DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 50, ARTICLE 28, PAGES 797–826
PUBLISHED 25 APRIL 2024
http://www.demographic-research.org/Volumes/Vol50/28/
DOI: 10.4054/DemRes.2024.50.28

Research Article

Mortality modelling with arrival of additional year of mortality data: Calibration and forecasting

Kam Kuen Mok **Chong It Tan**

Jinhui Zhang **Yanlin Shi**

This publication is part of the Special Collection in the Memory of Professor James W Vaupel (1945–2022), founder and long-time publisher of Demographic Research. The Special Collection is edited by Jakub Bijak, Griffith Feeney, Nico Keilman, and Carl Schmertmann.

Contents

Demographic Research: Volume 50, Article 28
Research Article

Mortality modelling with arrival of additional year of mortality data: Calibration and forecasting

Kam Kuen Mok[1]

Chong It Tan[2]

Jinhui Zhang[3]

Yanlin Shi[4]

Abstract

BACKGROUND
For commonly used mortality models, the existing estimates change with the recalibration of new data. This issue is also known as the lack of the new-data-invariant property.

OBJECTIVE
We adapt the Lee–Carter, age-period-cohort, Renshaw–Haberman, and Li–Lee models to achieve the new-data-invariant property. The resulting fitted or forecast mortality indexes are tractable and comparable when more recent data are modelled.

METHODS
Illustrated by mortality rates of the England and Wales populations, we explore the trade-off between goodness of fit and the new-data-invariant property. Using the adapted model and vector autoregressive framework, we explore the interdependencies of subregional mortality dynamics in the United Kingdom.

RESULTS
To compare the goodness of fit, we consider the four adapted models and the Cairns–Blake–Dowd model, which are invariant to new data without adaptation. The Renshaw–Haberman model is demonstrated to be the best-performing model. The in-sample and

[1] Department of Actuarial Studies and Business Analytics, Macquarie Business School, Macquarie University, Sydney, Australia.
[2] Department of Actuarial Studies and Business Analytics, Macquarie Business School, Macquarie University, Syndey, Australia.
[3] Department of Actuarial Studies and Business Analytics, Macquarie Business School, Macquarie University, Sydney, Australia. Email: colin.zhang@mq.edu.au.
[4] Department of Actuarial Studies and Business Analytics, Macquarie Business School, Macquarie University, Sydney, Australia.

backtesting results show that the proposed adaptation introduces only a small cost of reduced model fitting, which is robust across sensitivity analyses.

CONCLUSIONS
The adapted Renshaw–Haberman model is recommended to construct tractable mortality indexes.

CONTRIBUTION
From a methodological perspective, we adopt popular models to achieve a desirable new-data-invariant property. Our empirical results suggest that the adapted model can provide reliable forecast of mortality rates for use in demographic research.

1. Introduction

The seminal work of Vaupel, Villavicencio, and Bergeron-Boucher (2021) finds that, for the countries doing the best, life expectancy has increased by roughly 2.5 years per decade since 1840. Such a continuously increasing trend has a significant effect on various aspects of our society, such as healthcare systems, pension plans, and the insurance industry. Vaupel, Villavicencio, and Bergeron-Boucher (2021) also point out the uncertainty of future life expectancy, which may worsen the aging and longevity risks that many countries are already facing. Influential demographic research has been conducted to combat uncertainty and risks. For instance, Barbi et al. (2018) discuss the existence of extreme-age mortality plateaus, which helps to understand the biological limit of the human lifespan. Thinggaard et al. (2020) examine the survival status of the oldest ages and conclude that lifespan and health increase among these populations.

The accuracy of forecasting future life expectancy is attributed to the accuracy of projected future mortality rates. Such rates are usually obtained after calibrating historical experiences through the application of mortality models. In the mortality literature, the most celebrated model is proposed in the seminal work of Lee and Carter (1992). In addition, the Lee–Carter (LC) model assumes that the future mortality trend is a continuation of the past, as captured by the estimated time-varying parameters. Over the past few decades, numerous studies have focused on the extension of the LC model. As reviewed by Booth and Tickle (2008), some earlier extensions include adjustment in the temporal factor by matching life expectancy (Lee and Miller 2001), determining an optimum fitting period (Booth, Maindonald, and Smith 2002), incorporating more sophisticated time series specification in the forecasting of the temporal factor (Renshaw and Haberman 2003), and adopting a Poisson log-bilinear model in the estimation (Brouhns, Denuit, and Vermunt 2002). More recent extensions have attempted to simplify jump-off rates (Bergeron-Boucher et al. 2017), adopt a machine-learning method for temporal factor

forecasting (Marino, Levantesi, and Nigri 2023), and address the coherence issue within a multi-population framework (Kjærgaard et al. 2020; for a more detailed review of such recent developments, see Basellini, Camarda, and Booth 2023).

Under the LC framework, forecasting of the temporal factor is performed via a usual time series model, which incorporates the available mortality experience to date. The mortality projection is then largely determined by this forecasted temporal factor. Other factor-based mortality models employ a similar approach to obtain forecasts and projections. The accuracy of forecasting relies on the appropriate modelling of relevant temporal factors. Moreover, fitted/forecast temporal factors may be employed to construct model-based mortality indexes (Chan, Li, and Li 2014). Such indexes are useful for demonstrating the overall evolution of the mortality experience of the investigated population over time. Despite its popularity, as pointed out by Chan, Li, and Li (2014), the LC model does not have the so-called new-data-invariant property. Specifically, the produced mortality indexes of historical periods are not invariant when the sample period is extended to include novel mortality rates. This property is desirable; otherwise, the associated mortality indexes will be intractable, owing to the potential variety of historical values.

This study revisits the new-data-invariant property for a range of popular factor-based mortality models. As stated by Chan, Li, and Li (2014), the Cairns–Blake–Dowd (CBD) model (Cairns, Blake, and Dowd 2006) is the only factor model (without modification) that is invariant to new data. Tan et al. (2014) attempt to adapt various mortality models to achieve a new-data-invariant property. However, these adapted models can accommodate only static samples and are unable to accommodate novel mortality data that are received sequentially. Thus, this inability hinders the examined models from dynamically calibrating new data items to monitor and update new-data-invariant mortality indexes.

The first objective of this study is to revisit the adaptations of popular mortality models. The adapted specifications are expected to produce new-data-invariant mortality indexes when models are recalibrated with sequential arrival of new data. In addition to the LC model, we consider the age-period-cohort (APC) model (Cairns et al. 2009), the Renshaw–Haberman (RH) model (Renshaw and Haberman 2006), and augmented common factor model or Li–Lee (LL) model (Li and Lee 2005). Adaptations of these models are examined by preserving parameters that are already estimated from the base sample period (i.e., excluding new data) with the aim of minimising the number of needed identifiability constraints. Two different strategies are proposed, with different levels of needed restrictions on parameters.

According to Tan et al. (2014), the cost of such adaptations is reduced by model fitting compared with non-adapted models. The second objective of this study, therefore, is to explore the adaptation that best balances the trade-off between goodness of fit and the new-data-invariant property. Analysing the male populations in English and Wales

from 1951 to 2016 for the 60 to 89 age group, we find that the strategy of fixing historical time-varying parameters is an optimal adaptation. Specifically, for the LC, APC, RH, and LL models, such adaptation introduces a minimal level of reduction in in-sample model fitting. Further, in contrast to the resulting goodness of fit of models that are invariant to new data, we find that the adapted RH model outperforms the LC, APC, and LL counterparts, as well as the original CBD model. This indicates that the adapted RH model may be preferable for demographic research. We then compare systematically the adapted and original RH models. The minimal level of reduced goodness of fit or forecasting accuracy for the adapted model is verified via backtesting and a range of sensitivity analyses. Thus, we conclude that the adapted RH model is an optimal strategy for balancing the trade-off between model fitting and the new-data-invariant property.

The remainder of this paper is organised as follows: Section 2 explains the new-data-invariant property and explores adaptations for the LC, APC, RH, and LL models. Section 3 compares the model-fitting performance between the static indexes, as investigated by Tan et al. (2014) and the dynamic indexes proposed in Section 2. Section 4 presents the results of the model backtesting and sensitivity analyses. Section 5 concludes.

2. New-data-invariant property and factor-based mortality models

2.1 New-data-invariant property

To facilitate the discussion of the new-data-invariant property, we list the following notations in Table 1, which will be used throughout the remainder of this paper.

Table 1: List of notations

$m_{x,t} = \frac{D_{x,t}}{E_{x,t}}$	The central mortality rate at age x in year t
$D_{x,t}$	Observed number of deaths at age x in year t
$E_{x,t}$	The matching exposures at age x in year t
$q_{x,t}$	The initial mortality rate at age x in year t
α_x	The geometric average rate at age x
β_x	Age-specific loading at age x
κ_t	The temporal factor in year t
$e_{x,t}$	The error at age x in year t
γ_c	The cohort factor, where $c = t - x$ denotes year of birth
n_a	The number of ages covered in the sample age range
$\bar{x}$	The mean age over the sample age range
B_x	The common age-specific loadings at age x for multi-populations modelling
K_t	The common temporal factor in year t for multi-populations modelling
$m_{x,t,i}$	The central death rate at age x in year t for population i
$\alpha_{x,i}, \beta_{x,i}, \kappa_{t,i}$	The population-specific factors for population i

We explain the new-data-invariant property using the CBD model (also known as M5 model) developed by Cairns, Blake, and Dowd (2006) and deemed the only specification that achieves this property (Chan, Li, and Li 2014). The CBD model has the following specifications:

$$\ln(\frac{q_{x,t}}{1-q_{x,t}}) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}). \tag{1}$$

The new-data-invariant property requires two conditions for the CBD model. The first assumes a Poisson distribution for the number of deaths, expressed as $D_{x,t} \sim Poi(E_{x,t}, m_{x,t})$. This assumption is introduced in the literature on stochastic mortality models by Brouhns, Denuit, and Vermunt (2002) and then by Renshaw and Haberman (2003). The second condition assumes a constant force of mortality for each integer age interval to link $q_{x,t}$ and $m_{x,t}$ with $m_{x,t} = -\ln(1 - q_{x,t})$. With these two conditions, the log-likelihood of Equation (1) is then expressed as follows:

$$\sum_{x=x_0}^{x_1}\sum_{t=t_0}^{t_1} D_{x,t}\ln(E_{x,t}m_{x,t}) - E_{x,t}m_{x,t} - \ln(D_{x,t}!) = \sum_{t=t_0}^{t_1}\lambda(t), \tag{2}$$

where $[x_0, x_1]$ is the sample age range, $[t_0, t_1]$ is the sample period (presented in a continuous interval), and $\lambda(t)$ is the contribution to the log-likelihood from data in year t. Chan, Li, and Li (2014) show that if the likelihood is separable for $t \neq s$, including additional $\lambda(t)$ for $t = t_1 + 1$, $t_1 + 2$,..., will not change existing estimates for the temporal factors $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ for $t = t_0, \ldots, t_1$. The reason is that the estimate of $\kappa_t^{(i)}$ is dependent only on $m_{x,t}$, $D_{x,t}$ and $E_{x,t}$ with the same t. This occurs because the CBD model does not have any age-specific or cohort parameters, as demonstrated in Equation (1). Thus, the CBD model can be implemented using linear instead of bilinear features to estimate unknown parameters. Further, because the likelihood is unaffected by including new data, the estimates of parameters will stay unchanged. Therefore, the original CBD model has a new-data-invariant property.

Unfortunately, this property does not hold for models with age-specific or cohort coefficients. We take the LC model as an example, the specification of which is described in Equation (3) with log-likelihood, as displayed in Equation (4).

$$\ln(m_{x,t}) = \alpha_x + \beta_x\kappa_t + e_{x,t}, \tag{3}$$

$$\lambda(t) = \sum_{x=x_0}^{x_1} D_{x,t}(\alpha_x + \beta_x\kappa_t + \ln(E_{x,t})) - E_{x,t}e^{\alpha_x+\beta_x\kappa_t} - \ln(D_{x,t}!). \tag{4}$$

For $t \neq s$, evidently $\lambda(t)$ and $\lambda(s)$ share age-specific parameters α_x and β_x. With the arrival of new data, the additional contribution $\lambda(t)$ for $t = t_1 + 1, \ldots$ will invariably affect the original estimates of α_x and β_x. Similarly, even without considering the identifiability constraint, the estimate of temporal factor κ_t is unlikely to hold, which may introduce a change in the fitted trend. Consequently, the LC model is not invariant to the arrival of new data. Further, the identifiability constraint contradicts intuitively the new-data-invariant property. For the LC model, the identification constraint for κ_t is $\sum_{t=t_0}^{t_1} \kappa_t = 0$. If the new-data-invariant property holds, κ_{t_1+1} for a single addition to the sample must be zero to satisfy the constraint. This is clearly unrealistic and contradicts the assumption that future mortality dynamics will continue to be based on historical patterns. Similarly, other commonly employed factor models, such as the APC and RH models, are not invariant to new data.

According to Chan, Li, and Li (2014), the new-data-invariant property is desirable for constructing mortality indexes. Such indexes are commonly used to evaluate the evolution of the mortality experience for a given population. In addition, the historical index should not be affected by the inclusion of newly available data, such that the index is tractable. To realise the new-data-invariant property, Tan et al. (2014) propose modifications to the identifiability constraints of models, including the LC and APC. However, the developed adaptations are not appropriate for dynamically arriving data (i.e., yearly mortality rates received sequentially). This is because of the static nature of the adaptation proposed by Tan et al. (2014). Specifically, the entire data range $[t_0, t_1]$ is split into two static periods, $[t_0, t_m]$ and $(t_m, t_1]$, where t_m is an intermediate time point. Moreover, the estimation for the second period $(t_m, t_1]$ is conditioned on pre-estimated parameters over the first period $[t_0, t_m]$. Consequently, dynamically updating the data range (i.e., continuously changing t_m and t_1) is unlikely to produce invariant estimates. Alternatively, this study attempts to achieve a new-data-invariant property in a dynamic manner. The data range is split into multiple subranges, such as $[t_0, t_1]$, $(t_1, t_2]$, $(t_2, t_3]$, where t_i $(i > 1)$ is the sequentially updated time points. Thus, the examined model is recalibrated every year after the base period $[t_0, t_1]$, such that the existing estimates of temporal factors remain unchanged. The details of this adaptation for the four models examined are explained below.

2.2 Adapted mortality models

In this study, we consider three well-studied single-population mortality models and a popular multi-population model in the literature, all of which do not possess the new-data-invariant property in their original specifications: the LC model by Lee and Carter (1992), APC model by Cairns et al. (2009), RH model by Renshaw and Haberman (2006), and LL model by Li and Lee (2005). The specifications of these models are briefly out-

lined in Table 2. In this section, we explain the details of the adaptation of each model to achieve a new-data-invariant property. In particular, our aim is to simplify the needed identifiability constraints as much as possible during adaptation. Notably, identifiability constraints are imposed in the estimation to obtain a unique set of estimated parameters. Hence, a full re-estimation of the original model with the arrival of additional years of data would imply a change to historical time-varying parameters, contradictory to the desirable new-data-invariant property. The proposed adaptations to these mortality models would leave historical time-varying parameters unchanged and totally remove or partially alleviate the need to rely on identifiability constraints. In all applicable cases, the Poisson distribution will be used to model the number of deaths, that is $D_{x,t} \sim Poi(E_{x,t}, m_{x,t})$. The summary of mathematical description of the different model-specific data-invariance adjustments is stated in Table 3 to 6.

Table 2: Specifications of examined mortality models

Model 1: The Lee–Carter model	
$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + e_{x,t}$	(2 constraints)
Model 2: The age-period-cohort model	
$\ln(m_{x,t}) = \alpha_x + n_\alpha^{-1} \kappa_t + n_\alpha^{-1} \gamma_{t-x} + e_{x,t}$	(2 constraints)
Model 3: The Renshaw–Haberman model	
$\ln(m_{x,t}) = \alpha_x + \beta_x^{(1)} \kappa_t + \beta_x^{(2)} \gamma_{t-x} + e_{x,t}$	(4 constraints)
Model 4: The Li–Lee (or augmented common factor) model	
$\ln(m_{x,t,i}) = \alpha_{x,i} + B_x K_t + \beta_{x,i} \kappa_{t,i} + e_{x,t}$	(6 constraints)

Table 3: Specifications of variants of Model 1

	Number of constraints	Identifiability constraints	Preserved parameter with new arrival of data	Parameter re-estimated with new arrival of data
1a	2	$\sum_x \beta_x = 1 \sum_t \kappa_t = 0$		$\alpha_x, \beta_x, \kappa_t$
1b	0		$[\kappa_{t_0}, ..., \kappa_{t_1}]$	$\alpha_x, \beta_x, \kappa_{t_2}$
1c	0		$[\kappa_{t_0}, ..., \kappa_{t_1}], \beta_x$	α_x, κ_{t_2}

Table 4: Specifications of variants of Model 2

	Number of constraints	Identifiability constraints	Preserved parameter with new arrival of data	Parameter re-estimated with new arrival of data
2a	2	$\sum_t \kappa_t = 0$ $\sum_{t-x} \gamma_{t-x} = 0$		$\alpha_x, \kappa_t, \gamma_{t-x}$
2b	1	$\sum_{t-x} \gamma_{t-x} = 0$	$[\kappa_{t_0}, ..., \kappa_{t_1}]$	$\alpha_x, \kappa_{t_2}, \gamma_{t-x}$
2c	0		$[\kappa_{t_0}, ..., \kappa_{t_1}]$, $\gamma_{base\ period}$	$\alpha_x, \kappa_{t_2}, \gamma_{new\ cohort}$

Note: The non-zero $\gamma_{base\ period}$ refers to $[\gamma_{t_0-x_n+5}, ..., \gamma_{t_1-x_1-5}]$ since the first and last five cohorts are set to be 0, where x_1 is the youngest age, x_n is the oldest age. With the arrival of new data, the first five cohort $[\gamma_{t_0-x_n}, ..., \gamma_{t_0-x_n+4}]$ and the last five cohort $[\gamma_{t_2-x_1-4}, ..., \gamma_{t_2-x_1}]$ are set to be 0, the new cohort updated at t_2 will be $\gamma_{t_2-x_1-5}$ with one year of new arrival of data.

Table 5: Specifications of variants of Model 3

	Number of constraints	Identifiability constraints	Preserved parameter with new arrival of data	Parameter re-estimated with new arrival of data
3a	4	$\sum_t \kappa_t = 0$ $\sum_{t-x} \gamma_{t-x} = 0$ $\sum_x \beta_x^{(1)} = 1$ $\sum_x \beta_x^{(2)} = 1$		$\alpha_x, \kappa_t,$ $\beta_x^{(1)}, \beta_x^{(2)}, \gamma_{t-x}$
3b	1	$\sum_x \beta_x^{(2)} = 1$	$\alpha_x, [\kappa_{t_0}, ..., \kappa_{t_1}]$	$\kappa_t, \beta_x^{(1)}, \beta_x^{(2)}, \gamma_{t-x}$
3c	0		$\alpha_x, [\kappa_{t_0}, ..., \kappa_{t_1}]$, $\gamma_{base\ period}$	$\kappa_{t_2}, \beta_x^{(1)}, \beta_x^{(2)}$ $\gamma_{new\ cohort}$

Note: The non-zero $\gamma_{base\ period}$ refers to $[\gamma_{t_0-x_n+5}, ..., \gamma_{t_1-x_1-5}]$ since the first and last five cohorts are set to be 0, where x_1 is the youngest age, x_n is the oldest age. With the arrival of new data, the first five cohort $[\gamma_{t_0-x_n}, ..., \gamma_{t_0-x_n+4}]$ and the last five cohort $[\gamma_{t_2-x_1-4}, ..., \gamma_{t_2-x_1}]$ are set to be 0 , the new cohort updated at t_2 will be $\gamma_{t_2-x_1-5}$ with one year of new arrival of data.

Table 6: Specifications of variants of Model 4

	Number of constraints	Identifiability constraints	Preserved parameter with new arrival of data	Parameter re-estimated with new arrival of data
4a	6	$\sum_x B_x = 1$ $\sum_t K_t = 0$ $\sum_x \beta_{x,1} = 1$ $\sum_t \kappa_{t,1} = 0$ $\sum_x \beta_{x,2} = 1$ $\sum_t \kappa_{t,2} = 0$		$\alpha_{x,1}, \alpha_{x,2}$ B_x, K_t $\beta_{x,1}, \kappa_{t,1}, \beta_{x,2}, \kappa_{t,2}$
4b	0		$[K_{t_0}, ..., K_{t_1}]$ $[\kappa_{t_0,1}, ..., \kappa_{t_1,1}]$ $[\kappa_{t_0,2}, ..., \kappa_{t_1,2}]$	$\alpha_{x,1}, \alpha_{x,2}$ B_x, K_{t_2} $\beta_{x,1}, \kappa_{t_2,1}, \beta_{x,2}, \kappa_{t_2,2}$

2.2.1 Lee–Carter model

The LC model is one of the most popular mortality models used in demographic research. In terms of estimation, both the single value decomposition method and the maximum

likelihood estimation (MLE) method by Brouhns, Denuit, and Vermunt (2002) are commonly used. To ensure a unique parameter estimation, two identifiability constraints are needed: $\sum_x \hat{\beta}_x = 1$ and $\sum_t \hat{\kappa}_t = 0$, because $(\hat{\alpha}_x, \frac{1}{c}\hat{\beta}_x, c\hat{\kappa}_t)$ and $(\hat{\alpha}_x - d\hat{\beta}_x, \hat{\beta}_x, \hat{\kappa}_t + d)$ are two transformed sets of parameters (from $(\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t)$) that would leave the fitted mortality rates unchanged in the absence of the identifiability constraints. In the remainder of this paper, we consider three cases of the LC model, namely Models 1a, 1b, and 1c, depending on the recalibration strategies. The MLE method is employed to produce estimates over the base period $[t_0, t_1]$.

Model 1a is the standard LC model. With the arrival of new data, all parameters α_x, β_x, and κ_t are re-estimated with no additional restrictions other than the two identifiability constraints stated above. The resulting $\hat{\kappa}_t$ then changes each time the model is recalibrated to include the new data.

Model 1b aims to achieve adaptation while maintaining $\hat{\kappa}_t$ estimated from the sample of the base period $[t_0, t_1]$. With the additional data arrival in the period $(t_1, t_2]$, the estimated $\hat{\kappa}_t$ over $[t_0, t_1]$ is preserved, and we update the estimates of $\hat{\alpha}_x$ and $\hat{\beta}_x$ for all ages and obtain $\hat{\kappa}_{t_2}$ via the MLE. During the MLE procedure, we discard the identifiability constraints of $\hat{\beta}_x$ and $\hat{\kappa}_t$. This is because $\hat{\kappa}_t$ is unchanged for $t \in [t_0, t_1]$, and only one κ_t needs to be estimated, for $t = t_2$. Thus, it is easy to see that the identifiability issue no longer applies. Specifically, for $t \in [t_0, t_1]$, we have unknown α_x and β_x for $\ln(m_{x,t}) = \alpha_x + \beta_x\hat{\kappa}_t + e_{x,t}$, whereas $\hat{\kappa}_t$ is known. Therefore, β's and κ's no longer have a confounding effect on the likelihood. This recalibration process can then continue for each sequentially updated mortality data period.

Apparently, the $\hat{\beta}_x$ produced in the recalibration of Model 1b does not necessarily sum up to 1. This is concerning when cross-population comparison is needed, because the estimated β's are not 'standardised.' To address this issue, Model 1c preserves estimates of both κ_t and β_x obtained over $[t_0, t_1]$. Thus, only α_x and κ_t over the new data period are (re-)estimated. The MLE procedure is performed without the need to apply any identifiability constraints.

2.2.2 Age-period-cohort model

The APC model proposes including an additional cohort effect in the LC framework. For identifiability constraints, the two commonly required values are $\sum_t \hat{\kappa}_t = 0$ and $\sum \hat{\gamma}_{t-x} = 0$. Moreover, Cairns et al. (2009) point out that the tilting parameter δ is needed because adding $n_\alpha\delta((t - \bar{t}) - (x - \bar{x}))$ to γ_{t-x}, subtracting $n_\alpha\delta(t - \bar{t})$ from κ_t, and adding $\delta(x - \bar{x})$ to α_x will have no effect on the first two constraints. Cairns et al. (2009) also suggest that the first and last five cohort parameters $\hat{\gamma}_{t-x}$ can be set to zero to avoid getting unstable parameter estimates with the lack of mortality data. By doing so, the additional constraints stated in Cairns et al. (2009) are no longer needed for the APC model. In contrast hand, Hunt and Villegas (2015) propose an additional constraint

of $\sum_y \hat{y} - \bar{y}_{\gamma_y} = 0$ for $y = t - x$ to speed up the convergence issue (see, for example, Cairns et al. 2009) with minimal effect on the model fitting for both the APC and RH models. Finally, similar to the case of the LC, we consider three cases for the new data recalibration using the APC, namely, Models 2a, 2b, and 2c.

Model 2a is the standard APC model. All parameters α_x, κ_t, and γ_{t-x} will be re-estimated upon the arrival of new data, considering the two identifiability constraints. Model 2b keeps the $\hat{\kappa}_t$ obtained from the base period $[t_0, t_1]$ and estimates $\hat{\alpha}_x$, $\hat{\gamma}_{t-x}$, and $\hat{\kappa}_t$ for the new data. In this case, the constraint of $\sum_t \hat{\kappa}_t = 0$ is no longer needed for the same reason, as stated in the LC case. However, the constraint of the cohort factor $\sum \gamma_{t-x} = 0$ is still required. It is possible to have $\alpha^*_x = \alpha_x - bn_a^{-1}$ and $\gamma^*_{t-x} = \gamma_{t-x} + b$ where b is a non-zero constant, such that $\alpha_x + n_a^{-1}\kappa_t + n_a^{-1}\gamma_{t-x} = (\alpha_x - bn_a^{-1}) + n_a^{-1}\kappa_t + n_a^{-1}(\gamma_{t-x} + b) = \alpha^*_x + n_a^{-1}\kappa_t + n_a^{-1}\gamma^*_{t-x}$. That is, identifiability constraint is needed to obtain unique solutions. Model 2c preserves both the produced $\hat{\kappa}_t$ and $\hat{\gamma}_{t-x}$, and re-estimate α_x, as well as associated κ_t and γ_{t-x} for the new data. Because the cohort parameters over the base period are kept, the constraint $\sum \gamma_{t-x} = 0$ is no longer needed to ensure the uniqueness of the estimation. Therefore, no constraints are required for Model 2c.

2.2.3 Renshaw–Haberman model

The RH model also considers the cohort effect, and has a more flexible specification than that of the APC, with age-specific loadings for both the temporal ($\beta_x^{(1)}$) and cohort ($\beta_x^{(2)}$) effects. Altogether, four identifiability constraints are needed: $\sum_t \kappa_t = 0$, $\sum \gamma_{t-x} = 0$, $\sum_x \beta_x^{(1)} = 1$, and $\sum_x \beta_x^{(2)} = 1$. Similar to the APC counterpart, the first and last five cohort years are excluded from the estimation to avoid high uncertainty therein. However, this will not reduce the number of required identifiability constraints.

Three cases, namely Models 3a, 3b, and 3c, are examined. Model 3a is the standard RH model. Similar to the cases of the LC and APC, Model 3b preserves the estimates of κ_t over the based period $[t_0, t_1]$. The original RH model has four constraints. Thus, different from Models 1b and 2b, to reduce the needed constraints, we propose to maintain the estimates of α_x from the base period as well. Otherwise, two identifiability constraints, $\sum_x \beta_x^{(2)} = 1$ and $\sum \gamma_{t-x} = 0$, would still be required. To see this, it is possible to have multiple solutions for α_x and γ_{t-x} with $\alpha^*_x = \alpha_x - b\beta_x^{(2)}$ and $\gamma^*_{t-x} = \gamma_{t-x} + b$ for a non-zero constant b, such that $\alpha_x + \beta_x^{(1)}\kappa_t + \beta_x^{(2)}\gamma_{t-x} = (\alpha_x - b\beta_x^{(2)}) + \beta_x^{(1)}\kappa_t + \beta_x^{(2)}(\gamma_{t-x} + b) = \alpha^*_x + \beta_x^{(1)}\kappa_t + \beta_x^{(2)}\gamma^*_{t-x}$. Keeping the estimates of α_x requires only one identifiability constraint of $\sum_x \beta_x^{(2)} = 1$. This is so because non-unique solutions are still possible for $\beta_x^{*(2)} = \beta_x^{(2)}/c$ and $\gamma^*_{t-x} = c\gamma^*_{t-x}$, with a non-zero constant c. Finally, Model 3c keeps all estimates from the base period and needs only $\hat{\kappa}_t$ and $\hat{\gamma}_{t-x}$ to be estimated for the new data. Identifiability constraints are no longer needed in this case.

2.2.4 Li–Lee model

Unlike the three single-population mortality models, the LL (or augmented common factor) model is a multi-population counterpart. Compared with a single-population model, such as the LC, the advantage of the LL model is that the forecast mortality rates are coherent across the modelled populations (Li and Lee 2005). Thus, the ratio of forecast mortality rates at each age will not diverge infinitely across populations, which is particularly desirable for long-term projection. In the two-population case, the six identifiability constraints required are $\sum_x B_x = 1$, $\sum_t K_t = 0$ for the common factors, and $\sum_x \beta_{x,i} = 1$ and $\sum_t \kappa_{x,i} = 0$ for each of the population $i = 1, 2$.

For the LL model, we consider two cases. Model 4a is the standard LL model. Model 4b preserves all estimates of the temporal factors K_t, $\kappa_{t,1}$, and $\kappa_{t,2}$ over the base period. Similar to the reason stated for Model 1b, no constraints are required. We do not explore the case (i.e., a potential Model 4c) to further fix age-specific loadings (i.e., B_x, $\beta_{x,1}$, and $\beta_{x,2}$). As detailed in Section 3.3, this is to better balance the trade-off between the new-data-invariant property and model fitting.

2.3 Forecasting

In the demographic literature, temporal factors κ_t (Models 1–3) and K_t (Model 4) are normally assumed to be non-stationary (e.g., Lee and Carter 1992). A simple and common practice is to adopt a random walk with drift in terms of forecasting. For illustration, the specification of this model for κ_t is as follows:

$$\hat{\kappa}_t = \theta + \hat{\kappa}_{t-1} + \epsilon_t,$$

where θ is the drift parameter and ϵ_t is the error term. θ is estimated as the average of $\Delta\hat{\kappa}_t = \hat{\kappa}_t - \hat{\kappa}_{t-1}$ for all available $\hat{\kappa}_t$. Thus, $\theta = \Delta\kappa_t/n$, where n is the sample size of $\Delta\kappa_t$. After obtaining $\hat{\theta}$, from the base period $[t_0, t_1]$, the one-step-ahead forecast $\hat{\kappa}_{t_1+1}$ can be produced, which is used to derive the point forecast of the central death rate m_{x,t_1+1}, along with other estimated parameters.

As explained above, to achieve the new-data-invariant property, estimates of temporal factors over the base period are maintained for Models 1–4 b and c. As an alternative to the simple random walk with drift model, we allow for an autoregressive integrated moving average (ARIMA) structure to model $\hat{\kappa}_t$ as follows:

$$\Delta^d \hat{\kappa}_t = \theta + \sum_{i=1}^{p} \alpha_i \Delta^d \hat{\kappa}_{t-i} - \sum_{j=1}^{q} \beta_j \epsilon_{t-j} + \epsilon_t \,,$$

where Δ^d is the differencing operator, with $\Delta^1\hat{\kappa}_t = \hat{\kappa}_t - \hat{\kappa}_{t-1}$. α_i and β_j are associated AR and MA coefficients at the ith and jth orders, respectively. To choose the order of p, d, and q, we follow Hyndman and Khandakar (2008) and employ the corrected Akaike information criterion (AIC). For all models, we find that ARIMA(0,1,0) specification results in the smallest AIC and is then chosen as the final model. This coincides with a simple random walk with drift. Finally, the new-data-invariant property requires these models to be 'path dependent.' Thus, the h-step-ahead forecasts would be obtained as h sequential one-step-ahead forecasts. At the h-step for Models 1–4 b and c, $\hat{\theta}$ depends on all $\hat{\kappa}_t$ estimated over $[t_0, t_{h-1}]$. In contrast, for standard Models 1–4 a, $\hat{\theta}$ is kept constant and depends only on $\hat{\kappa}_t$ estimated over the base period.

3. Model-fitting performance

3.1 Data description

To illustrate the model performance, we employ the populations of England and Wales, the mortality data of which are collected from the Human Mortality Database (2020); these populations have been considered in most previous studies (e.g., Cairns et al. 2009). Following Booth et al. (2006) and the findings of Lee and Miller (2001) regarding different age patterns of change in mortality before and after 1950, we choose a range of data starting from 1951 to 2016. The age range of 60 to 89 years is used because we focus on analysing mortality at higher ages, as stated by Cairns et al. (2009). We consider the male population for the single-population models and both male and female populations for the multi-population model.

3.2 Model parameter estimation and discussions

To illustrate the recalibration performance, we set the base period to 1951–2000. Altogether, we plot the parameter estimates over three new data periods: 2001–2005, 2001–2010, and 2001–2015. Baseline results are also included for comparison. The results of Models 1a–c are shown in Figure 1. We observe upward shifts of $\hat{\kappa}_t$ and downward shifts of $\hat{\alpha}_x$ for Model 1a with a longer recalibrated data range. This is consistent with the overall improvement (decline) in mortality rates across all ages during the period 2001–2015. The age-specific factors $\hat{\beta}_x$ also 'rotate' slightly, which suggests that the older age group (70 to 90) experiences more mortality improvements. This is consistent with the findings of Li, Lee, and Gerland (2013). In Model 1b, no shifts of $\hat{\kappa}_t$ are observed over the base period because of the new-data-invariant property. The age-specific factors $\hat{\beta}_x$ in Model 1b rotate similarly to those in Model 1a, whereas $\hat{\alpha}_x$ do

not change much with the arrival of new data. In Model 1c, the estimated $\hat{\beta}_x$ over the base period is further fixed, and $\hat{\alpha}_x$ also changes marginally with additional years of data.

The parameter estimates of Models 2a–c are presented in Figure 2. Similar to the LC model, shifts are observed for all age, temporal, and cohort factors in Model 2a. After fixing the temporal effects over the base period, only the cohort parameters shift substantially in Model 2b. Finally, when both $\hat{\kappa}$'s and $\hat{\gamma}$'s are kept, even more limited variations are visualised for Model 2c.

We demonstrate the parameter estimates of Models 3a–c in Figure 3. Overall, the shifts of the three major factors ($\hat{\alpha}$'s, $\hat{\kappa}$'s, and $\hat{\gamma}$'s) in Model 3a are comparable to those in Model 2a. After fixing the $\hat{\alpha}_x$ and $\hat{\kappa}_t$ in Model 3b, and $\hat{\alpha}_x$ and $\hat{\kappa}_t$ and $\hat{\gamma}_{t-x}$ in Model 3c, over the base period, much smaller variations are observed in these three factors. Because the estimated age-specific loadings $\hat{\beta}_x^{(1)}$ and $\hat{\beta}_x^{(2)}$ are not fixed over the base period in any case, they exhibit highly similar patterns for Models 3a–c.

Finally, the estimates of Models 4a and 4b are plotted in Figures 4 and 5, respectively. In the multi-population case, the reference population is selected as the total mortality rates, with the male and female population combined. The population-specific $\hat{\alpha}$'s and the common factors $\hat{B}$'s and $\hat{K}$'s in Model 4a demonstrate comparable patterns to their single-population counterparts, as shown in Figure 1. In Model 4b, we keep $\hat{K}_t$, $\hat{\kappa}_{t,1}$, and $\hat{\kappa}_{t,2}$ over the base period. Consequently, only age-specific loadings rotate with the arrival of new data.

3.3 Model-fitting performance comparison

Tan et al. (2014) argue that the potential price to pay for the adaptation to achieve the new-data-invariant property is the reduced model fitting. In this section, we explore this using two widely employed performance comparison criteria: the Bayesian information criterion (BIC) and the mean absolute percentage error (MAPE).

The BIC is used to assess the goodness of fit of a fitted model and penalises non-parsimonious models. Based on the log-likelihood, the BIC is calculated as follows:

$$\text{BIC} = -2l + p\ln(n),$$

where l is the log-likelihood of a fitted model, p is the number of total parameters, and d is the number of observations. A smaller BIC indicates a better overall goodness of fit.

Figure 1: Estimation of α_x, β_x, and κ_t for Lee–Carter model

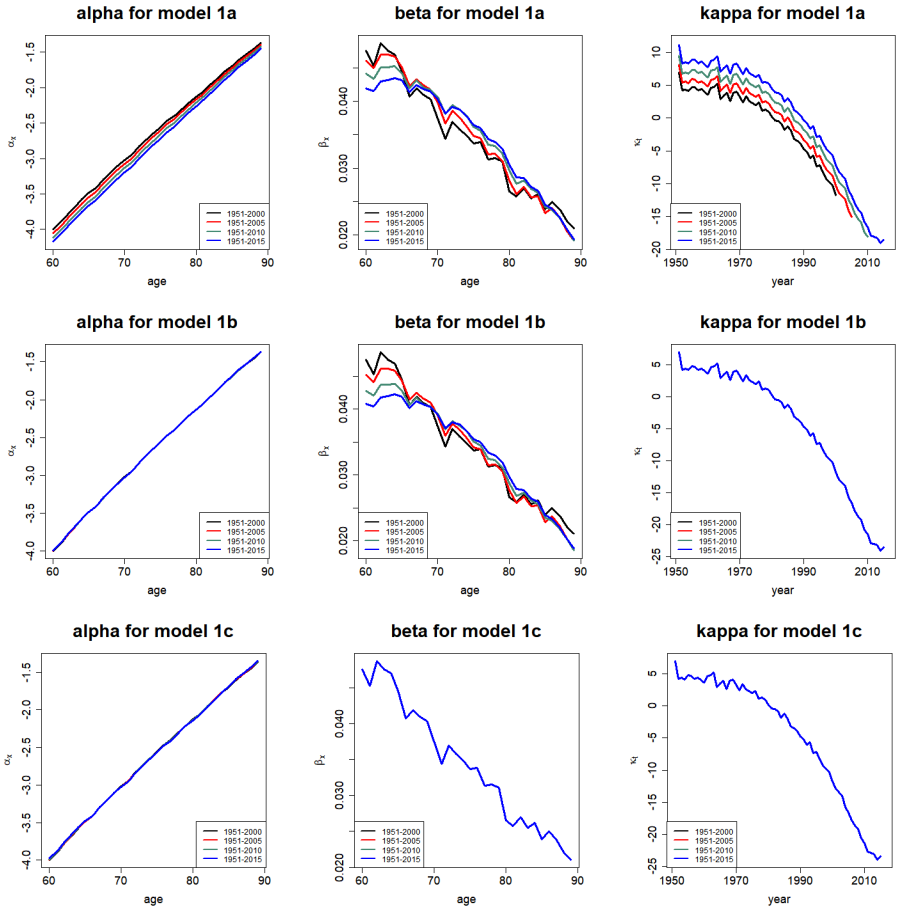


Figure 2: Estimation of α_x, β_x, and γ_{t-x} for age-period-cohort model

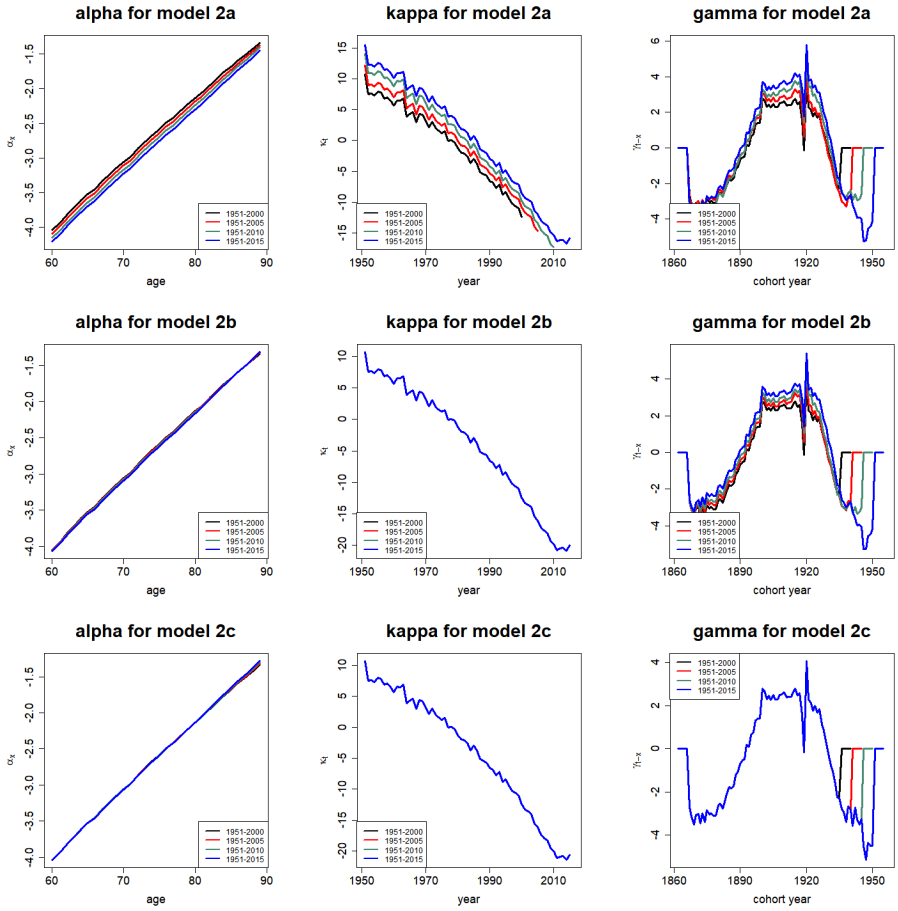


Figure 3: Estimation of α_x, $\beta_x^{(1)}$, κ_t,$\beta_x^{(2)}$, and γ_{t-x} for Renshaw–Haberman model

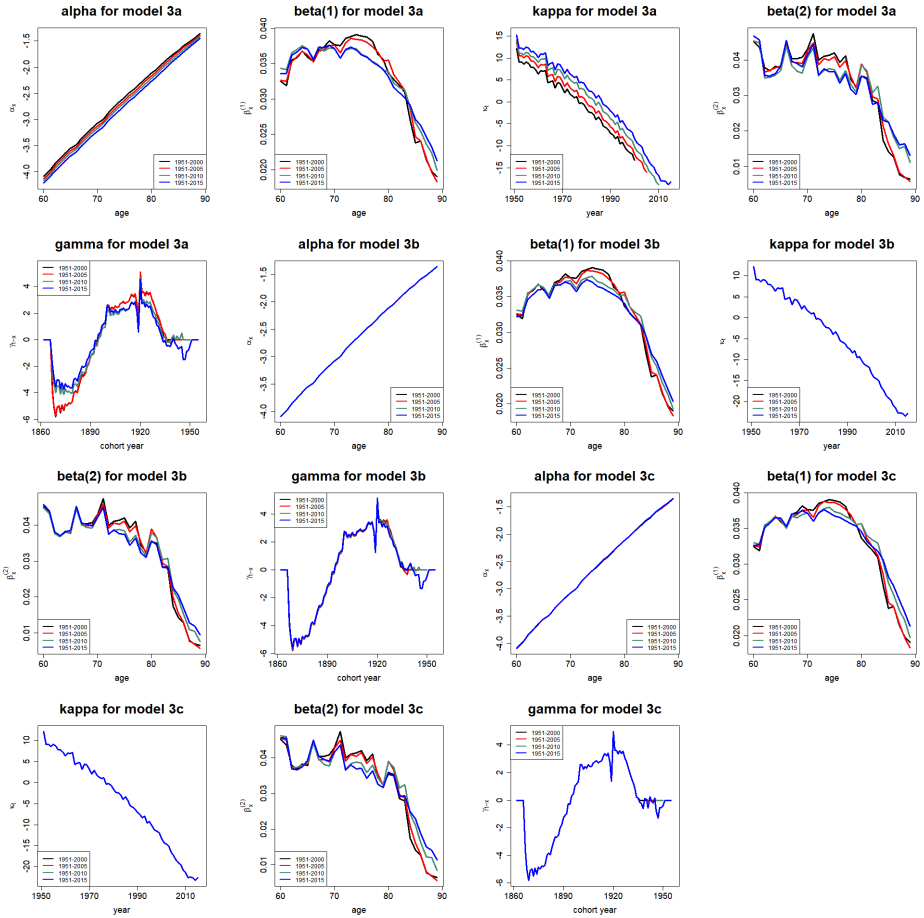


Figure 4: Estimation of α_x, B_x, K_t, $\beta_{x,1}$, and $\beta_{x,2}$, $\kappa_{t,1}$, and $\kappa_{t,2}$ for the original Li–Lee model

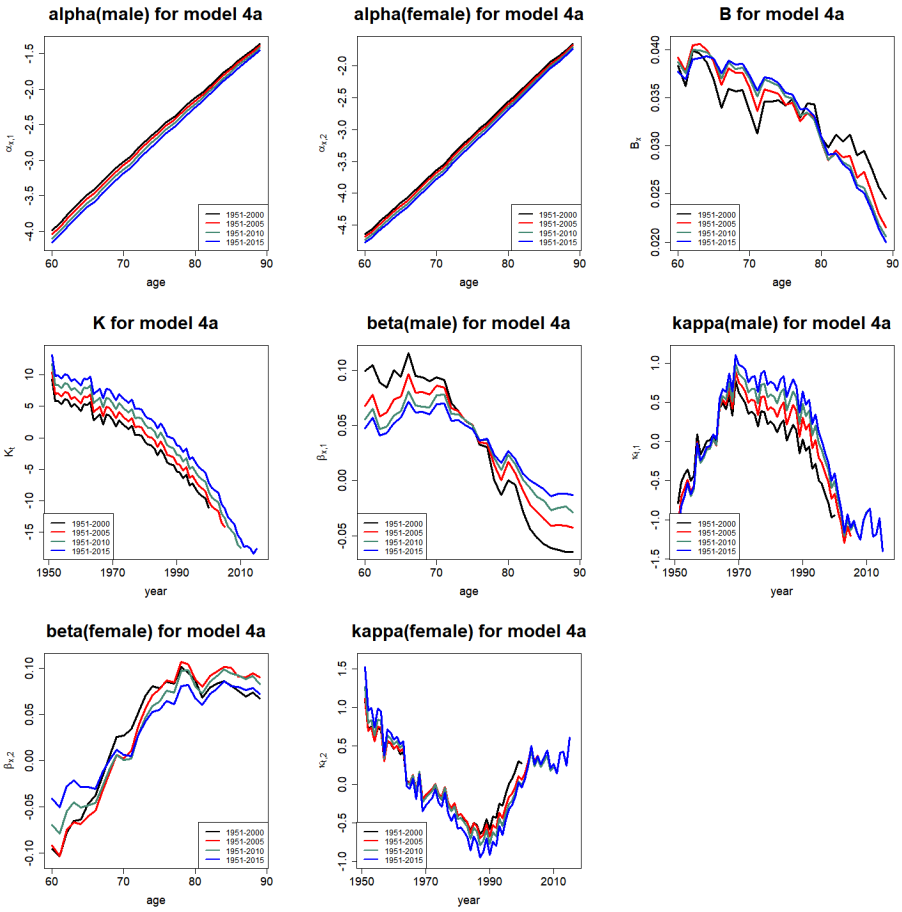
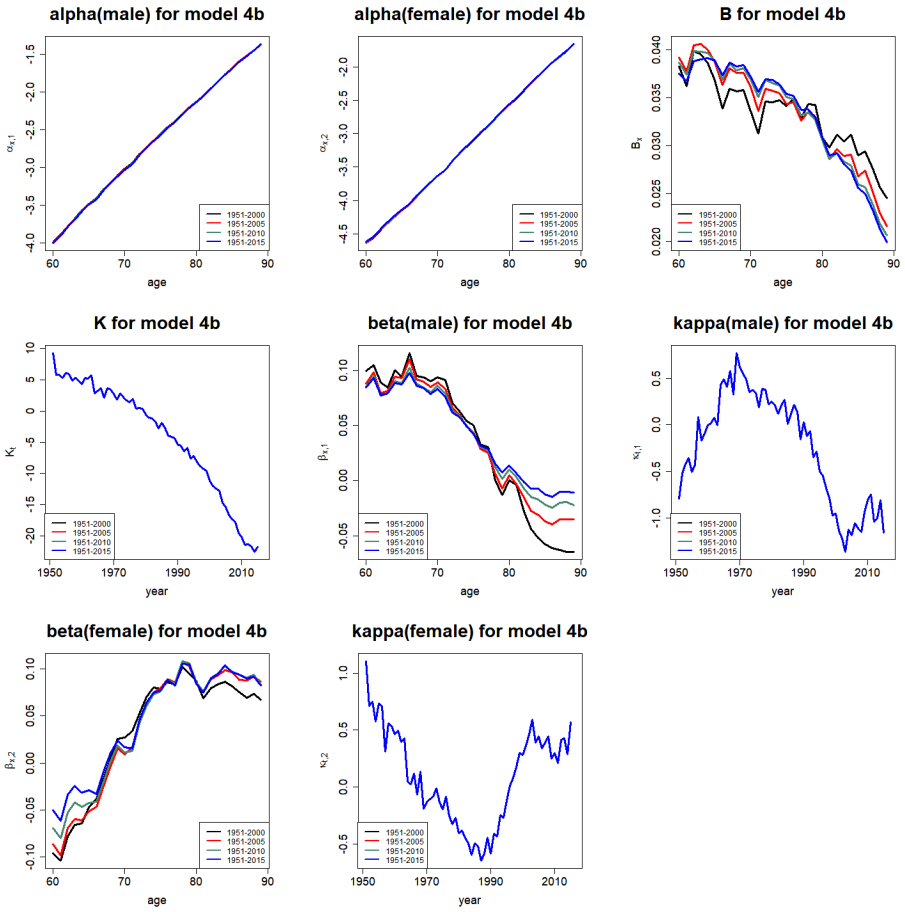


Figure 5: Estimation of α_x, B_x, K_t, $\beta_{x,1}$, and $\beta_{x,2}$, $\kappa_{t,1}$, and $\kappa_{t,2}$ for the adapted Li–Lee model



We present BICs for all models (including the CBD Model for comparison) across the base period and 16 recalibrated periods in Table 7.[5] For the LC models, the BIC is nearly identical for Models 1a and 1b, but is substantially larger for Model 1c. The results of the APC and RH models are largely consistent with those of the LC, except that the differences between Model 2a and Model 2b are noticeably greater than those between Model 1a and Model 1b. Because the LL is a multi-population model, BICs are not directly comparable to single-population models. In addition, BICs of the CBD model are noticeably larger than those of Models 1b, 2b, and 3b. Overall, Model 3b (adapted RH model) yields the smallest BICs among all new-data-invariant specifications, indicating its preferred in-sample performance.

Table 7: BIC values for Models 1–4

Period/Model	1a	1b	1c	2a	2b	2c	3a	3b	3c	4a	4b	CBD
1951–2000	25272	25272	25272	21867	21867	21867	20062	20062	20062	51918	51918	25629
1951–2001	25760	25760	25782	22506	22515	22557	20434	20437	20438	52902	52865	26044
1951–2002	26300	26300	26373	23113	23140	23249	20808	20813	20815	53948	53997	26489
1951–2003	26909	26911	27088	23723	23778	23987	21177	21186	21187	55063	55367	26993
1951–2004	27535	27537	27825	24320	24400	24713	21612	21622	21629	56261	56823	27510
1951–2005	28079	28081	28541	24954	25066	25509	22004	22020	22036	57498	58385	28014
1951–2006	28715	28717	29409	25394	25539	26115	22418	22440	22471	58869	60017	28590
1951–2007	29361	29363	30367	26060	26238	26970	22838	22878	22920	60373	61811	29202
1951–2008	29983	29985	31351	26763	26985	27887	23266	23328	23386	62008	63760	29847
1951–2009	30587	30590	32378	27393	27667	28718	23685	23773	23860	63514	65421	20494
1951–2010	31244	31248	33528	28003	28331	29518	24097	24220	24341	65007	67010	31122
1951–2011	31966	31971	34806	28506	28880	30222	24529	24677	24853	66701	68698	31817
1951–2012	32525	32531	35989	29046	29486	31007	24875	25031	25245	68368	70529	32536
1951–2013	33136	33142	37351	29721	30240	31978	25273	25438	25699	69994	72334	33373
1951–2014	33795	33803	38842	30260	30855	32805	25635	25822	26113	71626	74024	34187
1951–2015	34365	34374	40273	31154	31848	34162	26050	26226	26577	73331	76008	35144
1951–2016	34987	34997	41858	31902	32683	35377	26408	26591	26981	75101	77899	36067

As an alternative to the BIC, we use MAPE to compare the fitted central death rate with the actual rate, which is calculated as follows:

$$\frac{1}{n}\sum_{x,t}\left|\frac{\hat{m}_{x,t}-(D_{x,t}/E_{x,t})}{(D_{x,t}/E_{x,t})}\right|,$$

[5] Studies including Hunt and Villegas (2015) have documented the importance to set $\beta_x^{(2)} = 1$ to simplify the RH model. In our analyses (available upon request) analogous to those presented in Tables 3 to 6, those simplified RH models outperforms the LC (Models 1a–1c) but underperform the RH (Models 3a–3c) proposed in this paper. This further supports the usefulness of our final model.

where n is the number of observations, $D_{x,t}$ is the actual number of deaths at age x in year t, $E_{x,t}$ is the corresponding exposed to risk, and $\hat{m}_{x,t}$ is the fitted central death rate from the examined model. A smaller MAPE indicates a more accurate estimate.

In Table 8, we present MAPEs of all examined models over 17 periods. In line with our observations regarding the BIC, we find that the less stringent new-data-invariant models (1b, 2b, and 3b) are superior to their more restrictive equivalents (1c, 2c, and 3c). Compared to the standard models (1a, 2a, and 3a), the costs of achieving new-data-invariant property for Models 1b, 2b, and 3b are more acceptable, evidenced by the marginally increased MAPEs. Further, in contrast to all new-data-invariant specifications, Model 3b almost uniformly leads to the smallest MAPE, suggesting its best goodness-of-fit performance.

Table 8: MAPE values for Models 1–4

Period/Model	1a	1b	1c	2a	2b	2c	3a	3b	3c	4a	4b	CBD
1951–2000	2.30%	2.30%	2.30%	1.78%	1.78%	1.78%	1.31%	1.31%	1.31%	2.39%	2.39%	2.34%
1951-2001	2.31%	2.31%	2.31%	1.82%	1.82%	1.82%	1.31%	1.31%	1.31%	2.40%	2.38%	2.32%
1951–2002	2.32%	2.32%	2.33%	1.86%	1.86%	1.86%	1.30%	1.31%	1.31%	2.41%	2.40%	2.31%
1951–2003	2.34%	2.34%	2.37%	1.89%	1.89%	1.90%	1.30%	1.30%	1.30%	2.43%	2.43%	2.31%
1951–2004	2.36%	2.36%	2.40%	1.92%	1.92%	1.95%	1.31%	1.31%	1.32%	2.45%	2.47%	2.31%
1951–2005	2.37%	2.37%	2.44%	1.95%	1.96%	1.99%	1.31%	1.31%	1.32%	2.47%	2.51%	2.32%
1951–2006	2.40%	2.40%	2.49%	1.96%	1.97%	2.01%	1.32%	1.32%	1.33%	2.50%	2.55%	2.33%
1951–2007	2.43%	2.43%	2.54%	1.99%	2.00%	2.06%	1.33%	1.34%	1.34%	2.53%	2.60%	2.35%
1951–2008	2.45%	2.45%	2.59%	2.02%	2.04%	2.11%	1.34%	1.35%	1.36%	2.57%	2.66%	2.37%
1951–2009	2.47%	2.47%	2.65%	2.05%	2.06%	2.15%	1.35%	1.37%	1.38%	2.61%	2.70%	2.39%
1951–2010	2.50%	2.50%	2.71%	2.07%	2.09%	2.19%	1.35%	1.38%	1.40%	2.64%	2.73%	2.41%
1951–2011	2.53%	2.53%	2.78%	2.09%	2.11%	2.23%	1.36%	1.39%	1.41%	2.68%	2.77%	2.44%
1951–2012	2.54%	2.54%	2.84%	2.12%	2.14%	2.27%	1.35%	1.39%	1.42%	2.72%	2.81%	2.47%
1951–2013	2.56%	2.55%	2.91%	2.16%	2.19%	2.33%	1.35%	1.39%	1.43%	2.76%	2.85%	2.51%
1951–2014	2.57%	2.57%	2.99%	2.20%	2.23%	2.38%	1.35%	1.39%	1.43%	2.80%	2.89%	2.54%
1951–2015	2.58%	2.58%	3.06%	2.25%	2.29%	2.46%	1.36%	1.40%	1.45%	2.85%	2.94%	2.59%
1951–2016	2.60%	2.60%	3.14%	2.29%	2.34%	2.53%	1.35%	1.39%	1.45%	2.89%	2.98%	2.64%

To sum up, we demonstrate that for single-population models, Models 1b, 2b, and 3b are more preferred specifications to Models 1c, 2c, and 3c, respectively, concerning model-fitting performance. This may be explained by the fact that imposing more restrictions on a model will make its likelihood deviate more from the maximum and therefore, reduce the goodness of fit. Consequently, for the multi-population case, we consider fixing only the estimates of temporal factors as in Model 4b. As presented in Tables 7 and 8, the resulting BICs and MAPEs of Model 4b only marginally increase, which is consistent with the single-population findings. In general, we find that the RH models beat other specifications, whereas Model 3b consistently outperforms the new-data-invariant competitors. This finding is consistent with the both BIC and MAPE results. Therefore, to best balance the goodness of fit and new-data-invariant property, our analyses suggest

that the less restrictive adapted RH model may be the optimal approach. Next, we further explore the original and adapted RH models to verify this.

4. Model backtesting performance and sensitivity analyses

In addition to assessing the in-sample model-fitting performance, backtesting can serve as a useful tool for evaluating the projection accuracy of a model. Because the RH models outperform competitors, we focus on the performance of Models 3a, 3b, and 3c in this section.

To obtain backtesting results, we split the entire dataset into various training and testing sets. Specifically, our training sets start from 1951 to 2000, and we include one new arrival year of data, until 1951–2015 rates are included. For the testing sets, we consider three categories: one-, three-, and five-year projections. For instance, using the initial training set of 1951–2000, we project mortality rates over the next one (2001), three years (2001–2003), and five years (2001–2005). Similar to Section 3, the MAPE is used in all cases to measure backtesting performance. Thus, in this example, the one-, three-, and five-year MAPEs are obtained by comparing the projected rates against real data over 2001, 2001–2003, and 2001–2005, respectively.

More generally, let s be the last year of a training set, then $\hat{\kappa}_{s+1} = \theta + \hat{\kappa}_s$ is used to project $\hat{m}_{x,s+1}$ in the RH models. MAPE of a one-year projection is then calculated by $\sum_x (\hat{m}_{x,s+1} - m_{x,s+1})/m_{x,s+1}$, where $\hat{m}_{x,s+1}$ is the projected rate and $m_{x,s+1}$ is the actual rate. For the three-year and five-year projections, the last training sets are 1951–2013 and 1951–2011, respectively, with the testing sets of 2014–2016 and 2012–2016, respectively.

MAPE results are computed for Models 3a, 3b, and 3c and are displayed in Table 9. It can be concluded that, similar to the in-sample results, Models 3a and 3b again exhibit a similar level of MAPE, whereas MAPE of Model 3c deviates substantially from them.

Finally, we consider a range of sensitivity analyses. In Section 3, we focus on the RH models because of their outstanding performance. Three types of sensitivity analyses are conducted, and all analyses are based on the one-year projection described above. First, the base period (the starting training set) is altered to three sets: 1951–1995, 1951–2005, and 1956–2005. The corresponding MAPEs are listed in Table 10. Second, we test the sensitivity of the recalibration period. In the baseline analyses, all models are updated every year with the arrival of new data. Using the same starting training set of 1951–2000, we update the RH models every three years and five years with the newly arrived data. Third, all models are still updated yearly using a fixed-sample rolling window approach. Specifically, rather than expanding the sample size at each step, we use a fixed size of 50 years and change the starting and ending years to project the one-year rate. MAPE results for the rolling window and the two new recalibration periods are reported in Table 11.

Regarding the rolling window part, the dataset has been updated by removing the oldest year and incorporating the latest year's data. For the three-year recalibration period, we recalibrated our results using the most recent data every three years. Similarly, we recalibrated our results with the most recent data every five years. Additionally, to be consistent with Section 3, we also present the in-sample BIC of each training set in Tables 12 and 13. In summary, we find that Models 3a and 3b exhibit superior performance in both in-sample evaluation (as measured by the BIC) and backtesting (as measured by MAPE) compared with Model 3c, as indicated by various robustness checks. Therefore, we identify Model 3b as the optimal strategy to cope with the new-data-invariant property in demographic analyses.

Table 9: MAPE values for Model 3

	One-year projection			Three-year projection			Five-year projection		
Model	3a	3b	3c	3a	3b	3c	3a	3b	3c
1951–2000	1.67%	1.66%	1.66%	1.45%	1.45%	1.45%	5.02%	4.97%	4.92%
1951–2001	1.35%	1.36%	1.36%	2.90%	2.91%	2.93%	5.09%	5.22%	5.22%
1951–2002	1.32%	1.33%	1.28%	3.90%	3.98%	3.96%	6.41%	6.57%	6.51%
1951–2003	3.70%	3.71%	3.70%	6.16%	6.26%	6.30%	6.70%	6.79%	6.75%
1951–2004	1.71%	1.73%	1.79%	3.28%	3.37%	3.38%	5.59%	5.59%	5.51%
1951–2005	2.13%	2.17%	2.35%	2.84%	2.98%	3.03%	5.10%	5.07%	5.09%
1951–2006	1.97%	2.11%	2.20%	3.46%	3.45%	3.57%	5.79%	5.65%	5.56%
1951–2007	2.27%	2.48%	2.52%	3.28%	3.26%	3.45%	3.22%	3.36%	3.78%
1951–2008	2.89%	3.03%	3.19%	5.41%	5.39%	5.38%	2.70%	2.93%	3.48%
1951–2009	2.06%	2.36%	2.67%	1.74%	2.25%	2.81%	2.23%	2.70%	3.09%
1951–2010	2.73%	2.90%	3.13%	2.59%	2.69%	3.08%	4.94%	5.26%	5.20%
1951–2011	2.03%	2.29%	2.51%	3.17%	3.55%	3.60%	7.05%	7.47%	7.50%
1951–2012	2.13%	2.21%	2.61%	4.56%	4.80%	4.81%	-	-	-
1951–2013	1.36%	1.50%	1.93%	3.29%	3.64%	3.64%	-	-	-
1951–2014	3.55%	3.61%	3.77%	-	-	-	-	-	-
1951–2015	1.04%	1.20%	1.75%	-	-	-	-	-	-

Table 10: MAPE values for different base year for Model 3

Base 1951–1995				Base 1951–2005				Base 1956–2005			
Model	3a	3b	3c	Model	3a	3b	3c	Model	3a	3b	3c
1951–1995	2.28%	2.26%	2.26%	1951–2005	2.13%	2.11%	2.11%	1956–2005	2.13%	2.11%	2.11%
1951–1996	2.18%	2.26%	2.26%	1951–2006	1.97%	2.06%	2.07%	1956–2006	1.97%	2.06%	2.07%
1951–1997	1.18%	1.44%	1.46%	1951–2007	2.27%	2.38%	2.45%	1956–2007	2.27%	2.38%	2.45%
1951–1998	1.46%	1.95%	1.91%	1951–2008	2.89%	2.96%	3.00%	1956–2008	2.89%	2.96%	3.00%
1951–1999	3.20%	3.45%	3.52%	1951–2009	2.06%	2.27%	2.51%	1956–2009	2.06%	2.27%	2.51%
1951–2000	1.67%	2.07%	2.19%	1951–2010	2.73%	2.84%	3.03%	1956–2010	2.73%	2.84%	3.03%
1951–2001	1.35%	1.85%	1.96%	1951–2011	2.03%	2.25%	2.40%	1956–2011	2.03%	2.25%	2.40%
1951–2002	1.32%	1.79%	1.75%	1951–2012	2.13%	2.24%	2.54%	1956–2012	2.13%	2.24%	2.54%
1951–2003	3.70%	3.96%	4.10%	1951–2013	1.36%	1.44%	1.79%	1956–2013	1.36%	1.44%	1.79%
1951–2004	1.71%	1.94%	2.04%	1951–2014	3.55%	3.65%	3.82%	1956–2014	3.55%	3.65%	3.82%
1951–2005	2.13%	2.42%	2.57%	1951–2015	1.04%	1.14%	1.60%	1956–2015	1.04%	1.14%	1.60%
1951–2006	1.97%	2.22%	2.26%								
1951–2007	2.27%	2.51%	2.56%								
1951–2008	2.89%	3.17%	3.20%								
1951–2009	2.06%	2.27%	2.44%								
1951–2010	2.73%	2.98%	3.21%								
1951–2011	2.03%	2.19%	2.35%								
1951–2012	2.13%	2.14%	2.45%								
1951–2013	1.36%	1.50%	1.75%								
1951–2014	3.55%	3.43%	3.58%								
1951–2015	1.04%	1.20%	1.61%								

Table 11: MAPE values for various recalibration windows

Rolling window				Three-year recalibrate window				Five-year recalibrate window			
Model	3a	3b	3c	Model	3a	3b	3c	Model	3a	3b	3c
1951–2000	1.66%	1.65%	1.67%	1951–2000	1.67%	1.56%	1.61%	1951–2000	1.67%	1.62%	1.70%
1952–2001	1.31%	1.33%	1.27%	1951–2003	3.70%	3.62%	3.62%	1951–2005	2.13%	2.25%	2.28%
1953–2002	1.25%	1.33%	1.34%	1951–2006	1.97%	2.15%	2.32%	1951–2010	2.73%	2.69%	3.09%
1954–2003	3.91%	3.88%	3.74%	1951–2009	2.06%	2.39%	2.64%	1951–2015	1.04%	1.19%	1.77%
1955–2004	1.68%	1.85%	1.72%	1951–2012	2.13%	2.37%	2.74%				
1956–2005	2.18%	2.24%	2.23%	1951–2015	1.04%	1.22%	1.84%				
1957–2006	2.05%	2.15%	2.06%								
1958–2007	2.14%	2.52%	2.43%								
1959–2008	3.03%	3.08%	2.95%								
1960–2009	1.96%	2.30%	2.40%								
1961–2010	2.79%	2.76%	2.93%								
1962–2011	2.10%	2.29%	2.44%								
1963–2012	2.20%	2.27%	2.61%								
1964–2013	1.43%	1.37%	1.77%								
1965–2014	3.44%	3.68%	3.93%								
1966–2015	1.06%	1.07%	1.65%								

Table 12: BIC values for different base years for Model 3

Base 1951–1995				Base 1951–2005				Base 1956–2005			
Model	3a	3b	3c	Model	3a	3b	3c	Model	3a	3b	3c
1951–1995	18192	18192	18192	1951–2005	22004	22004	22004	1956–2005	19914	19914	19914
1951–1996	18572	18580	18580	1951–2006	22418	22422	22424	1956–2006	20317	20321	20323
1951–1997	18955	18994	18996	1951–2007	22838	22849	22853	1956–2007	20734	20742	20745
1951–1998	19316	19379	19383	1951–2008	23266	23289	23299	1956–2008	21160	21175	21182
1951–1999	19695	19780	19789	1951–2009	23685	23726	23750	1956–2009	21571	21599	21615
1951–2000	20062	20151	20158	1951–2010	24097	24162	24208	1956–2010	21982	22027	22057
1951–2001	20434	20534	20543	1951–2011	24529	24620	24704	1956–2011	22417	22479	22540
1951–2002	20808	20919	20935	1951–2012	24875	24977	25088	1956–2012	22762	22836	22918
1951–2003	21177	21286	21303	1951–2013	25273	25383	25538	1956–2013	23164	23246	23365
1951–2004	21612	21729	21745	1951–2014	25635	25758	25940	1956–2014	23529	23623	23768
1951–2005	22004	22127	22149	1951–2015	26050	26162	26401	1956–2015	23951	24039	24233
1951–2006	22418	22517	22543	1951–2016	26408	26526	26801	1956–2016	24311	24404	24632
1951–2007	22838	22953	22985								
1951–2008	23266	23402	23443								
1951–2009	23685	23833	23884								
1951–2010	24097	24264	24323								
1951–2011	24529	24693	24770								
1951–2012	24875	25024	25125								
1951–2013	25273	25414	25559								
1951–2014	25635	25774	25938								
1951–2015	26050	26175	26404								
1951–2016	26408	26534	26794								

Table 13: BIC values for various recalibration windows

Rolling window				Three-year recalibrate window				Five-year recalibrate window			
Model	3a	3b	3c	Model	3a	3b	3c	Model	3a	3b	3c
1951–2000	20062	20062	20062	1951–2000	20062	20062	20062	1951–2000	20062	20062	20062
1952–2001	20023	20039	20008	1951–2003	21177	21185	21190	1951–2005	22004	22018	22037
1953–2002	19990	20030	19969	1951–2006	22418	22440	22468	1951–2010	24097	24208	24270
1954–2003	19941	20001	19952	1951–2009	23685	23767	23840	1951–2015	26050	26213	26582
1955–2004	19913	20015	20008	1951–2012	24875	25032	25233				
1956–2005	19914	20040	19997	1951–2015	26050	26217	26637				
1957–2006	19998	20157	19977								
1958–2007	19901	20097	19987								
1959–2008	19987	20202	20068								
1960–2009	19985	20225	20121								
1961–2010	19971	20206	20177								
1962–2011	20004	20222	20253								
1963–2012	19988	20216	20219								
1964–2013	19911	20264	20147								
1965–2014	19884	20228	20168								
1967–2015	19920	20220	20273								
1966–2016	19908	20267	20295								

5. Concluding remarks

This paper explores the new-data-invariant property of a variety of popular mortality models. For each model, we propose two strategies with different levels of introduced restrictions on the parameters. Using the male population of England and Wales, we demonstrate that the less restricted adapted RH model can achieve the new-data-invariant property, with trivial loss in the goodness of fit and backtesting performance. In addition, the adjusted RH model outperforms the LC, APC, CBD, and LL counterparts and is, therefore, deemed optimal.

The proposed adjustments in this paper will provide easier tracking of mortality indexes, with new data sequentially received in the future. This significantly complements the CBD (or M5) model, as studied by Chan, Li, and Li (2014). However, future research should be conducted analysing other populations and/or different age groups. In particular, a case study of mortality interdependencies can be extended to include a more comprehensive collection of geographically and/or economically connected populations. The findings may facilitate related demographic research, such as that on the geographical spillovers of the mortality dynamics of associated populations. Further, this research focuses on the adaptations of single-population mortality models. Other popular multi-population frameworks, such as the joint-LC model and the coherent functional demographic model, may be further examined to adopt more cross-population information. Finally, long-term forecasts based on the adopted models should be investigated.

References

Barbi, E., Lagona, F., