

## Article

# LSTM-Based Coherent Mortality Forecasting for Developing Countries

Jose Garrido <sup>1</sup>, Yuxiang Shang <sup>2</sup> and Ran Xu <sup>2,\*</sup>

<sup>1</sup> Department of Mathematics and Statistics, Concordia University, Montreal, QC H3G 1M8, Canada; jose.garrido@concordia.ca

<sup>2</sup> Department of Financial and Actuarial Mathematics, Xi'an Jiaotong–Liverpool University, Suzhou 215123, China; yuxiang.shang20@alumni.xjtlu.edu.cn

\* Correspondence: ran.xu@xjtlu.edu.cn

**Abstract:** This paper studies a long short-term memory (LSTM)-based coherent mortality forecasting method for developing countries or regions. Many of such developing countries have experienced a rapid mortality decline over the past few decades. However, their recent mortality development trend is not necessarily driven by the same factors as their long-term behavior. Hence, we propose a time-varying mortality forecasting model based on the life expectancy and lifespan disparity gap between these developing countries and a selected benchmark group. Here, the mortality improvement trend for developing countries is expected to converge gradually to that of the benchmark group during the projection phase. More specifically, we use a unified deep neural network model with LSTM architecture to project the life expectancy and lifespan disparity difference, which further controls the rotation of the time-varying weight parameters in the model. This approach is applied to three developing countries and three developing regions. The empirical results show that this LSTM-based coherent forecasting method outperforms classical methods, especially for the long-term projections of mortality rates in developing countries.

**Keywords:** coherent mortality forecasting; LSTM; developing countries; life expectancy; lifespan disparity



**Citation:** Garrido, Jose, Yuxiang Shang, and Ran Xu. 2024.

LSTM-Based Coherent Mortality Forecasting for Developing Countries. *Risks* 12: 27. <https://doi.org/10.3390/risks12020027>

Academic Editor: Sule Sahin

Received: 13 December 2023

Revised: 19 January 2024

Accepted: 26 January 2024

Published: 1 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In the last few decades, human mortality has improved significantly in several countries, especially in developing regions. These mortality reductions can generate important longevity risks for life insurance companies and pension schemes. The study of these longevity improvements is fundamental in life insurance and annuity research in actuarial science literature.

There are statistical techniques for the forecasting of future mortality. A popular method, proposed by Lee and Carter (1992), is the so-called Lee–Carter (LC) model, where the log force of mortality  $\ln(m_{x,t})$  is represented as the sum of an age component  $a_x$  plus the product of an age-specific function  $b_x$  and a time component  $k_t$ . Obviously, such a model cannot be fitted as a regular regression model because of the product of parameter terms. In work by Lee and Carter (1992), a two-step method was applied, where singular value decomposition (SVD) is used to fit the model, followed by an autoregressive component, or random walk, to forecast the time component  $k_t$ .

In the last two decades, the LC method has been used in the practice frequently, and several papers propose various extensions of the method; interested readers can refer to Lee (2000), Pitacco (2004), Wong-Fupuy and Haberman (2004), as well as the Cairns–Blake–Dowd (CBD) model (Cairns et al. 2006), and references therein. This literature has shown how the original LC method can lack flexibility with regard to the effect of age. Renshaw and Haberman (2003) extended the LC method to a multi-factor version adding an age-specific enhancement. In addition, a single-factor model with cohort effects is proposed

by [Renshaw and Haberman \(2006\)](#). For a comprehensive review of the early literature on these various forecasting methods, refer to [Cairns et al. \(2008\)](#) and [Cairns et al. \(2011a\)](#).

Note that apart from the LC method and its extensions, another approach developed in the literature for mortality modeling is based on generalized linear models (GLMs); for example, see [Brouhns et al. \(2002\)](#), [Renshaw and Haberman \(2006\)](#), and [O'Hare and Li \(2012\)](#). For a comprehensive survey on fitting GLMs to mortality data, refer to [Currie \(2016\)](#). In addition, the Bayesian approach appears in the literature on mortality modeling. For example, [Czado et al. \(2005\)](#) and [Pedroza \(2006\)](#) extend the LC model to Bayesian analyses using Markov chain Monte Carlo (MCMC) methods. [Cairns et al. \(2011b\)](#) further extended Bayesian stochastic mortality modeling to two populations; see [Antonio et al. \(2015\)](#) for an application of a Bayesian method under multiple populations. For more recent studies on mortality modeling with a Bayesian approach, refer to [Li and Lu \(2018\)](#), [Li et al. \(2019\)](#), and [Wong et al. \(2023\)](#), and references therein.

The LC method, and many of its early extensions, focuses on a single population (for example, modeling only one gender at a time or combined genders, and only one country). In particular, let  $m_{x,t}$  denote the mortality rate of age  $x$  at time  $t$  for  $t = 0, 1, 2, \dots, T$  and  $x = 0, 1, 2, \dots, \omega$  for a given population. Then, let  $a_x$  denote the mean log mortality rate over time, that is,  $a_x = \frac{1}{T+1} \sum_{t=0}^T \ln(m_{x,t})$ . Using the LC method, one obtains

$$\ln(m_{x,t}) = a_x + b_x k_t + \epsilon_{x,t}, \quad (1)$$

where  $\epsilon_{x,t}$  is the mean zero random noise, and  $k_t$  is modeled using random walk with drift:

$$k_t = d + k_{t-1} + \delta_t, \quad \delta_t \sim N(0, \sigma^2), \quad \mathbb{E}(\delta_t \delta_s) = 0, \quad \text{for } t \neq s. \quad (2)$$

The literature shows that it is difficult for the LC method to forecast mortality rates for two genders at the same time in one population or in multiple populations and regions, where a certain divergence could be reached due to the differences in  $b_x$  and  $d$  (and in turn, different  $k_t$ ) in the model. For example, [Carter and Lee \(1992\)](#) suggested using the same  $k_t$ , but gender-specific  $b_x$  values to forecast male and female mortality rates separately in the U.S. [Lee and Nault \(1993\)](#) used the same  $k_t$  and  $b_x$  for mortality forecasting in each province of Canada, which works only when the  $b_x$  values of different provinces, as obtained from historical data, are similar.

However, some early discussions on the convergence of life expectancy around the world (see, e.g., [White 2002](#), [Wilmoth 1998](#), and [Vaupel and Schnabel 2004](#)) show that there is convergence in long-term life expectancy, diverging forecasts of mortality rates for different populations in a group of countries is unrealistic. Therefore, [Li and Lee \(2005\)](#) introduced the so-called coherent extension of the LC method for mortality forecasting of a group of populations (we call it the Li-Lee method), where the log mortality rates for each member in the group are decomposed into three parts, namely, member-specific  $a_{i,x}$ , common age and period effects  $B_x$  and  $K_t$ , and member-specific age and period effects,  $b_{i,x}$  and  $k_{i,t}$ . More precisely, the log mortality rates  $\ln(m_{i,x,t})$ , for member  $i$  in the group at age  $x$  and time  $t$ , can be expressed as

$$\ln(m_{i,x,t}) = a_{i,x} + B_x K_t + b_{i,x} k_{i,t} + \epsilon_{i,x,t}, \quad (3)$$

with

$$K_t = d_0 + K_{t-1} + v_t, \quad k_{i,t} = \alpha_{0,i} + \alpha_{1,i} k_{i,t-1} + \epsilon_{i,t}, \quad (4)$$

where  $a_{i,x}$  measures the average mortality level at age  $x$  in country  $i$ .  $K_t$  is the common period effect for all countries and is modeled by a random walk with drift  $d_0$ .  $B_x$  is the common age effect, i.e., the common mortality sensitivity at age  $x$ , with respect to  $K_t$ . In addition,  $k_{i,t}$  and  $b_{i,x}$  are the country-specific period and age effects, respectively, which measure the fluctuations around the common mortality patterns in the group for country  $i$ . Finally,  $\epsilon_{i,x,t}$ ,  $v_t$ , and  $\epsilon_{i,t}$  are normally distributed i.i.d. errors. It turns out that the additional

information provided by similar members/countries in the group can improve the forecast accuracy for individual countries.

On the other hand, according to Hanewald (2011), there is a strong long-term connection between the mortality dynamics and the gross domestic product (GDP) per capita and unemployment rate in a country, which points to the essential difference of the mortality improvements between developed countries and developing countries during the same period of time. Based on such an observation, Niu and Melenberg (2014) improved the LC model with an extra factor (namely, GDP) describing the economic growth; see Boonen and Li (2017) for the study under multiple populations. More recently, Ma and Boonen (2023) further argued that the consumer price index (CPI) is a more suitable factor, added to the LC model, to explain the mortality trends in a country. It explains the affordability of healthcare, food, and housing in that country. The above-mentioned studies verify, from a different point of view, that life expectancy (or any other similar indexes) can play an important role in mortality forecasting.

Moreover, as mentioned by Li et al. (2013), mortality decline decelerates in younger ages and accelerates at older ages in many developed countries. Such a “rotation” can generate problems in the results of long-term mortality projections using the LC method for developing countries that do not exhibit such a subtle rotation in their historical data, e.g., the projected mortality rates are low for younger ages. Hence, Li et al. (2013) developed a rotation-based LC method, where the out-of-sample  $b_x$  was assumed to be converging to an ultimate structure based on the development of life expectancy. But, as argued by Li and Lu (2017), mortality rates should change smoothly and continuously across ages; such a problem, known as age-coherent mortality forecasting, is present in the above-mentioned Li–Lee model. Thus, Gao and Shi (2021) proposed two alternative extensions to the ordinary LC method: the *LC-Geometric* and *LC-Hyperbolic* models. The goal was to achieve long-term age coherence in mortality forecasts while retaining the short-term rotation-type forecasting adopted by Li et al. (2013). Here, *Geometric* and *Hyperbolic* refer to the type of decay allowed in the autoregressive (AR) model.

Recently, borrowing the concept of “rotation” proposed by Li et al. (2013), Li et al. (2021) developed a so-called *rotation algorithm* for the coherent mortality projections of less developed countries, which included all regions of Africa, Asia (except Japan), Latin America, the Caribbean, Melanesia, Micronesia, and Polynesia. Using Li–Lee’s model, where “rotation” refers to the effects of age and time components in the projection phase for developing countries; rotation may occur in  $b_{i,x}$  and  $d_i$ , based on their own historical data and the corresponding data from a group of developed countries used as the benchmark group. In their model, the rotation algorithm is controlled by a life expectancy gap function, between the target developing country and the benchmark group, where the gap function can be fitted by (double) logistic functions with some selected threshold levels for convergence of the gap.

Therefore, reliable mortality projection methods, especially the long-term projections of age-specific mortality rates, are crucial for developing countries. However, the recent fast decline in aggregate mortality might not be a long-term behavior (according to Müller and Krawinkel (2005), Austin and McKinney (2012), and Jeuland et al. (2013), the main factors contributing to recent mortality improvements in developing countries (especially for infants, the young, and the working-age population) are modernization, improved healthcare coverage, better nutrition, and prevention of infectious diseases, which can obviously only last for a short period of time); in the long run, the mortality patterns of developing countries could more closely resemble those of more developed countries (see, e.g., Li and Lee 2005). Therefore, predicting long-term, age-specific mortality rates in a developing country by simply extrapolating its historical patterns may lead to implausible results.

Hence, a method with the aforementioned “rotation” helps find a balance between the historical mortality pattern of the developing country and the average mortality patterns of a group of developed countries (the benchmark group). Note that for different target countries, the method proposed by Li et al. (2021) needs to use expert judgments and fit

different gap functions with possible different gap thresholds. This can bring restrictions to the unified application of the method.

On the other hand, as mentioned by [Aburto et al. \(2020\)](#), populations with the same life expectancy level may experience substantial differences in the time of death. This indicates that the mortality pattern of a developing country can still be different from the benchmark group even if there is a convergence in the life expectancy gap. Hence, the lifespan disparity, which describes life expectancy lost due to death by an individual at different ages and times (see, e.g., [Vaupel and Romo \(2003\)](#) and [Zhang and Vaupel \(2009\)](#)), may provide additional information when examining the convergence of mortality development between the developing country and the benchmark group. Therefore, to continue the study of coherent mortality forecasting (especially long-term projections based on lifespan disparity) for developing countries, here we propose a unified coherent mortality forecasting method with time-dependent rotation weights based on a benchmark group. A deep neural network, in particular a long short-term memory (LSTM), is used for the projection of the life expectancy and lifespan disparity gaps between the target developing country and the corresponding benchmark group. The projected gaps are used in the control of the rotated time-varying weight parameters in the model during the projection phase for mortality forecasting of the developing country.

