality.org/. The data accessed were updated to April 2024.



(a) Mortality rates

(b) Mortality improvements

Figure 4: Time trend of mortality rates (left panel) and improvements (right panel) for the USA male population aged 65

the reference year and the reference age. This normalization procedure enhances numerical stability during training while maintaining the relative structure across variables. To ensure a rigorous evaluation, we adopt a temporal data-splitting strategy:

- **Training Period (1950–1999):** Models are estimated using data from all included countries. Within this period, an 80–20 random split is applied to separate training and validation sets, facilitating hyperparameter optimization and reducing the risk of overfitting.
- **Test Period (2000 onward):** Out-of-sample performance is assessed on data from 2000 to the most recent available year for each country. This fully held-out test set provides an unbiased measure of predictive accuracy.

We evaluate the accuracy of mean predictions using the mean absolute percentage error (MAPE) as the metric, defined as:

$$\text{MAPE}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \left| \frac{y_{j,i} - \hat{y}_{j,i}}{y_{j,i}} \right|,$$

where $y_{j,i}$ is the observed mortality and $\hat{y}_{j,i}$ is the predicted value, with j being the population index, and N_j is the total number of predictions for that population.

3.2 Model implementation

Our neural network models are implemented using TensorFlow 2.10. The optimization follows standard deep learning practices with specific adaptations for multi-quantile mortality forecasting. Specifically,

Quantile specification We choose five representative quantiles (2.5%, 25%, 50%, 75%, 97.5%), to capture the extreme tails, quartiles, and median of the mortality distribution. This specification balances computational efficiency with comprehensive coverage of mortality scenarios relevant for actuarial applications. Robustness of this choice is examined in Section 3.5, where we test alternative quantile granularities.

Optimization and Training We optimize the networks using backpropagation (Goodfellow et al., 2016) with Adam optimizer (Kingma and Ba, 2015) and standard hyperparameters: a learning rate of 0.001, a batch size of 32, and ReduceLROnPlateau scheduling. Training is conducted for up to 500 epochs with early stopping applied to prevent overfitting.

Hyperparameter Selection We conduct systematic grid search over dropout rates (0.4, 0.5, 0.6) and early stopping patience values (20, 35, 50 epochs). These relatively high dropout rates address the overfitting risks inherent in multi-quantile regression. In particular, when the number of quantiles K is very large (e.g., 51 or 101), the model must simultaneously learn complex distributional mappings while maintaining monotonicity. To ensure a robust performance evaluation, we employ an ensemble approach: each model configuration is trained 10 times to account for variability from random initialization and stochastic dropout, and all reported results are averaged across these 10 runs. Model selection for each architecture (SimpleNN, LCNN, RHNN) is based on validation loss, while the reported test MAPEs reflect out-of-sample performance on 2000–2023 data.

3.3 Performance Comparison

We compare the empirical performance of seven models including Lee-Carter (LC), Cairns-Blake-Dowd (CBD), Renshaw-Haberman (RH), Age-Period-Cohort (APC), Lee-Carter Neural Network (LCNN), Renshaw-Haberman Neural Network (RHNN), and simple Neural Network

(simpleNN). For each model, we compute the mean and standard deviation of MAPE in the test samples across all populations, and Table 1 presents a summary of the performance comparison among these prediction models. For all neural network models, we select the hyperparameter configurations that achieve the lowest validation loss through grid search: dropout $= 0.4$ with patience $= 35$ for LCNN and RHNN, and dropout $= 0.4$ with patience $= 20$ for SimpleNN. These configurations are identified from a grid search over dropout rates $(0.4, 0.5, 0.6)$ and patience values $(20, 35, 50)$, with the detailed validation results provided in Appendix B.

Table 1: Performance summary of different prediction models.

	LCNN	RHNN	SimpleNN	CBD	LC	RH	APC
Average MAPE	14.46%	12.54%	10.04%	18.81%	20.04%	34.30%	17.61%
Std.Dev MAPE	8.20%	7.64%	5.52%	6.96%	10.68%	43.13%	10.55%

Note: This table summarizes the average and standard deviation (Std. Dev) of MAPE across different populations.

Table 1 presents the performance summary of different prediction models. We observe that traditional stochastic models like LC and RH exhibit significantly higher average MAPEs of 20.04% and 34.30%, respectively, compared to their NN-based counterparts. Notably, the RH model demonstrates the highest variability in predictions, with a standard deviation of 43.13%. This extreme variability likely stems from non-convergence issues, highlighting the inherent instability of the traditional RH model under varying conditions. While the CBD and APC models outperform both the LC and RH models with average MAPEs of 18.81% and 17.61% respectively, they still underperform all three NN-based models in terms of prediction accuracy. SimpleNN achieves the best overall performance with both the lowest average MAPE (10.04%) and the smallest standard deviation (5.52%) among all models.

The neural network adaptations, LCNN and RHNN, achieve average MAPEs of 14.46% and 12.54% respectively, with standard deviations of 8.20% and 7.64%. These improvements over their traditional counterparts are substantial. Both LCNN and RHNN retain the linear representation (1), preserving the interpretability that facilitates better understanding and trust in promoting these more advanced models. The RHNN particularly benefits from neural network implementation, which eliminates the convergence problems while maintaining model interpretability.

A more detailed breakdown of model performance across 48 populations is provided in Table 2, and the results are consistent with our earlier observations. Overall, the proposed neural network frameworks (LCNN, RHNN, and SimpleNN) consistently outperform traditional stochastic mortality models (CBD, LC, RH, APC). Among them, SimpleNN achieves the best performance in 25 populations, followed by RHNN, which leads in 11 populations. It is worth noting that the traditional RH model produces extreme MAPE values exceeding 100% in several populations, thereby reinforcing the convergence issues discussed earlier and underscoring its limitations for practical implementation. In contrast, the enhanced RHNN framework yields markedly improved performance—substantially more stable, less variable across populations, and achieving the best results in multiple cases. Importantly, SimpleNN, despite its deliberately streamlined structure, remains highly competitive and frequently surpasses classical models, highlighting the strength of our proposed architectures even in their most parsimonious forms.

3.4 Model Interpretability

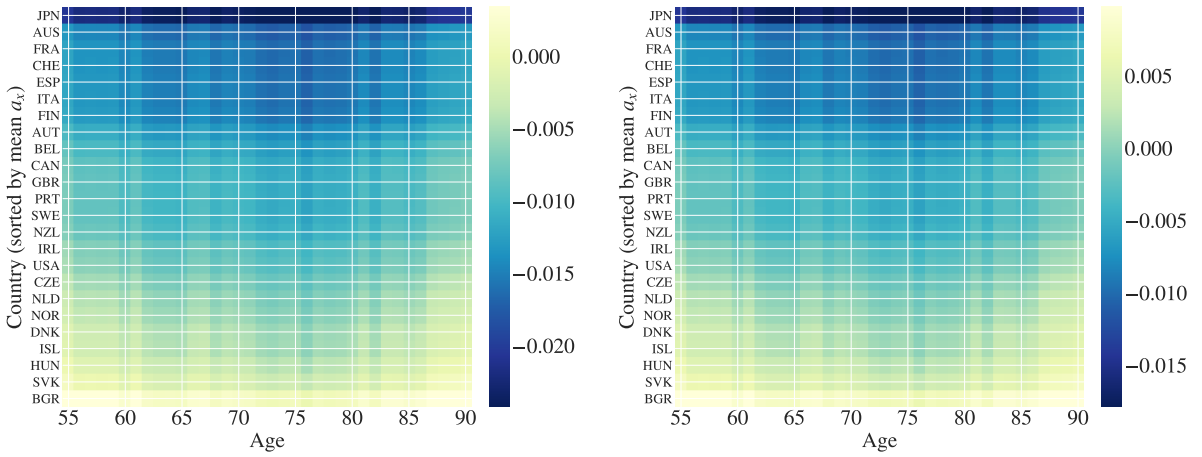
Developed within the same framework (1), all three NN-based models share the same basic architecture for age-specific mortality change patterns a_x, resulting in consistently similar negative values across countries. Recall that all three models model mortality changes rather than their levels, explaining the prevalence of negative values that represent mortality improvements. The models, as shown in Figure 5, unanimously identify Japan as experiencing the largest mortality improvements (the most negative a_x values), followed by other developed countries like Australia, France, and Switzerland, while a few Eastern European countries (e.g., Bulgaria, Hungary and Slovakia) show the slowest improvement rates.[2] This remarkable consistency across different neural network architectures validates the robustness of the approach.

Figure 6 shows that LCNN incorporates temporal variations through the period effect k_t^{LCNN}, with oscillating patterns capturing the non-uniform pace of mortality improvements over 1952-1999. These temporal dynamics represent baseline improvement rates, highlighting phases of accelerated progress (e.g., the 1960s) as well as temporary slowdowns (e.g., the 1990s). RHNN extends the framework by decomposing temporal variation into period effects k_t^{RHNN} and cohort effects γ_{t-x}^{RHNN}. The cohort parameters capture persistent impacts of early-life conditions,

[2]The estimated a_x values for LCNN and RHNN models show very similar patterns, and thus are provided in Appendix C.

Table 2: Performance comparison across different mortality models

Abbr.	Gender	LCNN	RHNN	SimpleNN	CBD	LC	RH	APC
AUS	Female	6.76%	**5.22%**	6.33%	12.17%	7.74%	8.94%	7.13%
AUS	Male	15.88%	13.02%	8.58%	14.71%	17.58%	17.11%	**7.93%**
AUT	Female	9.60%	**8.33%**	8.74%	18.94%	11.72%	12.65%	11.35%
AUT	Male	14.23%	11.29%	**7.78%**	14.78%	18.66%	12.83%	14.98%
BEL	Female	8.24%	**8.06%**	11.01%	18.03%	9.87%	72.46%	9.88%
BEL	Male	13.86%	10.95%	**7.01%**	16.21%	19.31%	11.24%	11.88%
BGR	Female	13.44%	11.45%	**9.45%**	20.02%	21.26%	32.41%	18.29%
BGR	Male	9.24%	8.64%	**8.15%**	13.71%	18.15%	83.43%	22.58%
CAN	Female	4.76%	**4.29%**	6.36%	10.79%	5.93%	91.29%	6.63%
CAN	Male	17.33%	14.57%	8.67%	18.51%	21.13%	7.95%	**7.45%**
CHE	Female	**6.08%**	6.49%	10.20%	17.82%	7.27%	83.91%	9.45%
CHE	Male	16.36%	13.57%	7.23%	15.07%	18.16%	**6.64%**	10.07%
CZE	Female	12.83%	10.55%	**6.97%**	14.95%	16.17%	17.25%	16.64%
CZE	Male	14.89%	12.84%	**7.45%**	17.78%	23.39%	25.47%	23.97%
DNK	Female	18.08%	15.66%	9.69%	18.78%	24.76%	**9.51%**	28.12%
DNK	Male	23.30%	20.23%	**11.58%**	26.62%	32.07%	12.20%	26.47%
ESP	Female	10.15%	**8.10%**	8.79%	19.63%	12.74%	26.26%	11.18%
ESP	Male	9.27%	7.14%	**5.66%**	13.89%	15.53%	86.60%	13.84%
FIN	Female	11.75%	**10.41%**	11.43%	20.04%	12.96%	17.11%	11.35%
FIN	Male	15.63%	12.82%	8.19%	16.17%	17.73%	9.15%	**7.87%**
FRA	Female	8.40%	**7.77%**	9.86%	20.43%	11.18%	21.28%	11.66%
FRA	Male	10.36%	7.61%	**6.93%**	15.76%	14.74%	11.67%	12.75%
GBR	Female	9.78%	7.49%	**6.54%**	12.13%	19.67%	7.23%	13.41%
GBR	Male	21.10%	17.97%	**11.40%**	20.66%	23.98%	261.44%	13.72%
HUN	Female	9.21%	7.26%	**6.86%**	17.05%	17.63%	35.13%	26.45%
HUN	Male	11.13%	9.02%	**4.90%**	20.55%	35.66%	54.21%	55.78%
IRL	Female	24.97%	22.28%	**15.03%**	26.52%	33.69%	58.07%	18.49%
IRL	Male	38.90%	35.67%	26.67%	47.37%	51.41%	**20.31%**	38.81%
ISL	Female	34.33%	33.68%	**30.80%**	31.69%	44.72%	43.36%	41.51%
ISL	Male	38.58%	36.21%	**30.58%**	32.05%	40.82%	32.30%	39.35%
ITA	Female	7.19%	**5.60%**	7.42%	13.62%	6.27%	12.56%	9.12%
ITA	Male	15.84%	13.48%	8.73%	18.69%	23.34%	**7.22%**	16.41%
JPN	Female	6.08%	**5.37%**	8.89%	16.71%	16.16%	10.10%	6.70%
JPN	Male	4.12%	**3.93%**	8.28%	8.25%	6.97%	15.38%	11.87%
NLD	Female	6.98%	**5.42%**	7.18%	14.19%	7.37%	17.04%	11.87%
NLD	Male	24.49%	21.19%	13.43%	31.10%	35.85%	**10.23%**	21.76%
NOR	Female	10.49%	8.95%	**6.39%**	14.38%	13.03%	14.47%	20.23%
NOR	Male	27.42%	23.97%	**15.22%**	31.97%	38.27%	19.51%	29.01%
NZL	Female	12.51%	11.07%	**8.67%**	14.53%	18.38%	13.54%	19.58%
NZL	Male	24.53%	21.71%	14.19%	24.01%	27.90%	**12.14%**	16.43%
PRT	Female	17.96%	14.83%	**9.42%**	22.32%	20.46%	105.19%	19.76%
PRT	Male	16.10%	12.76%	**9.21%**	19.93%	22.38%	104.53%	21.11%
SVK	Female	12.75%	10.78%	**6.81%**	14.42%	16.15%	29.24%	16.59%
SVK	Male	14.24%	12.27%	**6.63%**	23.14%	34.91%	29.61%	36.36%
SWE	Female	**4.82%**	6.04%	9.67%	13.44%	5.69%	6.00%	7.95%
SWE	Male	15.32%	12.49%	**5.96%**	15.28%	21.05%	15.96%	16.27%
USA	Female	**4.85%**	5.24%	9.74%	10.17%	8.40%	24.74%	6.70%
USA	Male	10.09%	7.98%	**7.42%**	14.02%	13.58%	9.49%	8.48%



(a) Female (b) Male

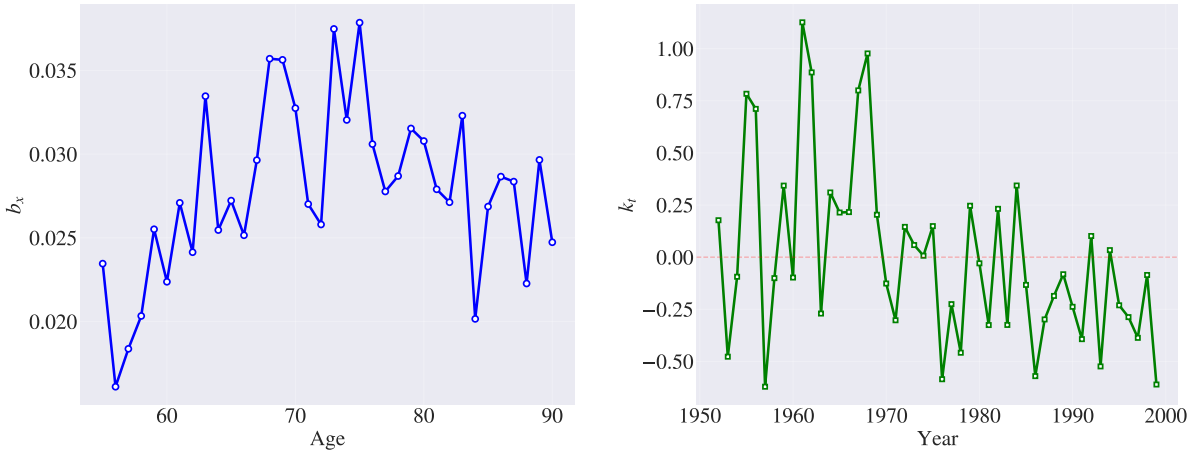
Figure 5: Heatmap of age-specific effects a_x by country with SimpleNN Model: Female (panel (a)) and Male (panel (b))

with birth cohorts exposed to major historical disruptions, such as the 1918-1920 pandemic or World Wars, exhibiting distinct mortality trajectories. Once cohort effects are incorporated, we observe in Figure 7 that the period effects k_t^{RHNN} in RHNN appear more stationary than in LCNN, indicating that much of the temporal trend in simpler models in fact reflects underlying cohort heterogeneity. Finally, the age-specific sensitivity coefficients, b_x^{LCNN} and b_x^{RHNN}, and η_x in RHNN, exhibit reasonable patterns across ages, consistent with demographic expectations.

3.5 Robustness Analysis

In the experiments in Section 3.3, we employ five quantiles (2.5%, 25%, 50%, 75%, 97.5%) in all NN-based models, representing the extreme tails, quartiles, and median of the mortality distribution. This specification balances computational efficiency with comprehensive coverage of mortality scenarios relevant for actuarial applications.

Because our simulation approach relies on interpolation between predicted quantiles (see Section 4), increasing the number of quantiles naturally yields a more accurate approximation of the full distribution. To establish a performance benchmark, we evaluate three levels of quantile granularity under the same training setup (dropout = 0.4, patience = 20): (1) 21 quantiles including 2.5%, 97.5%, and 19 evenly spaced quantiles from 5% to 95% in increments of 5%; (2) 51 quantiles consisting all odd percentiles from 1% to 99% in increments of 2%



(a) b_x^{LCNN} (b) k_t^{LCNN}

Figure 6: Estimated parameters of LCNN model: age modulation of time trend b_x^{LCNN} (panel (a)), and time-varying mortality index k_t^{LCNN} (panel (b)).

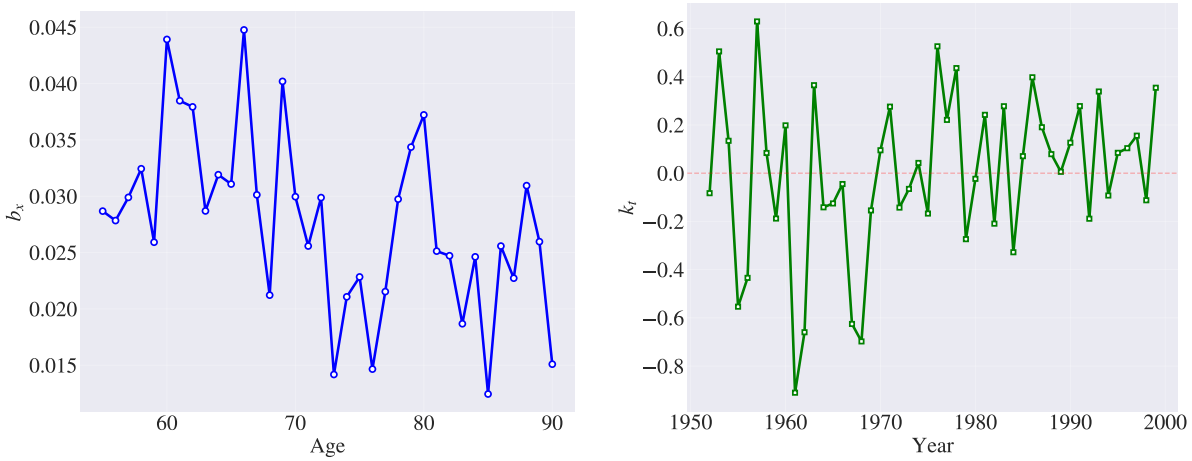
together with the median at 50%; and (3) 101 quantiles including 0.5%, 99.5%, and all integer percentiles from 1% to 99%.

Table 3 summarizes the results. We can see that different quantile configurations yield similar MAPE values of 9.74%, 9.82%, and 9.85%, demonstrating robust performance of our models. The population-level standard deviations also remain consistent, ranging from 5.2% to 5.8%. Since finer granularities yield only marginal gains, we adopt the 5-quantile specification in our main analysis to balance computational efficiency with accuracy.

Table 3: Performance across different quantile granularities

Quantile Specifications	Average MAPE	Std.Dev MAPE
(2.5%, 5%, 10%, . . . , 90%, 95%, 97.5%)	9.74%	5.20%
(1%, 3%, . . . , 49%, 50%, 51%, . . . 97%, 99%)	9.82%	5.76%
(0.5%, 1%, 2%, . . ., 98%, 99%, 99.5%)	9.85%	5.79%

To further validate the design of our composite loss function, we conduct an ablation study to examine the contribution of each loss component. Table 4 compares four configurations across different quantile specifications. The results demonstrate that the full composite loss consistently achieves the lowest MAPE across all quantile specifications, reducing error by approximately 2 -2.5 percentage points compared to the baseline pinball loss. While individual components show varying effectiveness across different quantile counts, their combination



(a) b_x^{RHNN} (b) k_t^{RHNN}

(c) η_x^{RHNN} (d) γ_{t-x}^{RHNN}

Figure 7: Estimated parameters of RHNN model: age modulation of time trend b_x^{RHNN} (panel (a)), time-varying mortality index k_t^{RHNN} (panel (b)), age modulation of cohort effect η_x^{RHNN}, and cohort effects γ_{t-x}^{RHNN}.

yields superior performance across all settings, achieving MAPE values below 10% consistently. These findings validate the importance of jointly addressing median stability and quantile ordering in multi-quantile optimization. The interaction effects between loss components across different quantile granularities present interesting avenues for future research.

Table 4: Results with different combinations of loss function components.

Loss Configuration	MAPE			
	K=5	K=21	K=51	K=101
Pinball only	12.13% (6.53%)	12.27% (6.58%)	11.80% (6.41%)	12.04% (6.51%)
Pinball + Anchor	10.12% (5.61%)	10.33% (6.01%)	12.13% (6.56%)	12.39% (6.64%)
Pinball + Monotone	11.38% (6.23%)	11.58% (6.33%)	11.88% (6.44%)	12.43% (6.62%)
Pinball + Anchor + Monotone	**10.04% (5.52%)**	**9.74% (5.20%)**	**9.82% (5.76%)**	**9.85% (5.79%)**

Note: Values in parentheses represent standard deviations across populations.

3.6 Uncertainty Quantification

In previous sections, we evaluated the predictive accuracy of the proposed models. For actuarial applications such as pricing and reserving, it is equally important to quantify the uncertainty surrounding forecasts, as decision-making often depends not only on point estimates but also on the reliability of the distributional outputs. To assess model uncertainty, we construct bootstrap confidence intervals based on 100 resampled training sets. Figure 8 presents the resulting 95% confidence intervals for all five quantiles and both genders. The intervals are consistently narrow across the entire range, indicating limited variability in the predictions. The stability of the bootstrap procedure is evident from the smooth trajectories of the bootstrap medians and the narrow confidence bands, suggesting that the model produces reliable predictions.

4 Longevity-linked derivative pricing

4.1 Pricing framework

In this section, we demonstrate the benefit of the proposed interpretable NN and simulation framework from the Methodology Section by applying it to longevity-linked derivative pricing. The standard approach of derivative pricing involves defining a risk-neutral probability measure to transform physical probabilities to risk-neutral probabilities, which implicitly accounts for

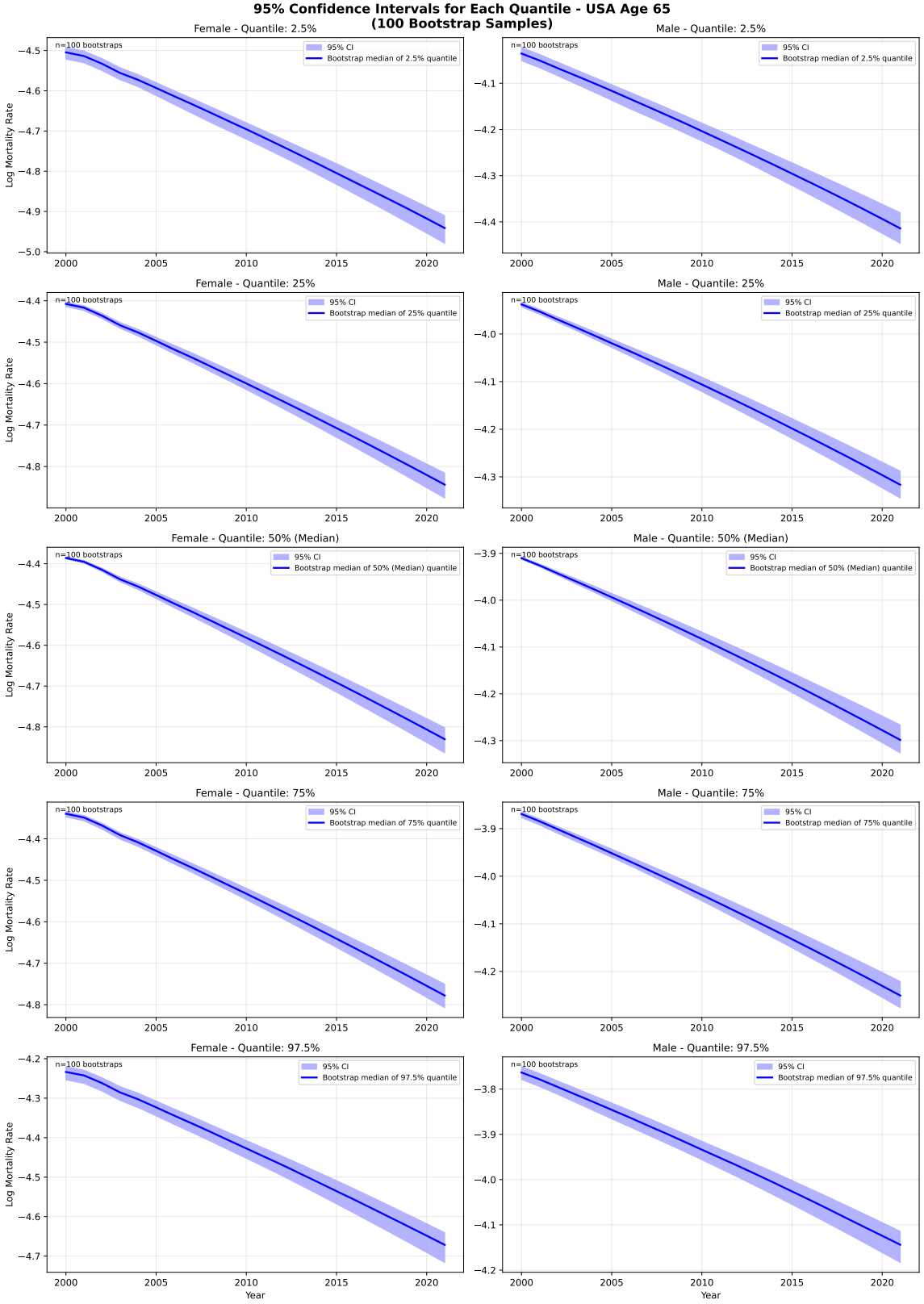


Figure 8: 95% bootstrap confidence intervals for all five quantiles - USA age 65

the market price of risk, and discounting future cash flows using the risk-free rate (e.g., Harrison and Kreps, 1979; Harrison and Pliska, 1981, 1983). Since mortality rates themselves are not directly tradable, the market for mortality-linked derivatives is likely incomplete. Consequently, there may exist multiple risk-neutral probability measures, resulting in a range of reasonable prices. In the actuarial domain, Gerber and Shiu (1994) initially propose the application of the Esscher transform for pricing derivative products in incomplete markets. Building on this, Wang (1996, 2000, 2002) introduce a methodology for transitioning from the physical probability measure $\mathbb{P}$ to a risk-neutral measure $\mathbb{Q}$ using the Wang transform, which later becomes the standard approach for pricing insurance-linked securities (e.g., Lin and Cox, 2005; Cox and Lin, 2007; Denuit et al., 2007; Huang et al., 2023, 2024). Mathematically it is defined as,

$$F^{\mathbb{Q}}(x) = \Phi\left[\Phi^{-1}(F^{\mathbb{P}}(x)) + \lambda\right], \tag{14}$$

where λ denotes the market price of risk; $F^{\mathbb{P}}(\cdot)$ and $F^{\mathbb{Q}}(\cdot)$ represent the cumulative distribution functions (CDF) of the same random variable under the physical probability measure $\mathbb{P}$ and the risk-neutral measure $\mathbb{Q}$ respectively; and $\Phi(\cdot)$ is the CDF of the standard normal distribution. In reality, the value of λ is determined by the existing derivatives written on the same underlying asset to preclude arbitrage opportunities. Equation (14) also highlights the importance of estimating the entire distribution of mortality rates in actuarial applications, as it directly influences derivative pricing and risk assessment.

We assume that the mortality rate for an individual aged x at time t remains constant over a one-year period, making the force of mortality equal to the central death rate $m_{x,t}$. Under this assumption, the survival probability under $\mathbb{P}$ that an individual aged x at time t will survive another N years is expressed by,

$$p_x(t, N) = \exp(-S_N), \tag{15}$$

where $S_N = \sum_{j=0}^{N-1} m_{x+j,t+j}$ represents the cumulative force of mortality over the survival period for the cohort. The cumulative probability function $F^{\mathbb{P}}(s)$ of $p_x(t, N)$ under the real probability measure $\mathbb{P}$ is defined as:

$$F^{\mathbb{P}}(s) = \mathbb{P}[p_x(t, N) \leq s]. \tag{16}$$

Then under the Wang risk-neutral measure $\mathbb{Q}$, the expected value of $p_x(t, N)$ can be expressed as (Denuit et al., 2007):

$$\mathbf{E}^{\mathbb{Q}}[p_x(t,N)] = \int_0^1 \left[1 - F^{\mathbb{Q}}(s)\right] ds = \int_0^1 \left[1 - \Phi\left[\Phi^{-1}\left(F^{\mathbb{P}}(s)\right) + \lambda\right]\right] ds. \tag{17}$$

4.2 Simulation of future mortality

Stochastic mortality models do not typically provide explicit expressions for the distribution function, so we employ the Monte Carlo simulation method to estimate the expected value of survival probabilities under $\mathbb{Q}$, as given by Equation (17) (Wang et al., 2015; Kung et al., 2021). For this analysis we select the simpleNN model proposed in Section 2.5 although the procedures for the other two models are very similar. The simulation process is summarized as follows:

Step 1: Fit the simpleNN model based on the optimal configuration identified by validation, as described in Section 3.2.

Step 2: Generate $M = 100{,}000$ future mortality paths for a specific population cohort over the tenor of the financial instrument. For each path, randomly sample from the predicted quantiles and construct $F^{\mathbb{P}}(\cdot)$, the CDF of survival probability under $\mathbb{P}$, based on (16) using the neural network predictions and their interpolated values.

Step 3: Apply the Wang Transform (14) to $F^{\mathbb{P}}(\cdot)$ to obtain $F^{\mathbb{Q}}(\cdot)$, and generate the corresponding survival probabilities under $\mathbb{Q}$ for each of the M paths.

Step 4: Apply (17) to calculate the relevant $\mathbb{Q}$-expected values for derivative pricing based on all M simulation paths.

4.3 Simulation and pricing results

We consider the pricing of a customized longevity swap issued at date t_0, whose underlying mortality rates are linked to future survival rates of a cohort aged x at date t. Suppose an insurer seeks to hedge against future mortality fluctuations in its portfolio, and therefore purchases a longevity swap from a counterparty with the following (per share) payouts. At the end of each year $t = t_0 + 1, t_0 + 2, ..., t_0 + N$,

Floating leg: The issuer of the swap will pay to its counterparty a random amount denoted as $c^{Float}(t) := p_x(t,N) - p_x^{ref}(t,N)$ as defined by (15), where $p_x^{ref}(t,N)$ represents a reference survival rate;

Fixed leg: In return, the issuer will receive a pre-determined fixed amount denoted by c^{Fix}, from its counterparty.

By the no-arbitrage pricing argument, c^{Fix} is determined such that,

$$\mathbf{E}^{\mathbb{Q}}\left\{\sum_{t=1}^{N} B(t_0, t_0+t)\left[c^{Float}(t) - c^{Fix}\right]\Big|\mathcal{F}_{t_0}\right\} = 0, \tag{18}$$

where $B(t_0, t_0+t)$ is the date-t_0 price of a zero-coupon discount bond maturing at t_0+t, and $\mathcal{F}_{t_0}$ represents all market information available up to time t_0. Equation (18) implies that,

$$c^{Fix} = \frac{\sum_{t=1}^{N} B(t_0, t_0+t)\mathbf{E}^{\mathbb{Q}}\left[c^{Float}(t)|\mathcal{F}_{t_0}\right]}{\sum_{t=1}^{N} B(t_0, t_0+t)}, \tag{19}$$

where the value of c^{Fix} is estimated through the Monte Carlo scheme discussed in Section 4.2.

In our baseline analysis, we specifically consider the U.S. males aged 65 at $t_0 = 2021$ as the reference population for the longevity swap. We also assume a tenor of $N = 25$ years and use the discounting factor $B(t_0, t_0+t)$ based on the U.S. Treasury yield curve data.[3] Figure 9 exhibits the projected mortality rate from the inception to the maturity of the considered swap. We observe that the average mortality increases over time as the cohort ages. In addition, the variations in the simulated mortality rates also widen over time, reflecting higher uncertainties in projections for more distant future periods.

Next, we examine the risk premium of the longevity swap under varying values of λ, compared to the baseline scenario where $\lambda = 0$. Following the parameter choices in Wang et al. (2013) and Kung et al. (2021), we adopt $\lambda \in \{-0.2, -0.15, -0.1\}$ and various terms of the longevity swap up to 25 years. In practice, the λ values represent risk preferences of the hedger, with more negative values indicating greater risk aversion. These values can be calibrated from existing derivatives in the market with the same underlying asset. As shown in Figure 10, all three cases produce a convex upward-curving trend in relation to the swap's tenor, indicating

[3]Data source: www.treasury.gov/resource-center/data-chart-center/interest-rates. The yield curve data as of December 31, 2021, were used.

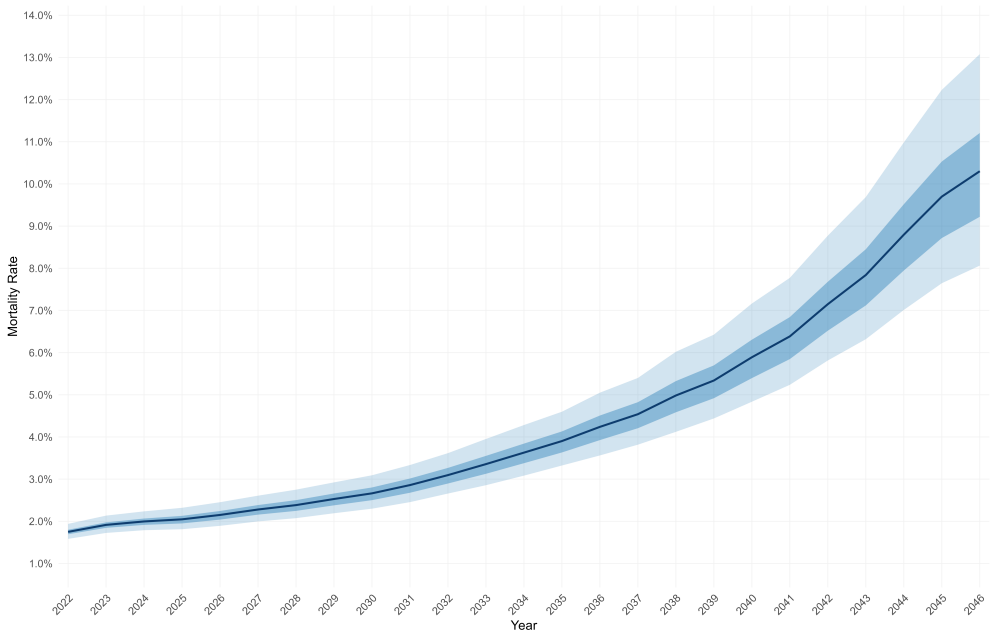


Figure 9: Mortality projections for U.S. males aged 65 from 2022 to 2046. The light-shaded region denotes the 90% projection interval, the dark-shaded region denotes the 50% interval, and the central curve indicates the mean projection.

the increased uncertainty associated with long-term mortality predictions. In addition, as one would expect, more risk-averse insurers are willing to pay higher risk premiums for the swap.

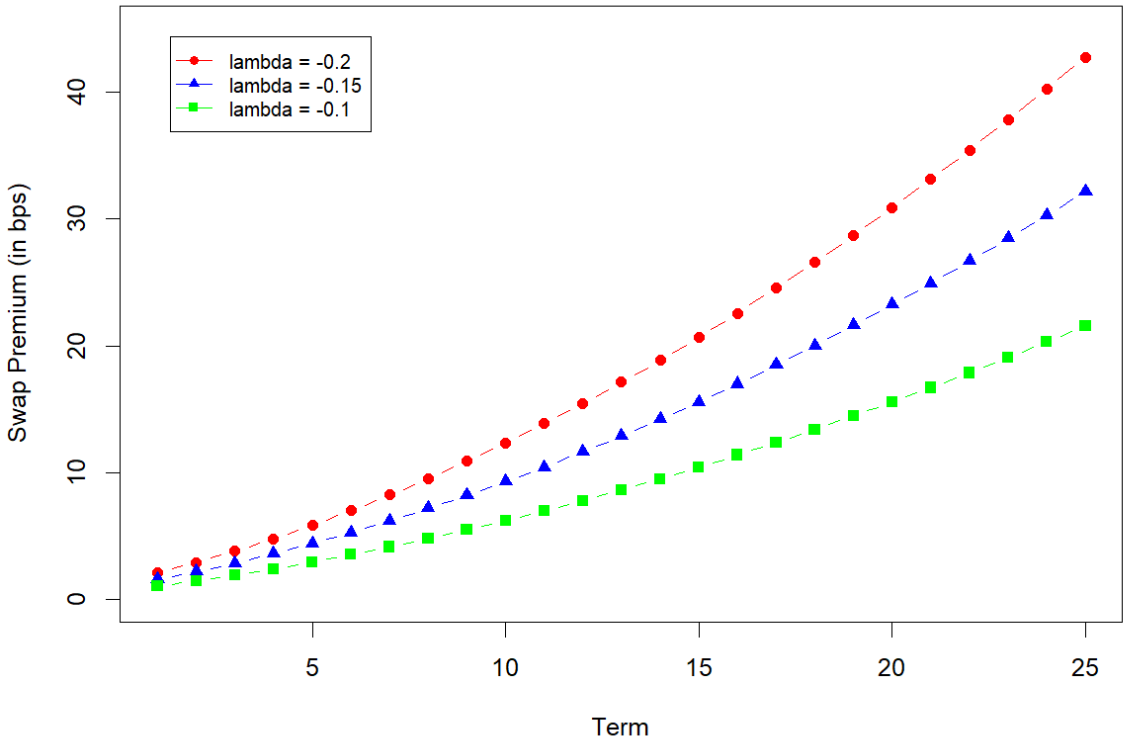
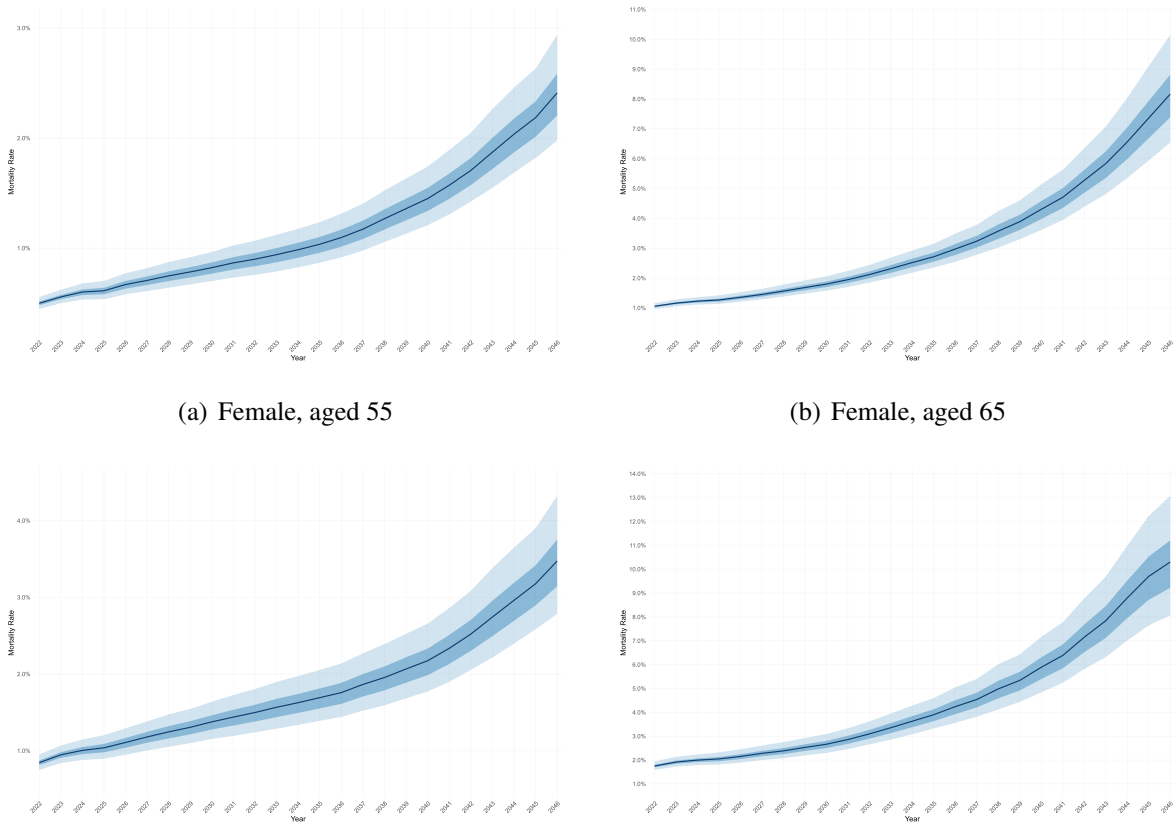


Figure 10: Swap premium curves estimated by the simpleNN model

4.4 Robustness checks

We finally perform some robustness checks on the underlying population to validate the results presented in 4.3. Specifically, in addition to the US male population aged 65 as of 2021, we also consider the U.S. females aged 55 and 65, as well as males aged 55 in 2021, repeating our analysis for each cohort. Figure 11 displays the mortality projections, while Figure 12 displays the risk premiums for each cohort considered.

We have several interesting observations from Figure 11. First, expected death rates increase as individuals age across all groups, with males consistently showing higher rates than females within the same age cohort, reflecting well-documented gender differences in mortality. Second,



(a) Female, aged 55

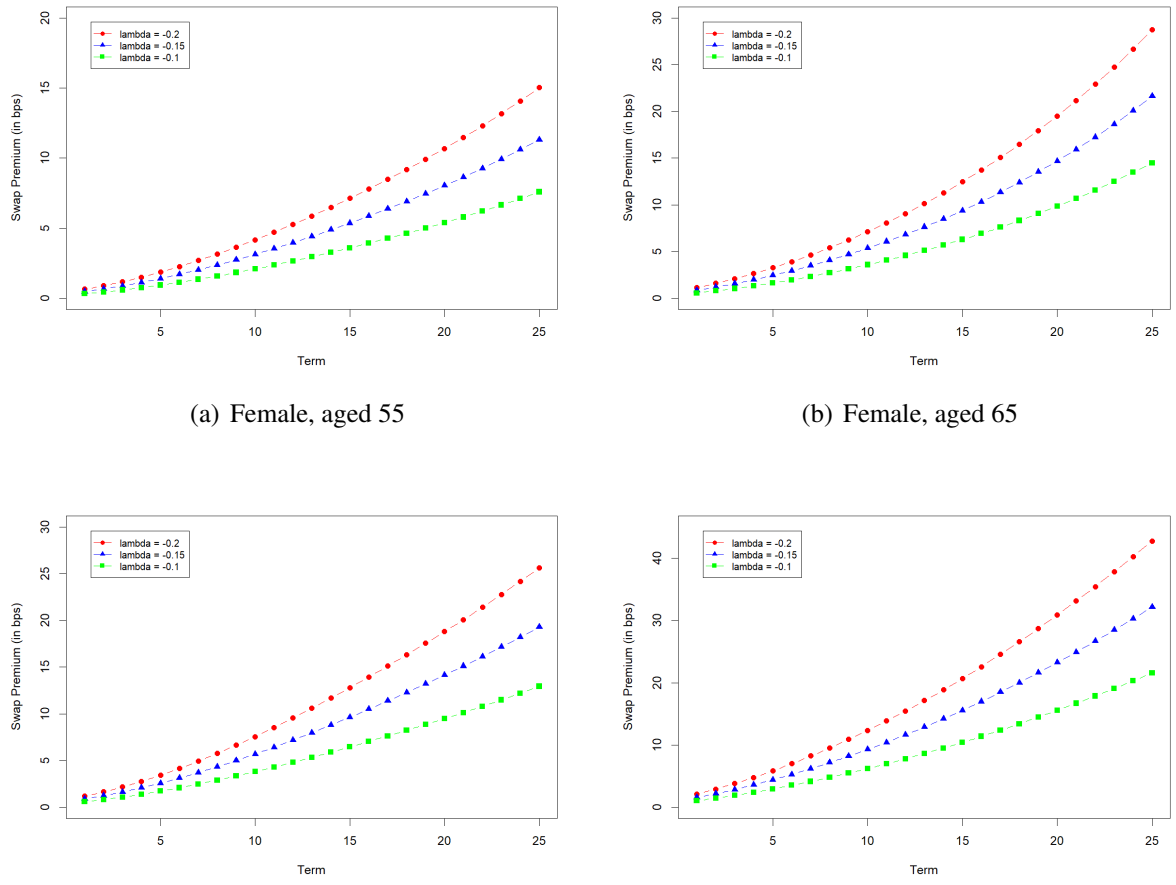
(b) Female, aged 65

(c) Male, aged 55

(d) Male, aged 65

Figure 11: Mortality projections for U.S. females aged 55 (panel (a)), females aged 65 (panel (b)), males aged 55 (panel (c)) and males aged 65 (panel (d)), from 2022 to 2046. The light-shaded region denotes the 90% projection interval, the dark-shaded region denotes the 50% interval, and the central curve indicates the mean projection.

the widening confidence intervals across all groups indicate increasing uncertainty in long-term mortality projections. Lastly, although males consistently exhibits higher predicted mortality rates compared to females, there are no significant differences in the uncertainty of predictions between genders. These trends highlight the importance of actuarial simulation in improving mortality modeling and understanding long-term trends.



(a) Female, aged 55

(b) Female, aged 65

(c) Male, aged 55

(d) Male, aged 65

Figure 12: Swap premium curves for U.S. females aged 55 (panel (a)), females aged 65 (panel (b)), males aged 55 (panel (c)) and males aged 65 (panel (d))

Figure 12 also reveals similar patterns in the longevity swap premium across different underlying populations. In particular, across all panels, the swap premium increases with the term length, reflecting the higher longevity risk associated with longer swap terms. In addition, higher absolute values of λ represents greater risk aversion and hence corresponds to greater

swap premiums. Comparing the demographic groups, it is observed that the longevity swaps written on older (male) populations tend to be more expensive than those written on younger (female) populations. These observed trends underscore the important interplay between different factors such as age, gender, contract term, and market price of risk in determining the pricing of longevity swaps.

5 Conclusion

This paper introduces three innovative quantile-based neural network models namely LCNN, RHNN, and simpleNN, which aim to strike a balance between model interpretability and predictive accuracy. Similar to classic stochastic mortality models, these models feature interpretable linear representations with transparently observable parameters. At the same time, numerical analysis based on mortality data from 24 countries demonstrates that these NN-based models significantly outperform their traditional counterpart in terms of both the mean and variance of MAPE.

A key innovation of our approach lies in directly modeling quantiles of mortality rates through the adoption of the pinball loss function during the learning process. This is complemented by a simulation scheme that interpolates these quantiles to generate future mortality uncertainties. This provides an efficient tool for producing distributional information on future mortality scenarios, which is highly useful for various actuarial tasks.

Finally, the application of these models to derivative pricing using the Wang transform highlights their value in efficiently characterizing probabilistic distribution functions of future predictions. This application offers important insights into the risk-neutral pricing of longevity derivatives, further highlighting the practical relevance and versatility of the proposed approach.

Acknowledgement The authors would like to thank the editors and the anonymous reviewers for their valuable comments, which have greatly improved an earlier version of the manuscript. Zhang acknowledges the research funding support from the Singapore Ministry of Education Academic Research Fund Tier 1 Grant (RG49/24). Zhu acknowledges the research funding support from the Singapore Ministry of Education Academic Research Fund Tier 3 (MOE-MOET32022-0006) and CyberSG R&D Programme Office (CRPO) Translation and Innovation

Grant (CRPO-GC4-SMU-002). Wang acknowledges support from the National Science and Technology Council (NSTC) under grant number 113-2410-H-110-005-MY3. Qiao acknowledges the support of the Global-Asia Insurance Partnership (GAIP) for his research fellowship.

References

Agarwal, R., Melnick, L., Frosst, N., Zhang, X., Lengerich, B., Caruana, R., and Hinton, G. E. (2021). Neural additive models: Interpretable machine learning with neural nets. *Advances in neural information processing systems*, 34:4699–4711.

Atance, D., Lledó, J., and Navarro, E. (2025). Modelling and forecasting mortality rates for a life insurance portfolio. *Risk Management*, 27(1):3.

Blake, D., Cairns, A., Coughlan, G., Dowd, K., and MacMinn, R. (2013). The new life market. *Journal of Risk and Insurance*, 80(3):501–558.

Bravo, J. M. (2021). Forecasting mortality rates with recurrent neural networks. *CAPSI 2021 Proceedings*, pages 1–19.

Cairns, A. J., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M. (2011a). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, 48(3):355–367.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, 13(1):1–35.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., and Khalaf-Allah, M. (2011b). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin: The Journal of the IAA*, 41(1):29–59.

Chen, H., MacMinn, R., and Sun, T. (2015). Multi-population mortality models: A factor copula approach. *Insurance: Mathematics and Economics*, 63:135–146.

Chen, H., MacMinn, R. D., and Sun, T. (2017). Mortality dependence and longevity bond pricing: a dynamic factor copula mortality model with the gas structure. *Journal of Risk and Insurance*, 84(S1):393–415.

Chen, Y. and Khaliq, A. Q. (2022). Comparative study of mortality rate prediction using data-driven recurrent neural networks and the Lee–Carter model. *Big Data and Cognitive Computing*, 6(4):134.

Chen, Z., Lu, Y., Zhang, J., and Zhu, W. (2024). Managing weather risk with a neural network-based index insurance. *Management Science*, 70(7):4306–4327.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.

Chuang, S.-L. and Brockett, P. L. (2014). Modeling and pricing longevity derivatives using stochastic mortality rates and the esscher transform. *North American Actuarial Journal*, 18(1):22–37.

Corsaro, S., Marino, Z., and Scognamiglio, S. (2024). Quantile mortality modelling of multiple populations via neural networks. *Insurance: Mathematics and Economics*, 116:114–133.

Cox, S. H. and Lin, Y. (2007). Natural hedging of life and annuity mortality risks. *North American Actuarial Journal*, 11(3):1–15.

Delong, Ł., Lindholm, M., and Wüthrich, M. V. (2022). Collective reserving using individual claims data. *Scandinavian Actuarial Journal*, 2022(1):1–28.

Denuit, M., Devolder, P., and Goderniaux, A.-C. (2007). Securitization of longevity risk: Pricing survivor bonds with wang transform in the lee-carter framework. *Journal of Risk and Insurance*, 74(1):87–113.

Dong, Y., Huang, F., Yu, H., and Haberman, S. (2020). Multi-population mortality forecasting using tensor decomposition. *Scandinavian Actuarial Journal*, 2020(8):754–775.

Dowd, K., Cairns, A. J., Blake, D., Coughlan, G. D., and Khalaf-Allah, M. (2011). A gravity model of mortality rates for two related populations. *North American Actuarial Journal*, 15(2):334–356.

Fissler, T., Merz, M., and Wüthrich, M. V. (2023). Deep quantile and deep composite triplet regression. *Insurance: Mathematics and Economics*, 109:94–112.

Gabrielli, A. and Wüthrich, M. V. (2018). An individual claims history simulation machine. *Risks*, 6(2):29.

Gao, G., Meng, S., and Wüthrich, M. V. (2019). Claims frequency modeling using telematics car driving data. *Scandinavian Actuarial Journal*, 2019(2):143–162.

Gao, G., Wang, H., and Wüthrich, M. V. (2022). Boosting poisson regression models with telematics car driving data. *Machine Learning*, pages 1–30.

Gao, G. and Wüthrich, M. V. (2018). Feature extraction from telematics car driving heatmaps. *European Actuarial Journal*, 8:383–406.

Gerber, H. U. and Shiu, E. S. (1994). Martingale approach to pricing perpetual american options. *ASTIN Bulletin: The Journal of the IAA*, 24(2):195–220.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge, MA.

Gupta, A., Shukla, N., Marla, L., Kolbeinsson, A., and Yellepeddi, K. (2019). How to incorporate monotonicity in deep networks while preserving flexibility? *arXiv preprint arXiv:1909.10662*. arXiv:1909.10662.

Haberman, S. and Renshaw, A. (2012). Parametric mortality improvement rate modelling and projecting. *Insurance: Mathematics and Economics*, 50(3):309–333.

Haberman, S. and Renshaw, A. (2013). Modelling and projecting mortality improvement rates using a cohort perspective. *Insurance: Mathematics and Economics*, 53(1):150–168.

Hainaut, D. (2018). A neural-network analyzer for mortality forecast. *ASTIN Bulletin: The Journal of the IAA*, 48(2):481–508.

Harrison, J. M. and Kreps, D. M. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20(3):381–408.

Harrison, J. M. and Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic processes and their applications*, 11(3):215–260.

Harrison, J. M. and Pliska, S. R. (1983). A stochastic calculus model of continuous trading: complete markets. *Stochastic processes and their applications*, 15(3):313–316.

Hsiao, H.-T., Wang, C.-W., Liu, I.-C., and Kung, K.-L. (2024). Mortality improvement neural-network models with autoregressive effects. *The Geneva Papers on Risk and Insurance-Issues and Practice*, pages 1–21.

Huang, S., Tan, K. S., Zhang, J., and Zhu, W. (2024). Epidemic financing facilities: Pandemic bonds and endemic swaps. *North American Actuarial Journal*, 28(3):626–657.

Huang, S., Zhang, J., and Zhu, W. (2023). Storm cat bond: Modeling and valuation. *North American Actuarial Journal*, pages 1–26.

Jarner, S. F. and Kryger, E. M. (2011). Modelling adult mortality in small populations: The saint model. *ASTIN Bulletin: The Journal of the IAA*, 41(2):377–418.

Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Koenker, R. (2005). Quantile regression. *Cambridge Univ Pr*.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of economic perspectives*, 15(4):143–156.

Koo, P. K. and Ploenzke, M. (2021). Improving representations of genomic sequence motifs in convolutional networks with exponential activations. *Nature machine intelligence*, 3(3):258–266.

Kung, K.-L., Liu, I.-C., and Wang, C.-W. (2021). Modeling and pricing longevity derivatives using skellam distribution. *Insurance: Mathematics and Economics*, 99:341–354.

Kuo, K. (2019). Deeptriangle: A deep learning approach to loss reserving. *Risks*, 7(3):97.

Laporta, A. G., Levantesi, S., and Petrella, L. (2024). Neural networks for quantile claim amount estimation: a quantile regression approach. *Annals of Actuarial Science*, 18(1):30–50.

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

Levantesi, S., Nigri, A., and Piscopo, G. (2022). Clustering-based simultaneous forecasting of life expectancy time series through long-short term memory neural networks. *International Journal of Approximate Reasoning*, 140:282–297.

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

Li, J. S.-H., Chan, W.-S., and Zhou, R. (2017). Semicoherent multipopulation mortality modeling: the impact on longevity risk securitization. *Journal of Risk and Insurance*, 84(3):1025–1065.

Li, J. S.-H. and Hardy, M. R. (2011). Measuring basis risk in longevity hedges. *North American Actuarial Journal*, 15(2):177.

Li, J. S.-H. and Liu, Y. (2020). The heat wave model for constructing two-dimensional mortality improvement scales with measures of uncertainty. *Insurance: Mathematics and Economics*, 93:1–26.

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, 42:575–594.

Lin, T., Wang, C.-W., and Tsai, C. C.-L. (2015). Age-specific copula-ar-garch mortality models. *Insurance: Mathematics and Economics*, 61:110–124.

Lin, Y. and Cox, S. H. (2005). Securitization of mortality risks in life annuities. *Journal of Risk and Insurance*, 72(2):227–252.

Lindholm, M. and Palmborg, L. (2022). Efficient use of data for lstm mortality forecasting. *European Actuarial Journal*, 12(2):749–778.

Marino, M., Levantesi, S., and Nigri, A. (2023). A neural approach to improve the Lee-Carter mortality density forecasts. *North American Actuarial Journal*, 27(1):148–165.

Mitchell, D., Brockett, P., Mendoza-Arriaga, R., and Muthuraman, K. (2013). Modeling and forecasting mortality rates. *Insurance: Mathematics and economics*, 52(2):275–285.

Miyata, A. and Matsuyama, N. (2022). Extending the Lee–Carter model with variational autoencoder: a fusion of neural network and bayesian approach. *ASTIN Bulletin: The Journal of the IAA*, 52(3):789–812.

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

Nigri, A., Levantesi, S., Marino, M., Scognamiglio, S., and Perla, F. (2019). A deep learning integrated Lee–Carter model. *Risks*, 7(1):33.

Noll, A., Salzmann, R., and Wüthrich, M. V. (2018). Case study: French motor third-party liability claims. *SSRN*.

Office of the Chief Actuary (2024). Mortality projections for social security programs in canada: Actuarial study no. 22. Technical report, Office of the Superintendent of Financial Institutions Canada.

Perla, F., Richman, R., Scognamiglio, S., and Wüthrich, M. V. (2021). Time-series forecasting of mortality rates using deep learning. *Scandinavian Actuarial Journal*, 2021(7):572–598.

Qiao, Y., Wang, C.-W., and Zhu, W. (2024). Machine learning in long-term mortality forecasting. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 1:1–23.

Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and economics*, 38(3):556–570.

Richman, R. (2022). Mind the gap–safely incorporating deep learning models into the actuarial toolkit. *British Actuarial Journal*, 27:e21.

Richman, R. and Wuthrich, M. V. (2019). Lee and Carter go machine learning: recurrent neural networks. *Available at SSRN 3441030*.

Richman, R. and Wüthrich, M. V. (2021). A neural network extension of the Lee–Carter model to multiple populations. *Annals of Actuarial Science*, 15(2):346–366.

Santolino, M. (2020). The lee-carter quantile mortality model. *Scandinavian Actuarial Journal*, 2020(7):614–633.

Schelldorfer, J. and Wuthrich, M. V. (2019). Nesting classical actuarial models into neural networks. *Available at SSRN 3320525*.

Schnürch, S. and Korn, R. (2022). Point and interval forecasts of death rates using neural networks. *ASTIN Bulletin: The Journal of the IAA*, 52(1):333–360.

Scognamiglio, S. (2022). Calibrating the Lee-Carter and the poisson Lee-Carter models via neural networks. *ASTIN Bulletin: The Journal of the IAA*, 52(2):519–561.

Takeuchi, I., Le, Q., Sears, T., and Smola, A. (2006). Nonparametric quantile estimation. *Journal of Machine Learning Research*.

Vaughan, J., Sudjianto, A., Brahimi, E., Chen, J., and Nair, V. N. (2018). Explainable neural networks based on additive index models. *arXiv preprint arXiv:1806.01933*.

Wang, C.-W., Huang, H.-C., and Liu, I.-C. (2013). Mortality modeling with non-gaussian innovations and applications to the valuation of longevity swaps. *Journal of Risk and Insurance*, 80(3):775–798.

Wang, C.-W. and Yang, S. S. (2013). Pricing survivor derivatives with cohort mortality dependence under the Lee-Carter framework. *Journal of Risk and Insurance*, 80(4):1027–1056.

Wang, C.-W., Yang, S. S., and Huang, H.-C. (2015). Modeling multi-country mortality dependence and its application in pricing survivor index swaps—a dynamic copula approach. *Insurance: Mathematics and Economics*, 63:30–39.

Wang, C.-W., Zhang, J., and Zhu, W. (2021). Neighbouring prediction for mortality. *ASTIN Bulletin: The Journal of the IAA*, 51(3):689–718.

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin: The Journal of the IAA*, 26(1):71–92.

Wang, S. S. (2000). A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance*, pages 15–36.

Wang, S. S. (2002). A universal framework for pricing financial and insurance risks. *ASTIN Bulletin: The Journal of the IAA*, 32(2):213–234.

Wang, Y., Gan, D., Sun, M., Zhang, N., Lu, Z., and Kang, C. (2019). Probabilistic individual load forecasting using pinball loss guided lstm. *Applied Energy*, 235:10–20.

Yang, Y., Lou, H., Wang, Z., and Wu, J. (2024). Pinball-huber boosted extreme learning machine regression: a multiobjective approach to accurate power load forecasting. *Applied Intelligence*, 54(17):8745–8760.

Zhang, J. (2024). Blended insurance scheme: A synergistic conventional-index insurance mixture. *Insurance: Mathematics and Economics*, 119:93–105.

Zhang, N., Chen, H., and Liu, J. (2022). Mortality forecasting using lstm-cnn model. *Available at SSRN 4261735*.

Zhou, R. (2019). Modelling mortality dependence with regime-switching copulas. *ASTIN Bulletin: The Journal of the IAA*, 49(2):373–407.

Zhou, R. and Ji, M. (2021). Modelling mortality dependence: An application of dynamic vine copula. *Insurance: Mathematics and Economics*, 99:241–255.

Zhou, R., Li, J. S.-H., and Tan, K. S. (2013). Pricing standardized mortality securitizations: A two-population model with transitory jump effects. *Journal of Risk and Insurance*, 80(3):733–774.

Zhou, R., Wang, Y., Kaufhold, K., Li, J. S.-H., and Tan, K. S. (2014). Modeling mortality rates with vector error correction models: applications to solvency II. *North American Actuarial Journal*, 18(1):150–167.

Zhu, W. (2024). A deep factor model for crop yield forecasting and insurance ratemaking. *North American Actuarial Journal*, 28(1):57–72.

Zhu, W., Tan, K. S., and Wang, C.-W. (2017). Modeling multicountry longevity risk with mortality dependence: A Lévy subordinated hierarchical Archimedean copulas approach. *Journal of Risk and Insurance*, 84(S1):477–493.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, 36:1108–1126.

A Data Specifications

B Hyperparameter Tuning Results

This appendix presents the detailed results from the hyperparameter grid search for the three neural network models. We evaluated combinations of dropout rates (0.4, 0.5, 0.6) and patience values (20, 35, 50) for early stopping. The relatively high dropout rates were specifically chosen because multi-quantile regression with 51 or 101 quantiles presents a substantially higher risk of overfitting compared to point forecasting. The model must simultaneously capture both the central trend and the full conditional distribution across numerous quantiles, creating a risk of prediction drift where individual quantile predictions may diverge from the underlying mortality pattern. Higher dropout rates help mitigate this risk by preventing the model from overfitting to individual quantiles and promoting more coherent distributional predictions.

Tables 6–8 present the validation loss values used for hyperparameter selection.

Table 6: Validation loss of SimpleNN under various hyperparameter settings. *Note*: Bold value indicates lowest (best) validation loss.

Dropout	Patience		
	20	35	50
0.4	**6.3531**	6.3731	6.3977
0.5	6.3997	6.3653	6.3932
0.6	6.3720	6.4010	6.3835

Table 5: Data summary of countries/regions in this study

Country	Abbreviation	Sample period	Sample Size
Australia	AUS	1950–2020	4680
Austria	AUT	1950–2019	4680
Belgium	BEL	1950–2022	4608
Bulgaria	BGR	1950–2021	4608
Canada	CAN	1950–2021	4824
Czechia	CZE	1950–2021	4824
Denmark	DNK	1950–2023	4752
Finland	FIN	1950–2022	4752
France	FRA	1950–2021	4752
Hungary	HUN	1950–2020	4752
Iceland	ISL	1950–2021	4752
Ireland	IRL	1950–2020	4752
Italy	ITA	1950–2020	4752
Japan	JPN	1950–2022	4896
Netherlands	NLD	1950–2021	4824
New Zealand	NZL	1950–2021	4824
Norway	NOR	1950–2022	4752
Portugal	PRT	1950–2022	4752
Slovakia	SVK	1950–2019	4680
Spain	ESP	1950–2021	4680
Sweden	SWE	1950–2022	4752
Switzerland	CHE	1950–2022	4752
U.K.	GBR	1950–2021	4680
U.S.A.	USA	1950–2021	4680

Table 7: Validation loss of LCNN under various hyperparameter settings. *Note*: Bold value indicates lowest (best) validation loss.

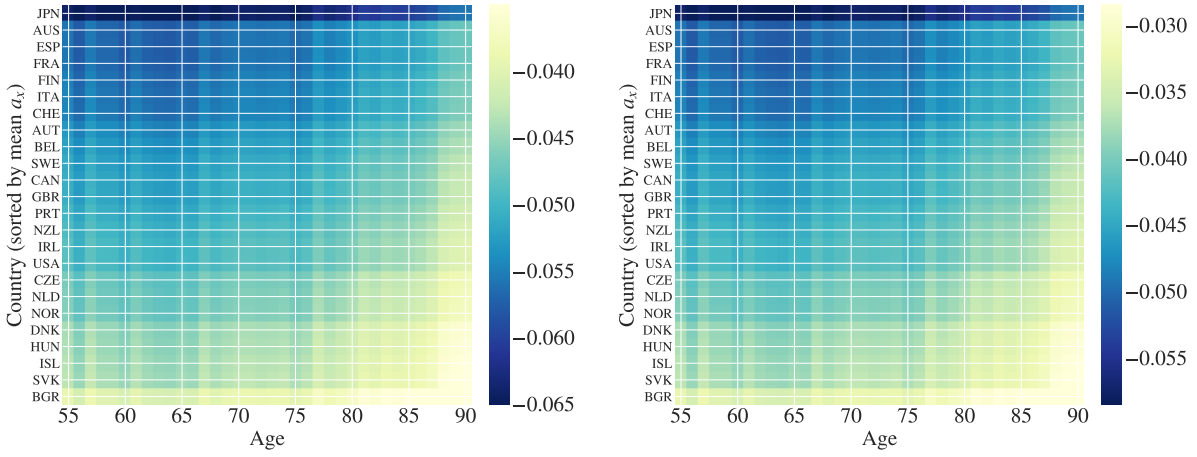
Dropout	Patience		
	20	35	50
0.4	6.2276	**6.2127**	6.2440
0.5	6.2226	6.2298	6.2415
0.6	6.2685	6.2890	6.2471

Table 8: Validation loss of RHNN under various hyperparameter settings. *Note*: Bold value indicates lowest (best) validation loss.

Dropout	Patience		
	20	35	50
0.4	6.1943	**6.1939**	6.2317
0.5	6.2191	6.2312	6.2128
0.6	6.2460	6.2363	6.2399

C Model Parameter Visualizations

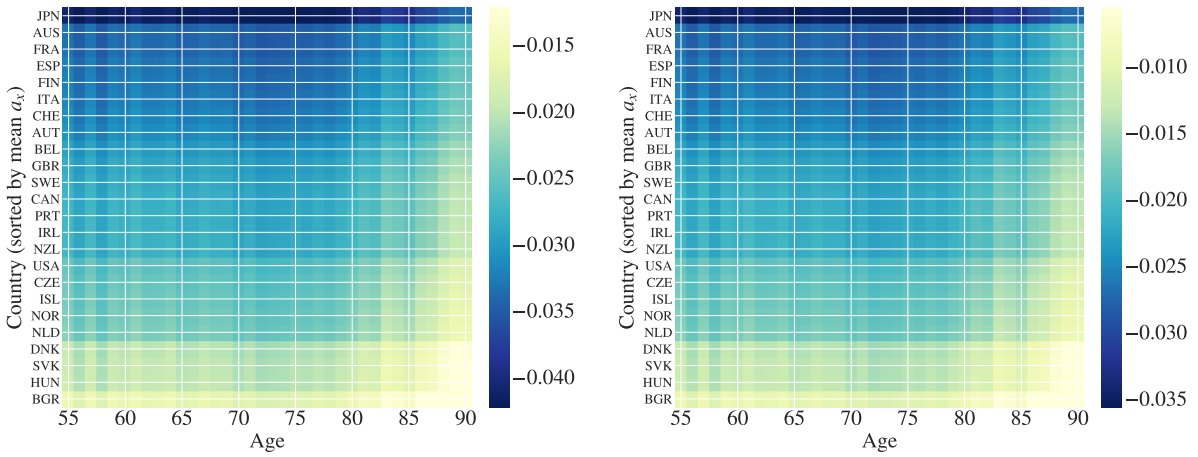
This appendix displays the estimated parameters a_x for LCNN and RHNN models.



(a) Female

(b) Male

Figure 13: Heatmap of age-specific effects a_x by country with LCNN Model: female (panel (a)) and male (panel (b))



(a) Female (b) Male

Figure 14: Heatmap of age-specific effects a_x by country with RHNN Model: female (panel (a)) and male (panel (b))