In the last decade, neural networks, especially deep neural networks (DNNs), have gained attention in the human mortality modeling and forecasting literature. The purpose of building neural network models for human mortality is to extend the modeling and forecasting ability passed by classical parametric models, such as the LC method and its many extensions. For example, [Hainaut \(2018\)](#) proposed a type of *neural network analyzer*, which uses an encoding and decoding network structure in the approximation of the non-linearity among ages, for each year in the data, for a single country. The model is essentially a feedforward neural network extension of the LC method, where a simple, fully connected feedforward neural network was used to learn the common nonlinearities in the lower dimensional structure of the log-forces of mortality, for different ages crossing the years.

[Nigri et al. \(2019\)](#) extended the classical LC method by introducing an LSTM model for the time series prediction in the forecasting phase of the LC method, where the related time series (i.e.,  $k_t$ ) were extracted by following the same SVD method by [Lee and Carter \(1992\)](#). Their results show the prediction power of LSTM, compared to the classical time series prediction method (e.g., ARIMA), especially in capturing the nonlinearities.

More recently, [Lindholm and Palmberg \(2022\)](#) discussed the procedures to efficiently use training data in mortality forecasting when applying an LSTM-based Poisson LC method. [Marino et al. \(2022\)](#) further confirmed that an LSTM model can improve the predictive power of the classical LC method by providing a rigorous analysis of the prediction interval for their so-called LC-LSTM model. Note that [Nigri et al. \(2021\)](#) also used the LSTM model for life expectancy and lifespan disparity forecasting. According to the above-mentioned literature, the LSTM model has shown great prediction power for the forecasting of period effects in the classical LC model, as well as the forecasting of life expectancy and lifespan disparity. It is interesting to further consider the question of whether such a powerful tool (LSTM) can help improve long-term mortality forecasting for developing countries.

On the other hand, deep feedforward neural networks (FNNs), as a different way of extending the LC method, may also be applied to mortality modeling; e.g., see [Richman and Wüthrich \(2021\)](#), where they treat human mortality modeling as a classical supervised learning problem. For the application of convolution neural networks (CNNs) in mortality modeling, refer to [Wang et al. \(2021\)](#) and [Schnürch and Korn \(2022\)](#). However, their neural networks and methods are fundamentally different from the LSTM-based neural network structure here; therefore, their results are not directly comparable. Furthermore, some non-neural network machine learning models have appeared in the literature to forecast human mortality rates; for example, see [Deprez et al. \(2017\)](#) and [Levantesi and Pizzorusso \(2019\)](#).

As a result, in this paper, we develop an LSTM-based coherent mortality forecasting method with time-varying rotation structures based on a benchmark group. It provides a unified and model-free mortality projection method for developing countries. The paper is organized as follows: Section 2 introduces some preliminaries, including neural networks, the LSTM, the classical LC, and the Li–Lee model, as well as the definitions of life expectancy and lifespan disparity. The LSTM-based coherent mortality forecasting model is presented in Section 3. Finally, the mortality data and the empirical results are presented in Section 4, followed by the conclusion and some remarks in Section 5.

## 2. Preliminaries

### 2.1. RNN with LSTM Architecture

Recurrent neural networks (RNNs) are a class of artificial neural networks (ANNs), that can store representations of recent input data through their feedback connections. RNNs have many significant applications in the areas of the speech process, non-Markov control, or time series prediction (see, e.g., Mozer 1991). For instance, let  $\{x_1, x_2, \dots, x_n\}$  denote the time sequence of input vectors, and  $\{h_1, h_2, \dots, h_n\}$  denote the time sequence of output vectors; for the simple RNN, the output vector at time-step  $t$  is defined as follows:

$$h_t = \phi(W_{hh}h_{t-1} + W_{hx}x_t + b_h),$$

where  $\phi$  is an activation function,  $W_{hh}$  and  $W_{hx}$  are the kernel weights for previous time step outputs and current inputs, respectively, and  $b_h$  is the corresponding bias.

However, with the conventional gradient-based *back-propagation through time* (BPTT) algorithm (see Williams and Zipser (1995)), simple RNNs suffer from the problem of vanishing or exploding gradients (Pascanu et al. 2013). Then, RNNs with an LSTM architecture, or simply LSTMs, were introduced by Hochreiter and Schmidhuber (1997) in order to overcome such vanishing gradient problems. Instead of using all the memory dynamically when processing the data, the LSTM architecture relies both on the memory block and a few gates for controlling data elaborations.

LSTMs have shown great power in natural language processing and time series predictions. According to Marino et al. (2022), the LSTM can be expressed in the following mathematical form. Let  $N_0$  denote the number of neurons within the input layer,  $N_p$  denote the number of neurons of the  $p$ -th hidden layer with  $n \in \{1, \dots, P\}$ , and  $N_{p+1}$  denote the number of neurons of the output layer, where  $P, N_0$ , and  $N_p$  for  $p \in \{1, \dots, P\}$  and  $N_{p+1} \in \mathbb{N}$ . Then, the activation of the  $p$ -th hidden layer may expressed as an affine mapping,  $A^{(p)} : \mathbb{R}^{N_{p-1}} \rightarrow \mathbb{R}^{N_p}$ , where  $\mathbb{R}^{N_{p-1}}$  refers to the output produced by the  $(p - 1)$ -th hidden layer. The output of an LSTM neuron at any time  $t$  in the  $p$ -th hidden layer can be expressed as follows:

$$h_t^{(p)} = o_t^{(p)} \odot \tanh(c_t^{(p)}),$$

where  $\odot$  denotes the element-wise product. The key to the LSTM lies in the following equations, which describe the outputs of four different gates in the architecture:

$$\begin{aligned} \text{Forget gate:} \quad & f_t^{(p)} = \sigma_f \circ A^{(p)} = \sigma(\langle W_f^{(p)}, h_t^{(p-1)} \rangle + \langle U_f^{(p)}, h_{t-1}^{(p)} \rangle + b_f^{(p)}), \\ \text{Input gate:} \quad & i_t^{(p)} = \sigma_i \circ A^{(p)} = \sigma(\langle W_i^{(p)}, h_t^{(p-1)} \rangle + \langle U_i^{(p)}, h_{t-1}^{(p)} \rangle + b_i^{(p)}), \\ \text{Output gate:} \quad & o_t^{(p)} = \sigma_o \circ A^{(p)} = \sigma(\langle W_o^{(p)}, h_t^{(p-1)} \rangle + \langle U_o^{(p)}, h_{t-1}^{(p)} \rangle + b_o^{(p)}), \\ \text{Memory state:} \quad & c_t^{(p)} = f_t^{(p)} \odot c_{t-1}^{(p)} + i_t^{(p)} \odot \tanh(\langle W_c^{(p)}, h_t^{(p-1)} \rangle + \langle U_c^{(p)}, h_{t-1}^{(p)} \rangle + b_c^{(p)}), \end{aligned}$$

where  $\sigma(x) = (1 + e^x)^{-1}$  is the sigmoid activation function,  $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$  is the hyperbolic tangent activation function,  $W_k^{(p)}$  for  $k = f, i, o, c$  are, respectively, the weight matrices for the four different gates of feedforward connections in the structure,  $U_k^{(p)}$  for



$k = f, i, o, c$  are the corresponding weight matrices for the gates of recurrent connections, and  $b_k^{(p)}$  for  $k = f, i, o, c$  are the bias terms in the model.

Then let  $\mathcal{D} = \{(x_t, \mathbf{y}_t), x_t \in \mathbb{R}^{N_0}, \mathbf{y}_t \in \mathbb{R}^{N_{p+1}}\}$  be a dataset where  $x_t$  and  $\mathbf{y}_t$  are the input variables and associated responses at time  $t$ , respectively. Hence, the LSTM is essentially a function, say  $g_{LSTM} : \mathbb{R}^{N_0} \rightarrow \mathbb{R}^{N_{p+1}}$ , with

$$\mathbf{y}_t = g_{LSTM}(x_t; \mathbf{W}) + \gamma_t = \psi \circ \left( \mathbf{h}_t^{(p)} \circ \mathbf{h}_t^{(p-1)} \circ \dots \circ \mathbf{h}_t^{(1)} \right) (x_t; \mathbf{W}) + \gamma_t, \quad (5)$$

where  $\psi : \mathbb{R}^{N_p} \rightarrow \mathbb{R}^{N_{p+1}}$  is the activation function at the output layer,  $\mathbf{W}$  is the set of all weight parameters in the network, and  $\gamma_t$  is a noise term, with zero mean and variance  $\sigma_t^2$ , independent of  $g_{LSTM}$ .

## 2.2. The Mortality Models

Here, we use two classical mortality projection methods, namely the LC models and Li–Lee model. Specifically, we use the LC method for a rough/first-step mortality forecast of the target developing countries. That is, for such single populations, the Lee–Carter method (Lee and Carter 1992) assumes that the logarithm of the crude death rates ( $m_{x,t}$ ) for each age  $x$  and year  $t$  satisfies (1) and (2), where  $a_x$  summarizes the average level of mortality throughout the time at age  $x$ ,  $k_t$  provides the overall level of mortality at year  $t$ , and  $b_x$  measures the age effect of mortality on different periods. Note that the following is also assumed in the LC method (for identification purposes):

$$a_x = \frac{1}{T+1} \sum_{t=0}^T \log(m_{x,t}), \quad \sum_{x=0}^{\omega} b_x = 1.$$

On the other hand, the Li–Lee method introduced by Li and Lee (2005) is an extension of the classical LC method, which can generate coherent mortality projections for multiple countries. In this study, the Li–Lee method (given by (3) and (4)) is used as the first step in mortality forecasting of the benchmark countries (i.e., a selected group of developed countries). Note that  $K_t$ , in general, can be fitted by a random walk with drift (i.e., non-stationary), whereas  $k_{i,t}$  for all  $i$  is assumed to be stationary; that is,  $k_{i,t}$  shall be fitted by a random walk without drift or first-order autoregressive model  $AR(1)$  with a coefficient that yields a bounded short-term trend; for more details, refer to Li and Lee (2005).

It is obvious that under the Li–Lee method, the long-term mortality trend is uniquely determined by the common period effect  $K_t$ , which makes the mortality forecasts coherent for all member countries in the group. In the empirical application, if the  $k_{i,t}$  of a country is non-stationary, then this country is considered non-coherent with other countries in the group, i.e., there is significant divergence between the historical mortality experience and the common mortality patterns  $B_x$  and  $K_t$ . Hence, we may need to exclude them from the selected benchmark group.

Note that in order to ensure comparability between the parameters in the LC and Li–Lee methods in our subsequent analysis, one should impose the same normalization constraints on the key parameters of the two methods (see, e.g., Li et al. 2018).

## 2.3. Life Expectancy and Lifespan Disparity

Most mortality forecasting methods aim to predict how many additional years of life people will gain in the future. Life expectancy at birth, which measures the central tendency, is frequently applied to evaluate the precision of mortality forecasting methods. However, as mentioned by Aburto et al. (2020), populations with the same life expectancy level might experience substantial differences in the time of death, i.e., life expectancy cannot detect distributional variations in lifespan. Therefore, lifespan disparity can serve as an additional indicator to evaluate mortality forecasting methods (see, e.g., Bohk-Ewald et al. 2017).

Let us introduce the notation and definitions of life expectancy and lifespan disparity. Let  $S(x, t)$  and  $\mu(x, t)$  denote the survival function and the force of mortality for an individual age  $x$  at time  $t$ , respectively, for a given population. These are assumed to be two continuous functions with respect to  $x$  and  $t$ . Also, denote by  $e_{x,t}$  the life expectancy for age  $x$  at time  $t$  as

$$e_{x,t} = \frac{\int_x^\infty S(y, t) dy}{S(x, t)}, \quad (6)$$

where

$$S(x, t) = \exp\left(-\int_0^x \mu(a, t) da\right),$$

and  $\mu(a, t)$  is the corresponding force of mortality at age  $a$  and time  $t$ .

To measure lifespan disparity, we take the average number of life years lost at birth (see [Vaupel and Romo 2003](#); [Zhang and Vaupel 2009](#)):

$$e_{0,t}^+ = \int_0^\infty e_{y,t} d(y, t) dy, \quad (7)$$

where  $d(y, t)$  are the deaths at age  $y$  and time  $t$ , and  $e_{y,t}$  is the remaining life expectancy at age  $y$  and time  $t$ . Obviously, (7) shows that lifespan disparity is an indicator representing the life expectancy lost due to death by an individual at age  $x$  at time  $t$ . Note that lifespan disparity can be described by other measures, such as the standard deviation, interquartile range, Gini coefficient, or prolate index. However, we shall use  $e_{0,t}^+$  for lifespan disparity in our analysis. Demographically, apart from its interpretation as the average life years lost or lost living potential, it also provides information about the capacity for further increases in life expectancy (see [Bohk-Ewald et al. 2017](#)).

### 3. LSTM-Based Coherent Method

In this section, we introduce our LSTM-based coherent mortality forecasting method with an embedded rotation in the time-varying model weight parameters during the projection phase for developing countries. The key to our method is a time-varying LC model, see (8) below, with an LSTM-based component that controls the rotation of the weight parameters in the model. Here, the rotation is referred to as the gradual change in the mortality development pattern during the projection phase (see [Li et al. 2021](#)). The time-varying parameters are defined through time-dependent weighted averages of the corresponding projected parameters based on the historical data from the target developing country and the selected benchmark group, respectively, where the weights are rotated based on how close the target developing country is to the benchmark group, under given criteria.

More specifically, we consider the following projection method for the logarithm of central death rate  $m_{x,t}^j$  for a particular developing country, say  $j$  at age  $x$  and year  $t$ , such that

$$\begin{aligned} \ln(m_{x,t}^j) &= a_x^j + b_{x,t}^j k_t^j + \varepsilon_{x,t}, \\ k_t^j &= d_t^j + k_{t-1}^j + \epsilon_t, \end{aligned} \quad (8)$$

where  $a_x^j$  can be estimated by the average mortality level at age  $x$  for developing country  $j$ . Here,  $k_t^j$  is the period effect and  $\varepsilon_{x,t}$  and  $\epsilon_t$  are two zero mean random noises.

The main difference between our method, in (8), and the classical LC method is the time-varying  $b_{x,t}^j$  that measures a time-dependent age effect on mortality at different periods, and  $d_t^j$  describes a time-dependent drift in the random walk model used to project  $k_t^j$ .

Then, we select a group of developed countries as the benchmark group, where the classical Li–Lee method is applied in the projection of the logarithm of the central death rates  $m_{i,x,t}$ , for member  $i$  in the group, that is,

$$\ln(m_{i,x,t}) = a_{i,x} + B_x K_t + b_{i,x} k_{i,t} + \epsilon_{i,x,t},$$

with

$$K_t = d_0 + K_{t-1} + v_t, \quad k_{i,t} = \alpha_{0,i} + \alpha_{1,i} k_{i,t-1} + \epsilon_{i,t}.$$

$B_x$  measures the common age effect in the benchmark group, and  $d_0$  gives the drift in the random walk model of the common period effect  $K_t$ .  $B_x$  and  $d_0$  are the two key components to be extracted from the benchmark group, using the Li–Lee model, to then be used in the rotation during the projection phase.

Here is how the rotation works in our method. As explained in Section 2, first select the life expectancy and lifespan disparity, respectively, as the criteria to describe the gap in terms of mortality levels, between the target developing country and the benchmark group. The varying gap will control the rotation of age and period effects in the projection phase for developing countries.

However, instead of selecting and fitting various (double) logistic functions as well as tailored threshold levels for the gap (see Li et al. 2021), we propose using a unified LSTM model for the gap forecasting. In addition, we use life expectancy, given by (6), and lifespan disparity, given by (7), respectively, in the construction of the gap function that describes the mortality distance between the target developing country and the benchmark group.

In particular, for notation simplicity, we let  $e_u^i$  and  $e_u^{ti}$  denote, respectively, the (projected) life expectancy and lifespan disparity at birth for the  $i$ -th member in the benchmark group, in year  $u$ , and define the corresponding average life expectancy and lifespan disparity at birth, in year  $u$ , for the whole benchmark group as follows:

$$e_u^b = \frac{1}{N} \sum_{i=1}^N e_u^i, \quad \text{for } u = \dots, T, \dots,$$

$$e_u^{tb} = \frac{1}{N} \sum_{i=1}^N e_u^{ti}, \quad \text{for } u = \dots, T, \dots,$$

where  $T$  is the number of years in the training/in-sample data and  $N$  is the total number of members in the benchmark group.

Let  $e_u^j$  and  $e_u^{tj}$  denote the corresponding life expectancy and lifespan disparity at birth for the target developing country/region  $j$ , in year  $u$ , for  $u = \dots, T, \dots$ . A unified LSTM model is introduced to forecast the life expectancy and lifespan disparity for both the target developing country and the benchmark group, such that the projected gaps in mortality levels between the developing country and the benchmark group can be expressed as the forecast for life expectancy or lifespan disparity difference.

More specifically, we construct LSTM models for the projection of  $e_t$  and  $e_t^{t\cdot}$  for both target developing countries/regions (i.e.,  $e_t^j$  and  $e_{x,t}^{tj}$ ) and the benchmark group (i.e.,  $e_t^b$  and  $e_t^{tb}$ ) as follows:

$$e_t = g_{LSTM}^e(e_{t-1}; W^e) + \epsilon_t^e, \quad e_t^{t\cdot} = g_{LSTM}^{e^t}(e_{t-1}^{t\cdot}; W^{e^t}) + \epsilon_t^{e^t}, \tag{9}$$

where  $\epsilon_t^e$  and  $\epsilon_t^{e^t}$  are zero mean errors.  $g_{LSTM}^e$  and  $g_{LSTM}^{e^t}$  are given by (5), respectively, for life expectancy and lifespan disparity. And  $W^e$  and  $W^{e^t}$  are the weight parameters in the corresponding LSTM models (see for example Nigri et al. 2021). The parameters in the LSTM model (i.e., the functional form of  $g_{LSTM}^e$  and  $g_{LSTM}^{e^t}$ ) are optimized using an  $L^2$  loss function, namely

$$\min_{W^e} \frac{1}{2} \sum_t (e_t - g_{LSTM}^e(e_{t-1}; W^e))^2, \quad \min_{W^{e^t}} \frac{1}{2} \sum_t (e_t^{t\cdot} - g_{LSTM}^{e^t}(e_{t-1}^{t\cdot}; W^{e^t}))^2.$$



Note that, to show the long-term projection power of our method, we need a sufficient number of years of mortality rates in the out-of-sample data, which results in a limited size of the in-sample mortality data for the training of the neural network model. Therefore, in this paper, we only consider a first-order autoregressive approach in the LSTM model, as illustrated in (9), where the neural network learns at each time step the relationship between two consecutive values during the training period (i.e., one-to-one structure). In addition, the method can be extended to more complex LSTM models with the structure of many-to-one or many-to-many, given that the available data are sufficiently large.

Next, we illustrate in detail our LSTM-based coherent mortality forecasting method. Consider the method based on lifespan disparity; for the case with life expectancy, one simply replaces all  $e_t^\dagger$  by  $e_t$  in the corresponding equations. As noted, at the core of the method is the time-varying LC model given in (8), where the term for the age effect,  $b_{x,t}^j$ , and the drift term,  $d_t^j$ , of the period effect, depend on time  $t$ . The time dependence is described through a set of time-varying weights, in terms of lifespan disparity gaps, linking the mortality improvements between the target developing country and the benchmark group. Let  $\hat{b}_x^j$  and  $\hat{d}^j$  denote the estimated age effect term and the drift parameter of the period effect term for the target developing country  $j$ , based on the classical LC method.

Now, let  $\hat{B}_x$  denote the estimated common age effect term and  $\hat{d}_0$  denote the drift parameter of the common period effect term, obtained by using Li–Lee’s method for the benchmark group. These two parameters provide information on the common mortality improvements of the benchmark group. Then, the next step is to specify how the time-varying  $b_{x,t}^j$  and  $d_t^j$  in (8) are defined in the mortality projection phase. To be specific, at the beginning of the projection phase, one can simply rely on the historical mortality data of the developing country when forecasting the short-term mortality rates.

In addition, denote the lifespan disparity at birth at time  $t$  projected through the LSTM model given in (9) as  $e_t^{\dagger j}$  and  $e_t^{\dagger b}$  for the developing country/region  $j$  and the benchmark group, respectively. Then, define the lifespan disparity gap at time  $t$  between the target country/region and the benchmark group as follows:

$$g_t^\dagger := e_t^{\dagger b} - e_t^{\dagger j}. \tag{10}$$

Then, for intermediate or long-term projections, include data from the benchmark group so that the long-term mortality development of the developing country converges gradually to the common trend in this benchmark group. Hence, redefine the age effect and drift terms of the period effect in the LC model as

$$\begin{aligned} b_{x,t+1}^j &:= (1 - \omega_t)\hat{b}_x^j + \omega_t\hat{B}_x, \\ d_{t+1}^j &:= (1 - \omega_t)\hat{d}^j + \omega_t\hat{d}_0, \end{aligned} \tag{11}$$

where

$$\omega_t = \left\{ \frac{1}{2} \left( 1 + \sin \left[ \frac{\pi}{2} \left( 2 \max \left( \frac{g_T^\dagger - g_t^\dagger}{g_T^\dagger}, 0 \right) - 1 \right) \right] \right) \right\}^p, \tag{12}$$

for each age  $x$  and  $t = T, T + 1, \dots$ , and  $g_t^\dagger$  is given by (10).  $\omega_t$  denotes the time-varying weights that link the projected time-dependent age and period effect parameters, in year  $t + 1$ , to the weighted average of the estimated  $\hat{b}_x^j$  and  $\hat{d}^j$ , respectively, with  $\hat{B}_x$  and  $\hat{d}_0$  in the first step (see, e.g., Li et al. 2013). To simplify the analysis, we apply here the same weight parameter for both  $b^j$  and  $d^j$  in (11). In addition,  $p \in [0, 1]$  in (12) is a tuning parameter that controls the functional form of  $\omega_t$ . We set  $p = 1$  in our analysis such that  $\omega_t$  has a considerably low rate of change when its value is close to zero or one. Note that when  $t = T$ , we have  $\omega_t = 0$ , which means that at the beginning of the projection phase, the method relies only on the historical data from the developing country. For  $t > T$ , the lifespan disparity gap decreases, such that  $\omega_t$  increases smoothly to one if the lifespan

disparity gap diminishes in the future projection phase. Note that if the projected lifespan disparity gap for a particular developing country diverges (e.g.,  $g_t^+ > g_T^+$  for  $t > T$ ), we simply forecast the mortality based on its own historical data (that is  $\omega_t = 1$  for all  $t > T$ ).

Finally, we summarize the model structure schematically in Figure 1 below. To be specific, the LSTM-based coherent mortality forecasting model contains two parts: (1) on the left of Figure 1 is a neural network component that contains an input layer, two layers of the LSTM structure (both LSTM layers contain 128 neurons with accompanied dropout layers), and two (fully connected) dense layers that contain 64 and 32 neurons, respectively, with accompanied dropout layers for the output. Note that adding dense layers in the model can provide flexibility in the control of the non-linearity of the model. (2) On the right of Figure 1, the projected life expectancy or lifespan disparity is transferred into a component of the rotation algorithm for the calculating of time-varying weights, and then the projected weights are applied to the time-varying LC model for the forecasting of mortality rates.

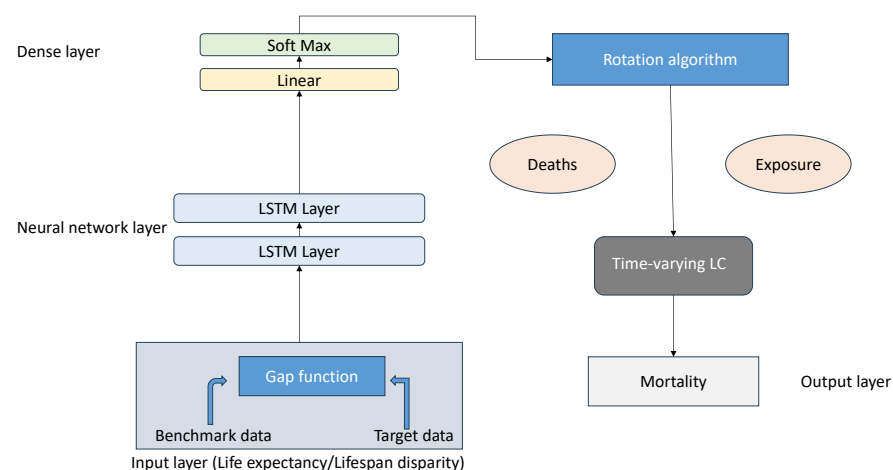


Figure 1. LSTM-based coherent mortality forecasting model.

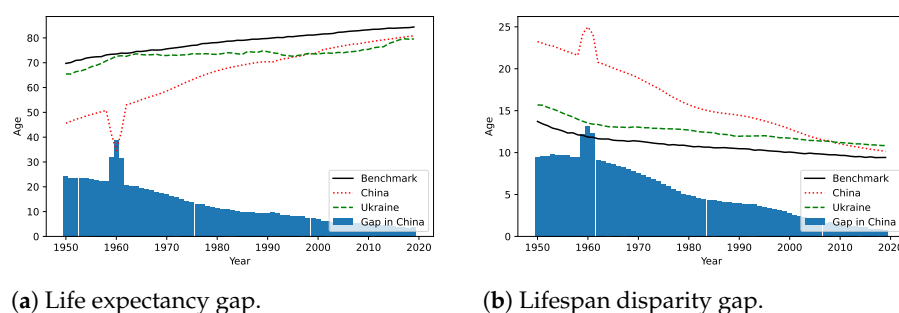
#### 4. Empirical Analysis

This section presents the application of our LSTM-based coherent mortality forecasting method to three developing countries, namely China, Brazil, and Nigeria, which are the most populous countries in their respective continents and also belong to the emerging/emerged markets in the world. According to BBVA (2014), China and Brazil are classified as *EAGLEs*, i.e., emerging and growth-leading economies that are expected to have GDP increments larger than the average of G7 economies, excluding the US, in the next ten years. Nigeria is classified as *NEST*, i.e., an emerging country that is expected to have GDP increments lower than the average of the G7—excluding the US but higher than Italy's—in the next ten years. In addition, we apply our method to three developing regions, namely less developed region(s) (LDR), less developed regions excluding China (LDRexChina), and less developed regions excluding the least developed countries (LDRexLDC). The United Nations defines the less developed countries/regions as all regions of Africa, Asia (except Japan), Latin America, and the Caribbean, plus Melanesia, Micronesia, and Polynesia, and categorizes 45 countries as the least developed countries (UN Source: <https://unctad.org/topic/least-developed-countries/list> (accessed on 21 November 2023)), including 33 countries in Africa, 8 countries in Asia, 1 in the Caribbean, and 3 in the Pacific. To proceed with the empirical results, we first introduce the mortality data used in the analysis.

#### 4.1. Mortality Data

In this study, the benchmark group is made up of nine selected developed countries, namely Denmark, Finland, France, the Netherlands, Switzerland, Sweden, the UK, the US, and Japan. The mortality rates of these countries are obtained from the Human Mortality Database. In particular, we use the central death rates in the one-age and one-year blocks, i.e., ages equal to 0, 1, 2, 3, . . . , 97, 98, 99, and years ranging from 1950 to 2019.

The mortality data for the six target developing countries/regions mentioned above are not included in the Human Mortality Database. Hence, the corresponding data are obtained from the population division of the United Nations (UN Source: <https://population.un.org/wpp/Download/Standard/Mortality/> (accessed on 21 November 2023)). Note that, a necessary condition for the application of our method is that the life expectancy or lifespan disparities of the target countries/regions converge to the ones of the benchmark group. Hence, a preliminary study is needed to select developed countries that can form a proper benchmark group. Figure 2 illustrates the convergence of life expectancy and lifespan disparity at birth between China and the benchmark group. The life expectancy and lifespan disparity at birth in the Ukraine do not converge. According to Figure 2, one can recognize a spike around the year 1960 in both the life expectancy and lifespan disparity in China. Such mortality outliers are due to the so-called *Great Chinese Famine* of 1959 to 1961. Hence, in order to reduce the effects of such extreme outliers from China, in the following analysis, we use only the data from the year 1962 to 2019 whenever the data of China are involved (i.e., the cases with China, LDR, and LDRexLDC).



(a) Life expectancy gap.

(b) Lifespan disparity gap.

**Figure 2.** China and Ukraine vs. benchmark group.

#### 4.2. LSTM for Life Expectancy and Lifespan Disparity

As discussed above, in order to construct time-varying weights that depend on the convergence of the life expectancy and lifespan disparity of a developing country to those of the benchmark group, one needs to develop projection models for the corresponding life expectancy or lifespan disparity gap.

Note that in the literature, (see Li et al. 2021), the forecasting of the life expectancy gap based on statistical methods uses different functions (logistic or double logistic) for developing countries/regions. Also, exogenous thresholds need to be introduced to test convergence in the model. The situation is even more complex if the lifespan disparity gap is also introduced in the method.

Here, instead of fitting various functions with thresholds to the life expectancy and lifespan disparity gaps, we use a unified LSTM model (see, e.g., Nigri et al. 2021) for the forecasting and identification of the gaps, compared with the benchmark group, for both the life expectancy and lifespan disparity of all six countries/regions. The projected life expectancy and lifespan disparity using the LSTM model for the six target countries/regions are presented in Appendix A.1. For similar results regarding the projection of life expectancy and lifespan disparity of other countries selected from the Human Mortality Database, refer to Nigri et al. (2021).

### 4.3. Empirical Results

In this empirical study, we carry out an out-of-sample test when training the model with the above-mentioned mortality dataset. For the purpose of long-term predictions, the data are divided into two parts, where the first 35 years of data, from 1950 to 1984 (from 1962 to 1984 for China, LDR, and LDRexLDC), are used for training, and the rest of the data, from 1985 to 2019, are set aside as the test data. To avoid possible overfitting to the training dataset, 20% of the training data are selected randomly as a validation part at each epoch.

To assess our models' projections accuracy, the criteria used are mean square error, root mean square error, and mean absolute error for the projected log-mortality rates in the test data. The forecasting results for the LSTM-based, time-varying LC method, which includes life expectancy and lifespan disparity, respectively, are compared with the traditional LC and Li–Lee methods. All the experiments were performed using Keras with TensorFlow in Python, for the LSTM model, the R package "StMoMo" for LC, and the Li–Lee method for the initial mortality data processing. Note that in the following tables, we use LSTM–ex to denote our forecasting model based on life expectancy, and use LSTM–disp to denote our model based on lifespan disparity.

### 4.4. Results for China, Brazil, and Nigeria

The first empirical results are for the application of our model to the mortality data of China, Brazil, and Nigeria. These selected target developing countries represent demographic trends in their continents. This strategy removes the effect of different ethnic groups on life expectancy or lifespan disparity and demonstrates the generality of the model. The six-year average projection errors are listed in Tables 1–3; more detailed results are presented in Appendix A.2, in terms of projection errors for each year, for males and females, respectively.

**Table 1.** Six-year (average) prediction errors for China.

		1985– 1991	1992– 1998	1999– 2005	2006– 2012	2013– 2019	Total
MSE	LC	0.0881	0.0965	0.0856	0.0799	0.0912	0.0883
	LL	0.0204	0.0284	0.0317	0.0363	0.0513	0.0336
	LSTM–ex	0.0092	0.0234	0.0245	0.0351	0.0501	0.0284
	LSTM–disp	0.0073 *	0.0166 *	0.0119 *	0.0083 *	0.0104 *	0.0109 *
MAE	LC	0.2105	0.2294	0.2129	0.215	0.2331	0.2202
	LL	0.1148	0.132	0.1294	0.1308	0.1543	0.1323
	LSTM–ex	0.0594	0.0942	0.1094	0.1383	0.1738	0.115
	LSTM–disp	0.0586 *	0.0849 *	0.0677 *	0.0569 *	0.0681 *	0.0672 *
RMSE	LC	0.2483	0.2906	0.2894	0.2818	0.289	0.2798
	LL	0.1388	0.1673	0.1733	0.1793	0.2014	0.172
	LSTM–ex	0.0959	0.1531	0.1565	0.1873	0.2238	0.1685
	LSTM–disp	0.0854 *	0.1288 *	0.1091 *	0.0911 *	0.1021 *	0.1044 *

\* indicates the smallest value of MSE/MAE/RMSE.

Tables 1–3 (see Figures A5–A7) show a clear cumulative error in long-term forecasts, which reveals the difficulty of long-term mortality forecasting, especially for the classical LC method. In our method, especially that based on lifespan disparity, this accumulation error is reduced to some extent, making long-term forecasting more reliable.

The results clearly show that the classical LC method underperforms, as it is based on only the historical mortality data of the target country or region. If the current mortality development trend in a developing country is not sustainable, mortality rates will gradually approach those of developed countries (like the benchmark group selected here). Hence, in such a setting, projections based solely on national mortality data are not reasonable. From the above results, the LSTM-based, time-varying LC method with lifespan disparity controlling the rotation in the time-dependent weights, is the most accurate one among the

four methods examined here, especially for long-term projections. On the other hand, it is interesting to observe that Nigeria has the most significant projection error reduction when transferring from the classical LC method to our LSTM-based time-varying LC method.

**Table 2.** Six-year (average) prediction errors for Brazil.

		1985– 1991	1992– 1998	1999– 2005	2006– 2012	2013– 2019	Total
MSE	LC	0.0094	0.0259	0.0848	0.1661	0.2691	0.1111
	Li-Lee	0.0061	0.0198	0.0647	0.1236	0.1999	0.0829
	LSTM-ex	0.0049	0.0195	0.0624	0.1125	0.1701	0.0739
	LSTM-disp	0.0038 *	0.0024 *	0.0157 *	0.0205 *	0.0566 *	0.0198 *
MAE	LC	0.0762	0.1191	0.2081	0.3084	0.4108	0.2245
	Li-Lee	0.0642	0.1033	0.1729	0.2442	0.3212	0.1811
	LSTM-ex	0.0506	0.1075	0.1996	0.2563	0.3161	0.1861
	LSTM-disp	0.0501 *	0.0324 *	0.0839 *	0.0837 *	0.1536 *	0.0807 *
RMSE	LC	0.0955	0.1461	0.2663	0.3924	0.5119	0.3239
	Li-Lee	0.0773	0.1335	0.2419	0.3457	0.4455	0.2845
	LSTM-ex	0.0697	0.1397	0.2496	0.3354	0.4121	0.2718
	LSTM-disp	0.0614 *	0.0488 *	0.1165 *	0.1239 *	0.2281 *	0.1332 *

\* indicates the smallest value of MSE/MAE/RMSE.

**Table 3.** Six-year (average) prediction errors for Nigeria.

		1985– 1991	1992– 1998	1999– 2005	2006– 2012	2013– 2019	Total
MSE	LC	0.0035	0.0082	0.0362	0.0961	0.1726	0.0633
	Li-Lee	0.0007	0.0022	0.0102	0.0286	0.0542	0.0192
	LSTM-ex	0.0045	0.0291	0.0334	0.0422	0.0521	0.0323
	LSTM-disp	0.0011 *	0.0008 *	0.0002 *	0.0006 *	0.0013 *	0.0008 *
MAE	LC	0.0432	0.0689	0.1631	0.2771	0.3778	0.1861
	Li-Lee	0.0209	0.0369	0.0835	0.1469	0.2059	0.0988
	LSTM-ex	0.0583	0.1582	0.1686	0.1955	0.2164	0.1594
	LSTM-disp	0.0247 *	0.0209 *	0.0102 *	0.0144 *	0.0203 *	0.0181 *
RMSE	LC	0.0507	0.0759	0.1825	0.3059	0.4116	0.2481
	Li-Lee	0.0224	0.0391	0.0959	0.1663	0.2302	0.1361
	LSTM-ex	0.0645	0.1681	0.1822	0.2045	0.2262	0.1796
	LSTM-disp	0.0306 *	0.0278 *	0.0121 *	0.0221 *	0.0331 *	0.0261 *

\* indicates the smallest value of MSE/MAE/RMSE.

#### 4.5. Results for LDR, LDRexChina, LDRexLDC

Finally, in order to demonstrate the projection accuracy of our method, the following illustration is for the mortality data of three developing regions, denoted as less developed region(s) (LDR), less developed regions excluding China (LDRexChina), and less developed regions excluding least developed country (LDRexLDC); see Tables 4–6 and also Figures A8–A10 in Appendix A.3).

Note that the error fluctuations in the prediction results for most of these less developed regions are reduced significantly (see Figures A8–A10), which is reasonable since the less developed regions contain larger populations (i.e., more stable) compared to individual countries. Overall, the LSTM-based, time-varying LC method with lifespan disparity as the control of the rotation in time-varying weights provides the most accurate projections within the four methods examined here.

It is worth mentioning that both our LSTM-based-time-varying LC method and the Li-Lee method incorporate mortality trend corrections based on a benchmark group. However, the empirical results show that, for developing countries or regions, such corrections are better modeled through the projection of life expectancy or lifespan disparity difference with an LSTM model, especially for long-term forecasts.



**Table 4.** Six-year (average) prediction errors for LDR.

		1985– 1991	1992– 1998	1999– 2005	2006– 2012	2013– 2019	Total
MSE	LC	0.0094	0.0307	0.0675	0.1033	0.1578	0.0738
	LL	0.0032	0.0114	0.0293	0.0524	0.0829	0.0358
	LSTM–ex	0.0011 *	0.00905	0.02425	0.0402	0.05095	0.0251
	LSTM–disp	0.0019	0.00135 *	0.0008 *	0.0031 *	0.00335 *	0.0021 *
MAE	LC	0.0808	0.1537	0.2269	0.2772	0.3472	0.2172
	LL	0.0441	0.0906	0.1408	0.1847	0.2443	0.1409
	LSTM–ex	0.0253*	0.071	0.1161	0.1576	0.1807	0.1101
	LSTM–disp	0.0344	0.0275 *	0.0216 *	0.0427 *	0.0372 *	0.033 *
RMSE	LC	0.0969	0.1736	0.2585	0.3223	0.397	0.2497
	LL	0.0529	0.1078	0.1688	0.2273	0.2876	0.1689
	LSTM–ex	0.0322 *	0.0897	0.1511	0.1975	0.2229	0.1387
	LSTM–disp	0.0433	0.0349 *	0.0285 *	0.0554 *	0.0549 *	0.0434 *

\* indicates the smallest value of MSE/MAE/RMSE.

**Table 5.** Six-year (average) prediction errors for LDRexChina.

		1985– 1991	1992– 1998	1999– 2005	2006– 2012	2013– 2019	Total
MSE	LC	0.0091	0.0271	0.0711	0.1358	0.2241	0.0934
	Li-Lee	0.0037	0.0126	0.0365	0.0738	0.1249	0.0504
	LSTM–ex	0.0008 *	0.0056	0.0146	0.0202	0.0205	0.0123
	LSTM–disp	0.0023	0.0008 *	0.0011 *	0.0031 *	0.0032 *	0.0021 *
MAE	LC	0.0556	0.1115	0.1861	0.2527	0.3404	0.1893
	Li-Lee	0.0732	0.1378	0.2191	0.3067	0.4022	0.2278
	LSTM–ex	0.0233 *	0.0562	0.0898	0.1136	0.1115	0.0789
	LSTM–disp	0.0368	0.0221 *	0.0222 *	0.0411 *	0.0369 *	0.0317 *
RMSE	LC	0.0938	0.1643	0.2664	0.3677	0.4727	0.3051
	Li-Lee	0.0604	0.1112	0.1889	0.2692	0.3511	0.2227
	LSTM–ex	0.0283 *	0.0747	0.1203	0.1422	0.1428	0.1109
	LSTM–disp	0.0476	0.0283 *	0.0332 *	0.0541 *	0.0516 *	0.0447 *

\* indicates the smallest value of MSE/MAE/RMSE.

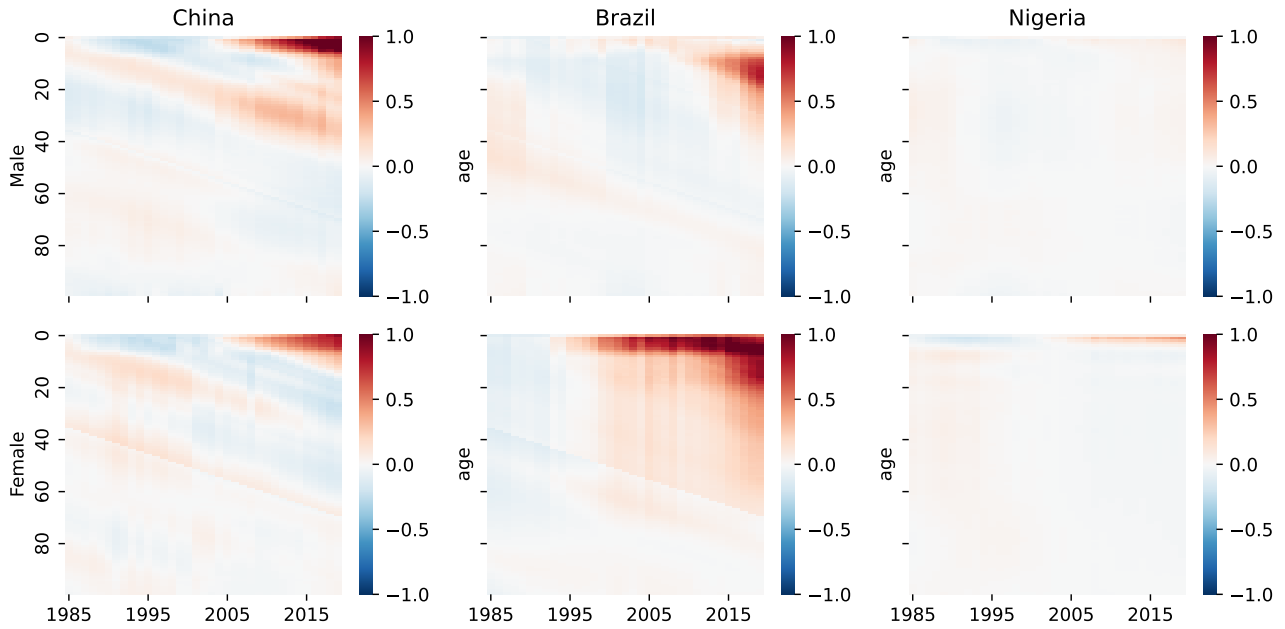
**Table 6.** Six-year (average) prediction errors for LDRexLDC.

		1985– 1991	1992– 1998	1999– 2005	2006– 2012	2013– 2019	Total
MSE	LC	0.0199	0.0328	0.05	0.0733	0.1025	0.0557
	LL	0.0074	0.0126	0.0182	0.0306	0.0454	0.0228
	LSTM–ex	0.0014 *	0.0065	0.0254	0.0493	0.064	0.0293
	LSTM–disp	0.0044	0.0032 *	0.0009 *	0.0061 *	0.0022 *	0.0034 *
MAE	LC	0.1183	0.1675	0.203	0.2406	0.2903	0.204
	LL	0.0722	0.0983	0.1165	0.1485	0.1897	0.1251
	LSTM–ex	0.0301 *	0.059	0.1176	0.1726	0.1956	0.115
	LSTM–disp	0.0551	0.045 *	0.023 *	0.0639 *	0.037 *	0.0448 *
RMSE	LC	0.1399	0.181	0.2224	0.2689	0.32	0.2264
	LL	0.0855	0.1073	0.1313	0.1714	0.2117	0.1414
	LSTM–ex	0.0368 *	0.0772	0.1582	0.2217	0.2526	0.1493
	LSTM–disp	0.0661	0.0541 *	0.0297 *	0.0782 *	0.0467	0.055 *

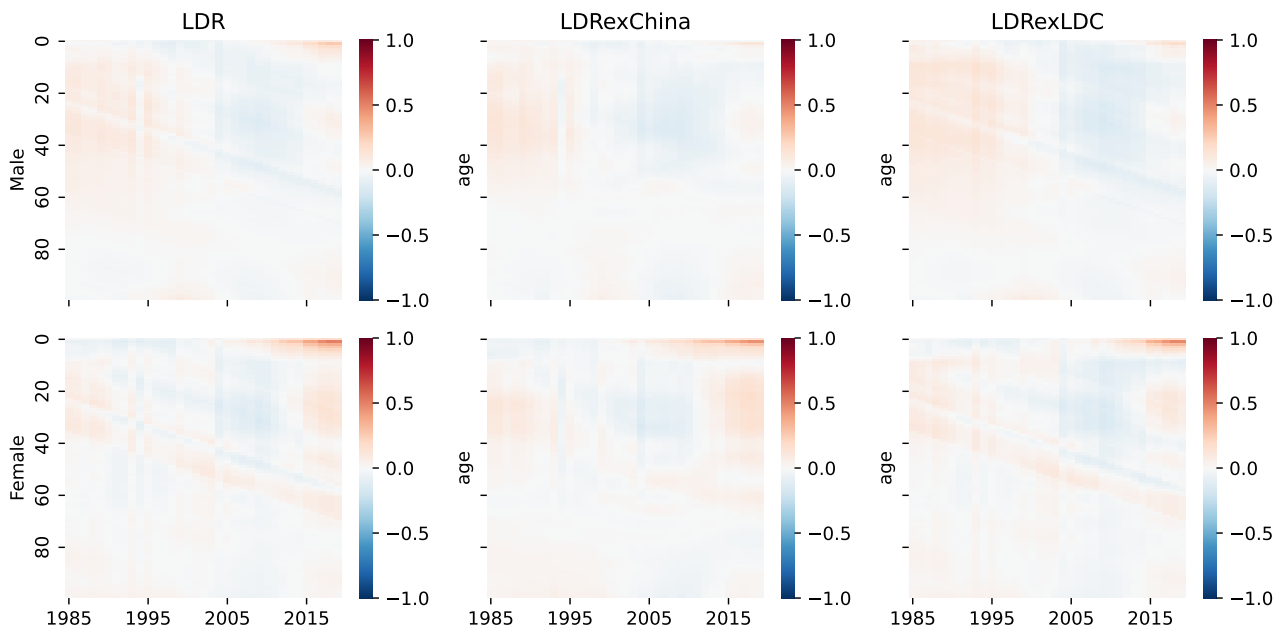
\* indicates the smallest value of MSE/MAE/RMSE.

To end this section, we draw heatmaps that show the relative prediction errors (i.e., (predicted value – actual value)/actual value) across all ages and years for the out-of-sample data. The results are presented in Figures 3 and 4. For most ages and years in the six target developing countries/regions (except young males in China and females in Brazil),

our model performs well. However, we also observe some cohort effects in the results, especially for the data from China. A possible improvement could be to extend (11) and (12) to include age dependency in our model. This is a non-trivial extension that will be left for future studies.



**Figure 3.** Relative prediction errors for three developing countries (males at the **top**, females at the **bottom**).



**Figure 4.** Relative prediction errors for three developing regions (males at the **top**, females at the **bottom**).

## 5. Conclusions

Mortality improvements are linked to social progress, for instance, in terms of health, nutrition, education, hygiene, and access to medical assistance. It is difficult to accurately predict mortality development trends, especially over a long-term period. For developing countries or regions, it is particularly important to provide accurate long-term predictions of mortality rates for each age in the population, given that the current mortality data might not reveal sustainable development trends in the long-run.

The proposal here is an LSTM-based coherent mortality forecasting method for developing countries, where the life expectancy and lifespan disparity gaps between the target developing country and the selected benchmark group are used for long-term projections. In particular, we allow the mortality development pattern of a developing country to be a weighted average of trends generated by its own historical data and the selected benchmark group. And the rotation in the time-varying weights is controlled by the projected life expectancy and lifespan disparity gaps between the developing country and the benchmark group. In addition, we introduce a unified deep neural network model with an LSTM architecture for the long-term forecasting of the gaps in life expectancy and lifespan disparity for all six developing countries and regions in our analysis.

We apply this LSTM-based coherent mortality forecasting method to three developing countries, China, Brazil, and Nigeria, and three developing regions defined by the United Nations, namely LDR, LDRexChina, and LDRexLDC. The empirical results show that the LSTM-based coherent forecasting method with lifespan disparity outperforms the classical LC and Li-Lee methods, as well as the one with life expectancy, especially for long-term projections.

**Author Contributions:** Conceptualization, J.G., Y.S. and R.X.; methodology, Y.S., R.X.; software, Y.S. and R.X.; validation, J.G., Y.S. and R.X.; formal analysis, Y.S. and R.X.; investigation, Y.S. and R.X.; resources, J.G., Y.S. and R.X.; data curation, Y.S. and R.X.; writing—original draft preparation, Y.S. and R.X.; writing—review and editing, J.G., Y.S. and R.X.; visualization, Y.S. and R.X.; supervision, J.G. and R.X.; project administration, J.G. and R.X.; funding acquisition, J.G. and R.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Natural Sciences and Engineering Research Council (NSERC) of Canada grant numbers RGPIN-2017-06643 and DGDND-2017-00096 (Garrido and Xu, during his stay at Concordia University), by the National Natural Science Foundation of China grant numbers 12201506 and 72171055, and by the XJTLU Research Development Funding grant number RDF-20-01-02 (Shang and Xu at XJTLU).

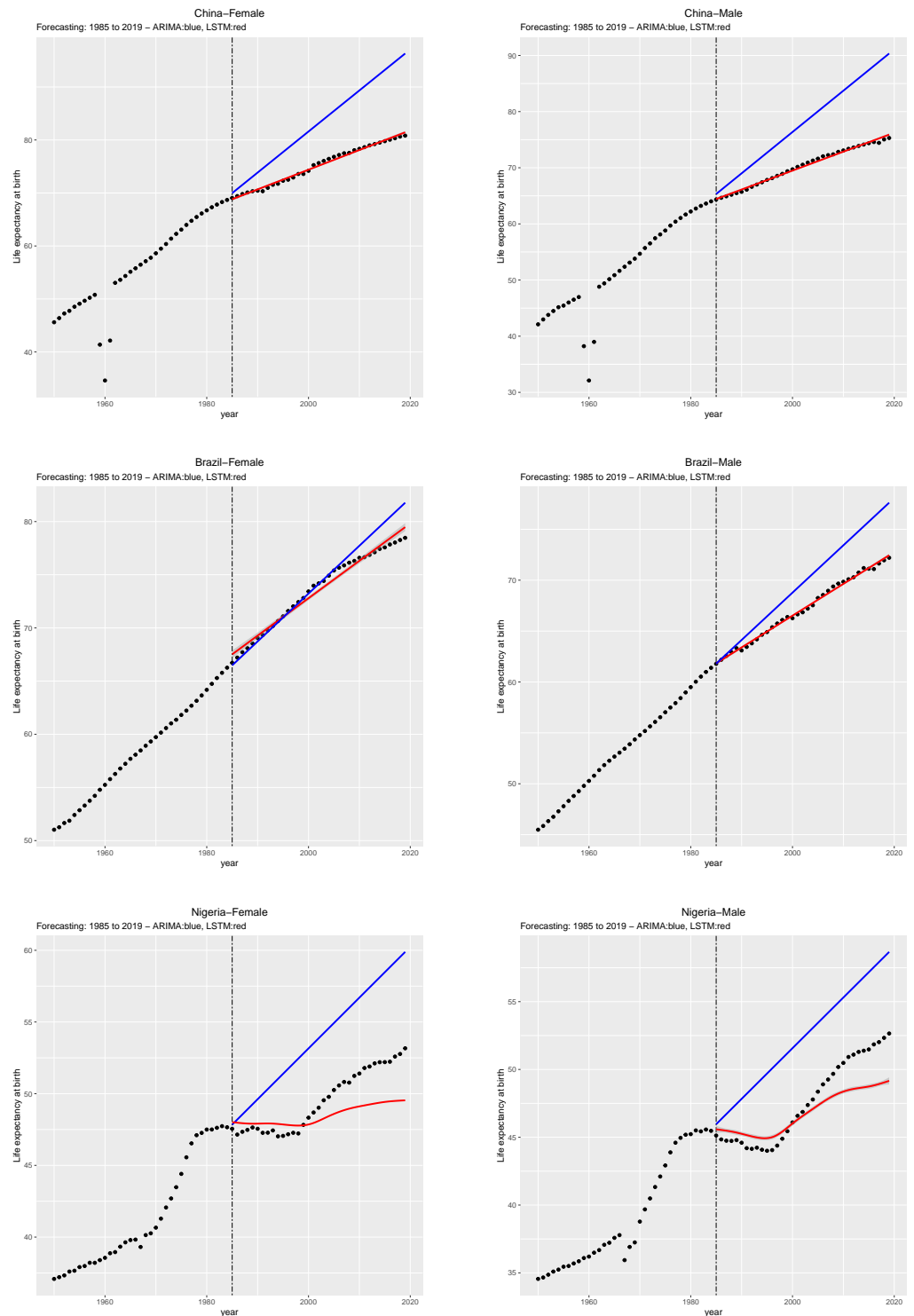
**Data Availability Statement:** The mortality data for the benchmark group is publicly available on the Human Mortality Database (<https://www.mortality.org/>, accessed on 21 November 2023); the mortality data for the developing countries and regions are publicly available on United Nations data source (<https://population.un.org/wpp/Download/Standard/Mortality/>, accessed on 21 November 2023)

**Acknowledgments:** The authors thank the three anonymous referees for their constructive comments, which helped improve the manuscript.

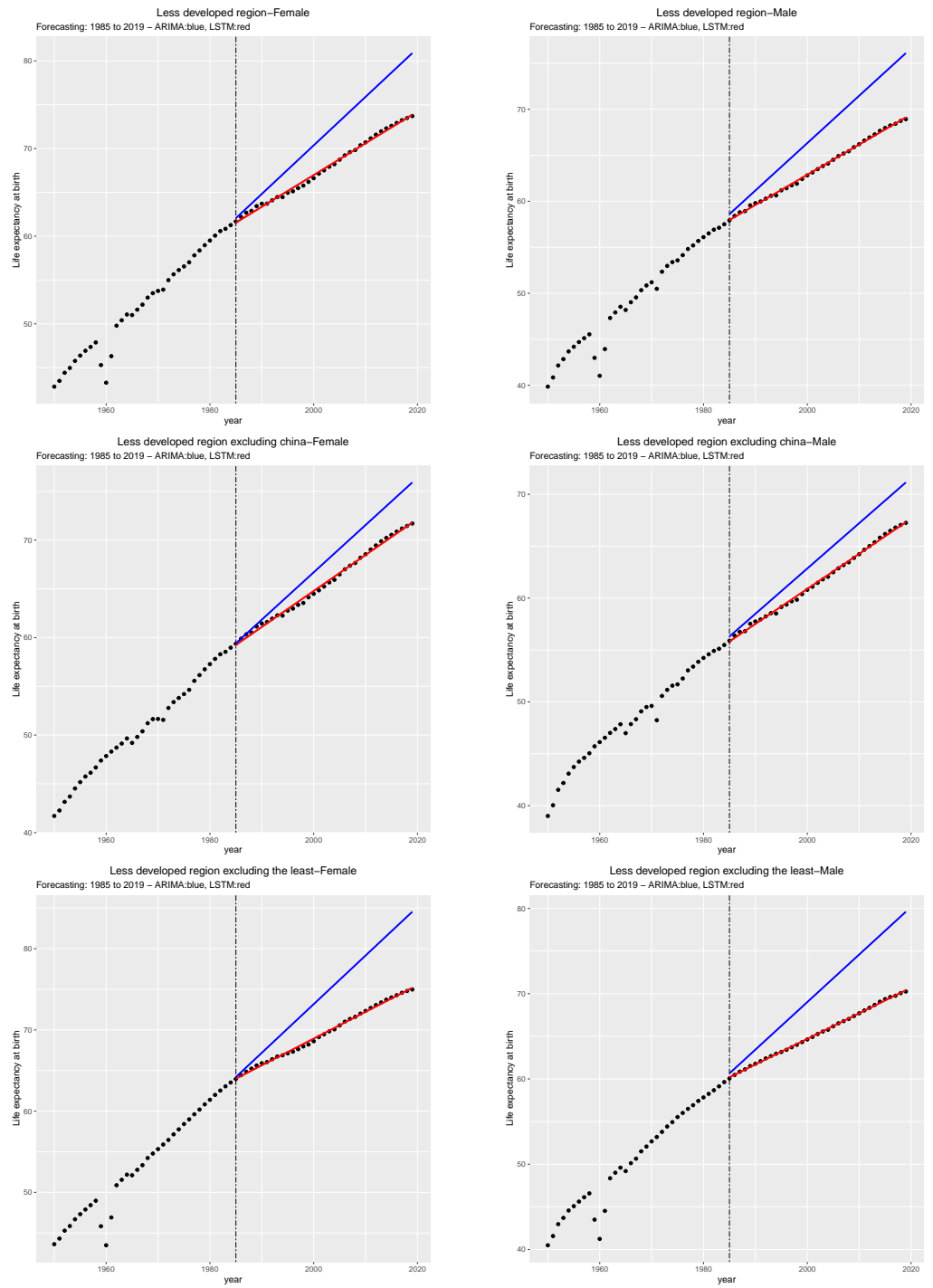
**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

### Appendix A.1. Life Expectancy and Lifespan Disparity Forecasts

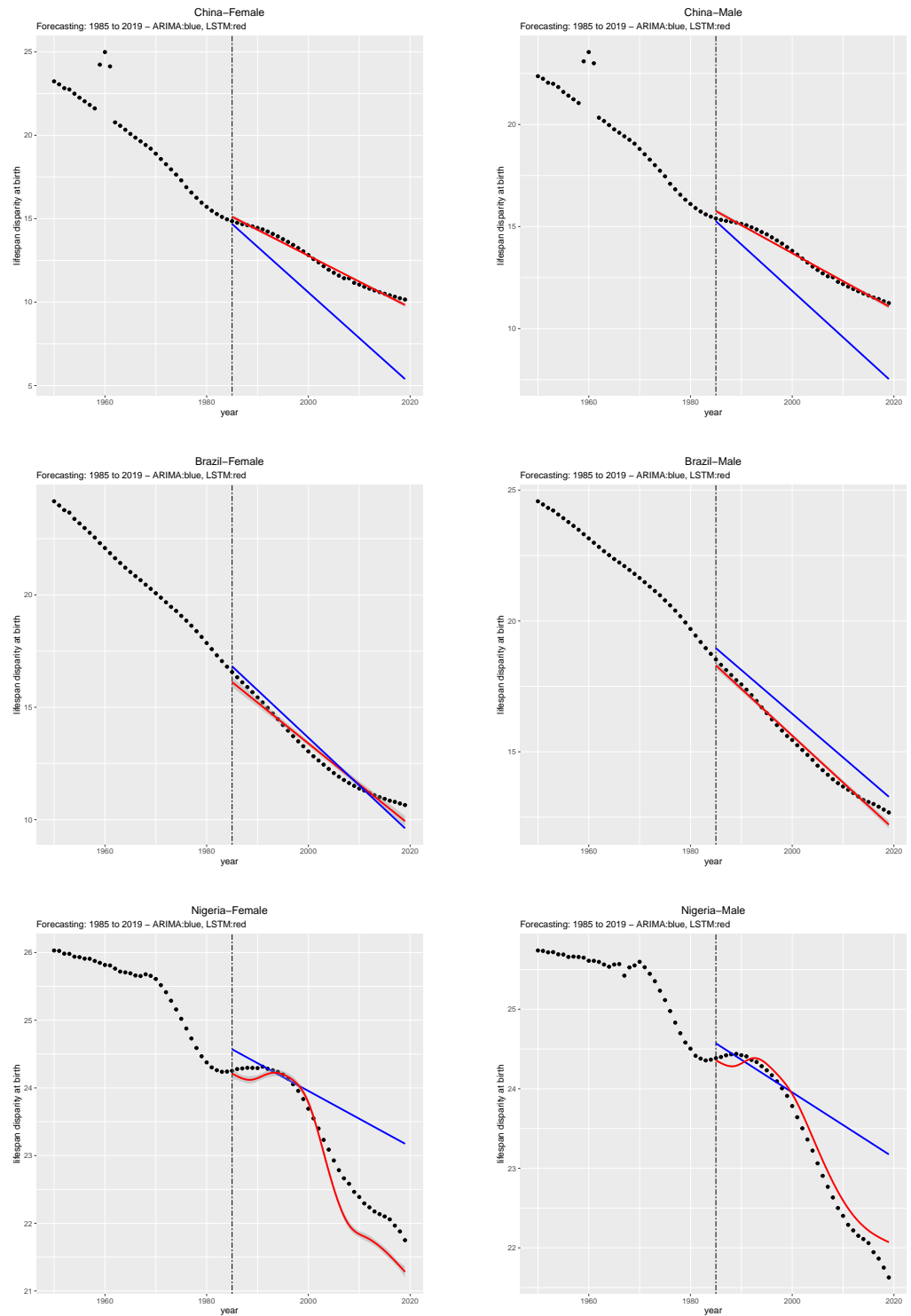


**Figure A1.** Historical (dotted lines) and forecast (blue for ARIMA, red for LSTM) values of  $e_{0,t}$  for three target countries.

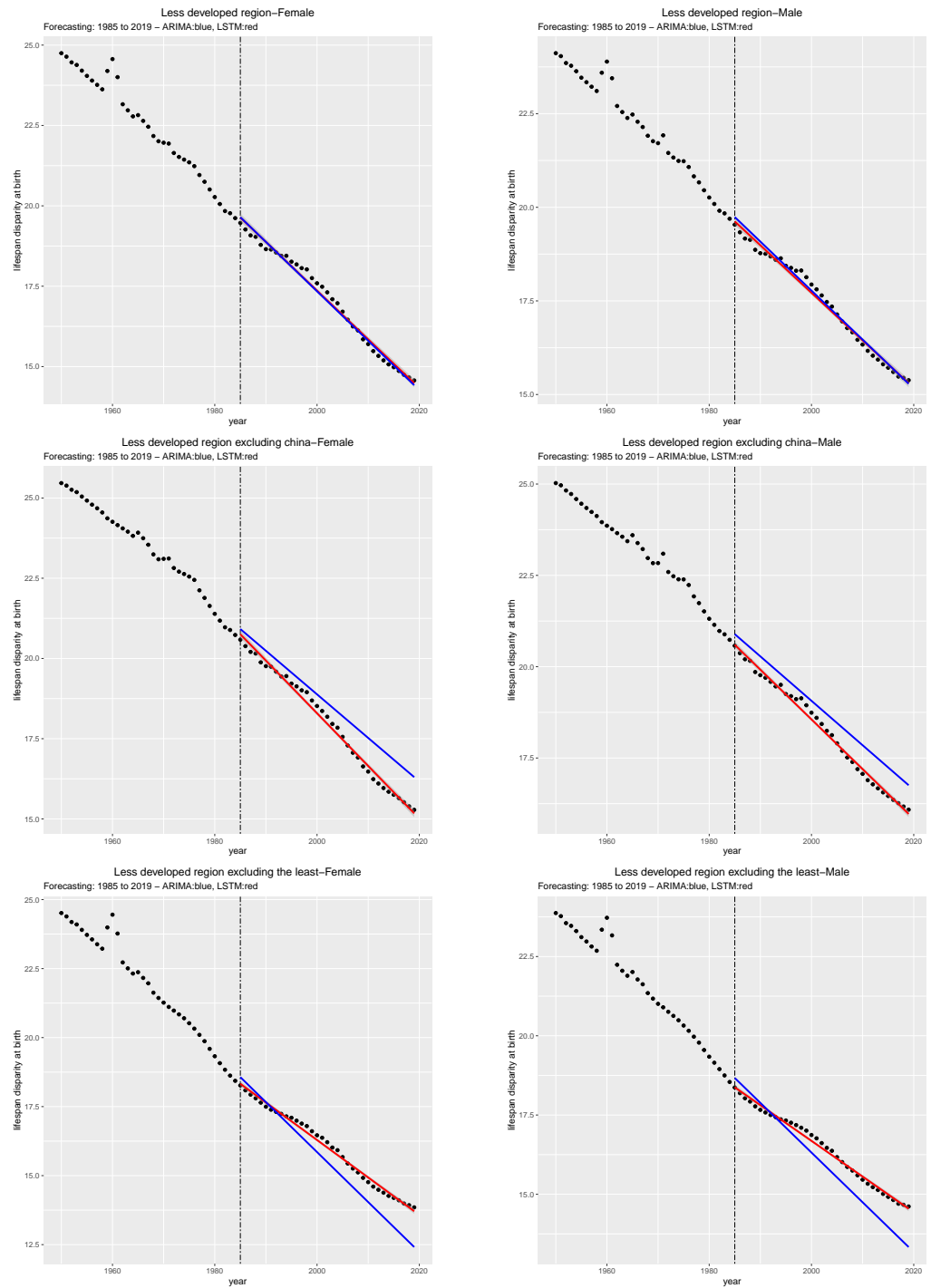


**Figure A2.** Historical (dotted lines) and forecast (blue for ARIMA, red for LSTM) values of  $e_{0,t}$  for three target regions.





**Figure A3.** Historical (dotted lines) and forecast (blue for ARIMA, red for LSTM) values of  $e_{0,t}^+$  for three target countries.



**Figure A4.** Historical (dotted lines) and forecast (blue for ARIMA, red for LSTM) values of  $e_{0,t}^+$  for three target regions.

Appendix A.2. Projection Errors for China, Brazil, and Nigeria

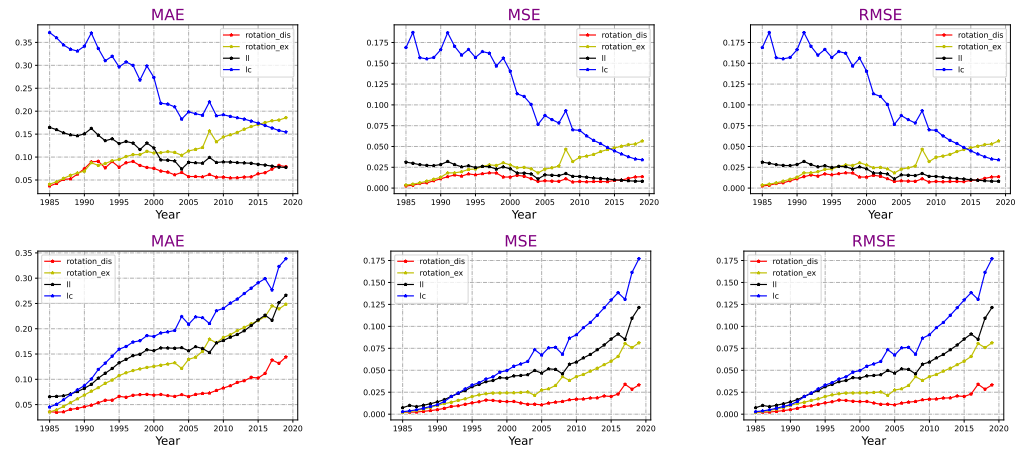


Figure A5. Forecasting errors by year for China (females at the top, males at the bottom).

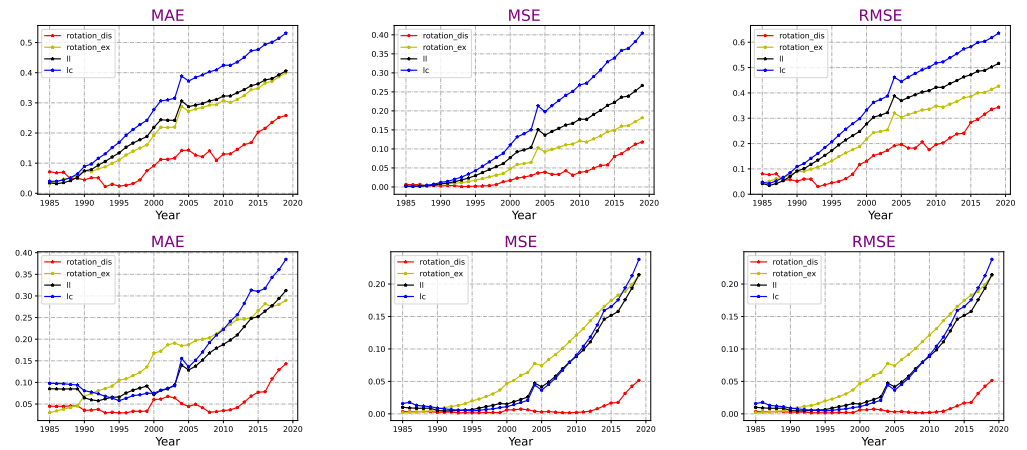


Figure A6. Forecasting errors by year for Brazil (females at the top, males at the bottom).

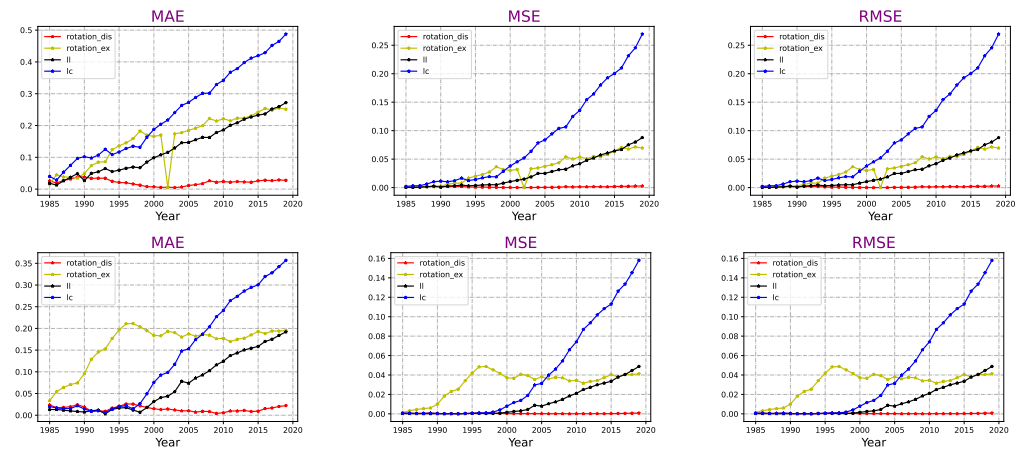


Figure A7. Forecasting errors by year for Nigeria (females at the top, males at the bottom).

Appendix A.3. Projection Errors for LDR, LDRexChina, and LDRexLDC

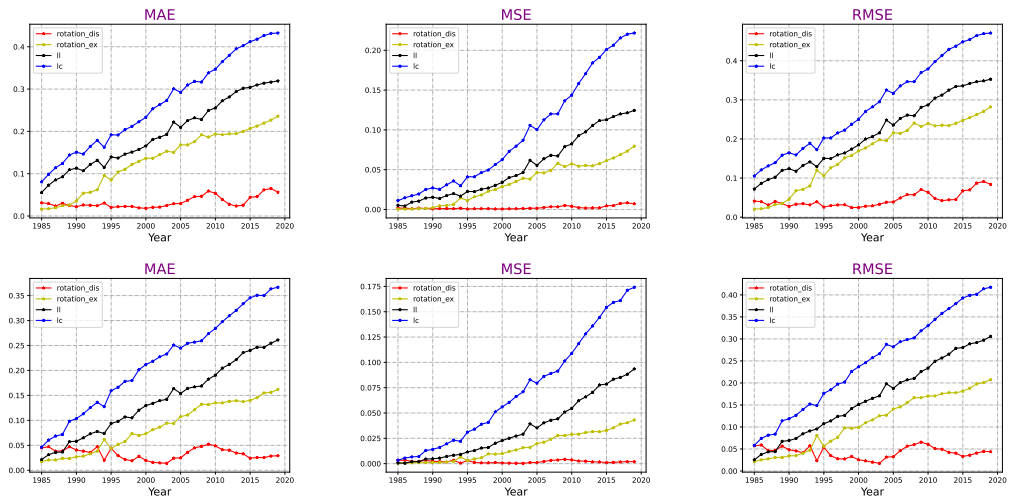


Figure A8. Forecasting errors by year for less developed regions (LDR) (females at the top, males at the bottom).

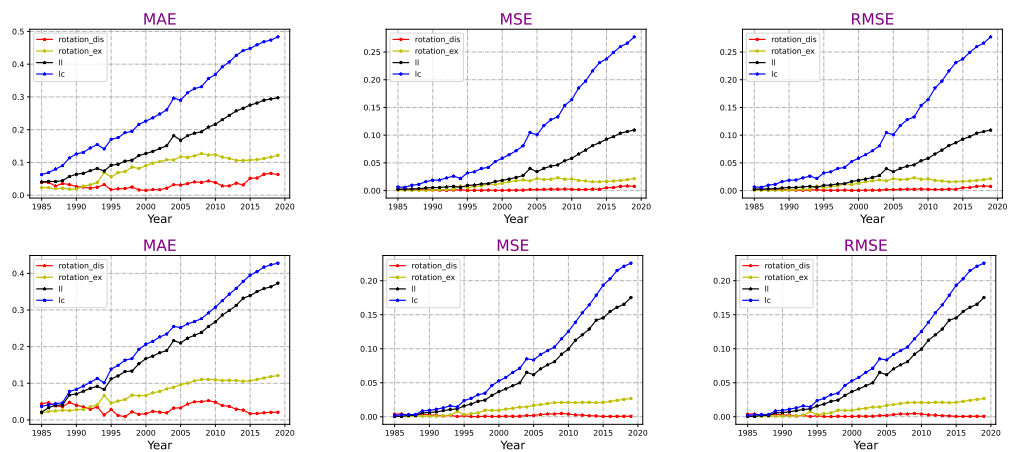


Figure A9. Forecasting errors by year for less developed regions, excluding China (LDRexChina) (females at the top, males at the bottom).

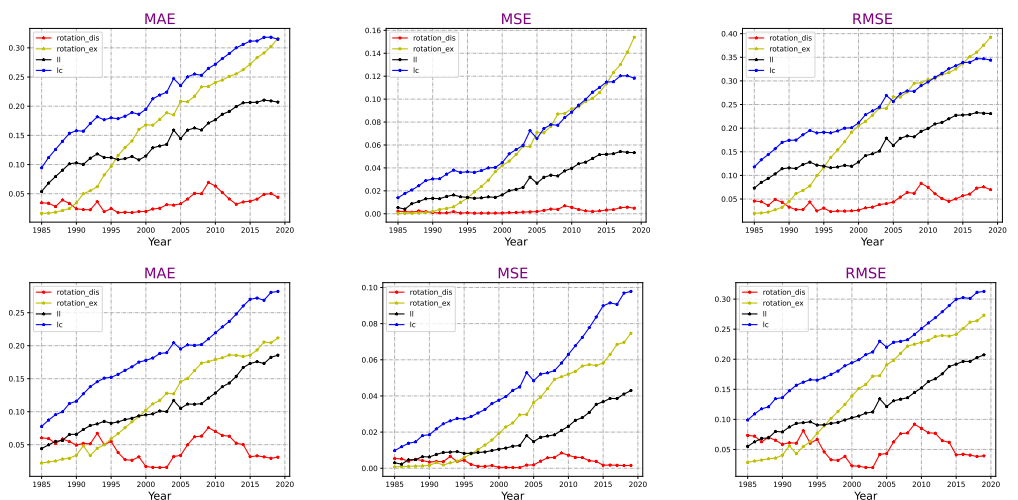


Figure A10. Forecasting errors by year for less developed regions, excluding least developed country (LDRexLDC) (females at the top, males on the bottom).

## References

- Aburto, José Manuel, Francisco Villavicencio, Ugofilippo Basellini, Søren Kjærgaard, and James W. Vaupel. 2020. Dynamics of life expectancy and life span equality. *Proceedings of the National Academy of Sciences* 117: 5250–59. [CrossRef]
- Antonio, Katrien, Anastasios Bardoutsos, and Wilbert Ouburg. 2015. Bayesian Poisson log-bilinear models for mortality projections with multiple populations. *European Actuarial Journal* 5: 245–81. [CrossRef]
- Austin, Kelly F., and Laura A. McKinney. 2012. Disease, war, hunger, and deprivation: A cross-national investigation of the determinants of life expectancy in less-developed and sub-Saharan African nations. *Sociological Perspectives* 55: 421–47. [CrossRef]
- BBVA. 2014. Eagles Economic Outlook Annual Report. Available online: [https://www.bbva.com/wp-content/uploads/2014/05/2014\\_EAGLEs\\_Economic\\_Outlook-Annual.pdf](https://www.bbva.com/wp-content/uploads/2014/05/2014_EAGLEs_Economic_Outlook-Annual.pdf) (accessed on 13 January 2024).
- Bohk-Ewald, Christina, Marcus Ebeling, and Roland Rau. 2017. Lifespan disparity as an additional indicator for evaluating mortality forecasts. *Demography* 54: 1559–77. [CrossRef]
- Boonen, Tim J., and Hong Li. 2017. Modeling and forecasting mortality with economic growth: A multipopulation approach. *Demography* 54: 1921–46. [CrossRef]
- Brouhns, Natacha, Michel Denuit, and Jeroen K Vermunt. 2002. A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics* 31: 373–93. [CrossRef]
- Cairns, Andrew J. G., David Blake, and Kevin Dowd. 2006. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance* 73: 687–718. [CrossRef]
- Cairns, Andrew J. G., David Blake, and Kevin Dowd. 2008. Modelling and management of mortality risk: A review. *Scandinavian Actuarial Journal* 2008: 79–113. [CrossRef]
- Cairns, Andrew J. G., David Blake, Kevin Dowd, Guy D. Coughlan, David Epstein, and Marwa Khalaf-Allah. 2011a. Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics* 48: 355–67. [CrossRef]
- Cairns, Andrew J. G., David Blake, Kevin Dowd, Guy D. Coughlan, and Marwa Khalaf-Allah. 2011b. Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin: The Journal of the IAA* 41: 29–59.
- Carter, Lawrence R., and Ronald D. Lee. 1992. Modeling and forecasting US sex differentials in mortality. *International Journal of Forecasting* 8: 393–411. [CrossRef]
- Currie, Iain D. 2016. On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal* 2016: 356–83. [CrossRef]
- Czado, Claudia, Antoine Delwarde, and Michel Denuit. 2005. Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics* 36: 260–84. [CrossRef]
- Deprez, Philippe, Pavel V. Shevchenko, and Mario V. Wüthrich. 2017. Machine learning techniques for mortality modeling. *European Actuarial Journal* 7: 337–52. [CrossRef]
- Gao, Guangyuan, and Yanlin Shi. 2021. Age-coherent extensions of the Lee–Carter model. *Scandinavian Actuarial Journal* 2021: 998–1016. [CrossRef]
- Hainaut, Donatien. 2018. A neural-network analyzer for mortality forecast. *ASTIN Bulletin: The Journal of the IAA* 48: 481–508. [CrossRef]
- Hanewald, Katja. 2011. Explaining mortality dynamics: The role of macroeconomic fluctuations and cause of death trends. *North American Actuarial Journal* 15: 290–314. [CrossRef]
- Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9: 1735–80. [CrossRef]
- Jeuland, Marc A., David E. Fuente, Semra Ozdemir, Maura C. Allaire, and Dale Whittington. 2013. The long-term dynamics of mortality benefits from improved water and sanitation in less developed countries. *PLoS ONE* 8: e74804. [CrossRef] [PubMed]
- Lee, Ronald D. 2000. The Lee–Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal* 4: 80–91. [CrossRef]
- Lee, Ronald D., and Lawrence R. Carter. 1992. Modeling and forecasting us mortality. *Journal of the American Statistical Association* 87: 659–71.
- Lee, Ronald D., and Francois Nault. 1993. Modeling and forecasting provincial mortality in Canada. Paper presented at the World Congress of the IUSSP, Montreal, QC, Canada, August 24–September 1.
- Levantesi, Susanna, and Virginia Pizzorusso. 2019. Application of machine learning to mortality modeling and forecasting. *Risks* 7: 26. [CrossRef]
- Li, Hong, and Yang Lu. 2017. Coherent forecasting of mortality rates: A sparse vector-autoregression approach. *ASTIN Bulletin: The Journal of the IAA* 47: 563–600. [CrossRef]
- Li, Hong, and Yang Lu. 2018. A Bayesian non-parametric model for small population mortality. *Scandinavian Actuarial Journal* 2018: 605–28. [CrossRef]
- Li, Hong, Yang Lu, and Pintao Lyu. 2018. *Modeling and Forecasting Chinese Population Dynamics in a Multi-Population Context*. SOA Research Reports. Schaumburg: Society of Actuaries.
- Li, Hong, Yang Lu, and Pintao Lyu. 2021. Coherent mortality forecasting for less developed countries. *Risks* 9: 151. [CrossRef]
- Li, Johnny Siu-Hang, Kenneth Q. Zhou, Xiaobai Zhu, Wai-Sum Chan, and Felix Wai-Hon Chan. 2019. A Bayesian approach to developing a stochastic mortality model for China. *Journal of the Royal Statistical Society Series A: Statistics in Society* 182: 1523–60. [CrossRef]



- Li, Nan, and Ronald D. Lee. 2005. Coherent mortality forecasts for a group of populations: An extension of the Lee–Carter method. *Demography* 42: 575–94. [[CrossRef](#)] [[PubMed](#)]
- Li, Nan, Ronald D. Lee, and Patrick Gerland. 2013. Extending the lee-Carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography* 50: 2037–51. [[CrossRef](#)]
- Lindholm, Mathias, and Lina Palmborg. 2022. Efficient use of data for LSTM mortality forecasting. *European Actuarial Journal* 12: 749–778. [[CrossRef](#)]
- Ma, Qingxiao, and Tim J. Boonen. 2023. Longevity risk modeling with the consumer price index. *North American Actuarial Journal* 1–18. [[CrossRef](#)]
- Marino, Mario, Susanna Levantesi, and Andrea Nigri. 2023. A neural approach to improve the Lee–Carter mortality density forecasts. *North American Actuarial Journal* 27: 148–165. [[CrossRef](#)]
- Mozer, Michael C. 1991. Induction of multiscale temporal structure. Paper presented at the Advances in Neural Information Processing Systems 4, NIPS Conference, Denver, CO, USA, December 2–5.
- Müller, Olaf, and Michael Krawinkel. 2005. Malnutrition and health in developing countries. *Cmaj* 173: 279–86. [[CrossRef](#)]
- Nigri, Andrea, Susanna Levantesi, and Mario Marino. 2021. Life expectancy and lifespan disparity forecasting: A long short-term memory approach. *Scandinavian Actuarial Journal* 2021: 110–33. [[CrossRef](#)]
- Nigri, Andrea, Susanna Levantesi, Mario Marino, Salvatore Scognamiglio, and Francesca Perla. 2019. A deep learning integrated Lee–Carter model. *Risks* 7: 33. [[CrossRef](#)]
- Niu, Geng, and Bertrand Melenberg. 2014. Trends in mortality decrease and economic growth. *Demography* 51: 1755–73. [[CrossRef](#)]
- O’Hare, Colin, and Youwei Li. 2012. Explaining young mortality. *Insurance: Mathematics and Economics* 50: 12–25. [[CrossRef](#)]
- Pascanu, Razvan, Tomas Mikolov, and Yoshua Bengio. 2013. On the difficulty of training recurrent neural networks. Paper presented at International Conference on Machine Learning, Atlanta, GA, USA, June 16–21, pp. 1310–18.
- Pedroza, Claudia. 2006. A Bayesian forecasting model: Predicting US male mortality. *Biostatistics* 7: 530–50. [[CrossRef](#)] [[PubMed](#)]
- Pitacco, Ermanno. 2004. Survival models in a dynamic context: A survey. *Insurance: Mathematics and Economics* 35: 279–98. [[CrossRef](#)]
- Renshaw, Arthur E., and Steven Haberman. 2003. Lee–Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics* 33: 255–72. [[CrossRef](#)]
- Renshaw, Arthur E., and Steven Haberman. 2006. A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics* 38: 556–70. [[CrossRef](#)]
- Richman, Ronald, and Mario V. Wüthrich. 2021. A neural network extension of the Lee–Carter model to multiple populations. *Annals of Actuarial Science* 15: 346–66. [[CrossRef](#)]
- Schnürch, Simo, and Ralf Korn. 2022. Point and interval forecasts of death rates using neural networks. *ASTIN Bulletin: The Journal of the IAA* 52: 333–60. [[CrossRef](#)]
- Vaupel, James W., and Vladimir Canudas Romo. 2003. Decomposing change in life expectancy: A bouquet of formulas in honor of Nathan Keyfitz’s 90th birthday. *Demography* 40: 201–16. [[CrossRef](#)]
- Vaupel, James W., and Sabine Schnabel. 2004. Forecasting best-practice life expectancy to forecast national life expectancy. Paper presented at the 2004 Annual Meeting of the Population Association of America, Boston, MA, USA, April 1–3.
- Wang, Chou-Wen, Jinggong Zhang, and Wenjun Zhu. 2021. Neighbouring prediction for mortality. *ASTIN Bulletin: The Journal of the IAA* 51: 689–718. [[CrossRef](#)]
- White, Kevin M. 2002. Longevity advances in high-income countries, 1955–96. *Population and Development Review* 28: 59–76. [[CrossRef](#)]
- Williams, Ronald J., and David Zipser. 1995. Gradient-based learning algorithms for recurrent networks and their computational complexity. In *Back-Propagation: Theory, Architectures, and Applications*, 1st ed. Edited by Yves Chauvin and David E Rumelhart. London: Psychology Press, Taylor & Francis Group, chp. 13. [[CrossRef](#)]
- Wilmoth, John R. 1998. Is the pace of Japanese mortality decline converging toward international trends? *Population and Development Review* 24: 593–600. [[CrossRef](#)]
- Wong, Jackie S. T., Jonathan J. Forster, and Peter W. F. Smith. 2023. Bayesian model comparison for mortality forecasting. *Journal of the Royal Statistical Society Series C: Applied Statistics* 72: 566–86. [[CrossRef](#)]
- Wong-Fupuy, Carlos, and Steven Haberman. 2004. Projecting mortality trends: Recent developments in the United Kingdom and the United States. *North American Actuarial Journal* 8: 56–83. [[CrossRef](#)]
- Zhang, Zhen, and James W. Vaupel. 2009. The age separating early deaths from late deaths. *Demographic Research* 20: 721–30. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.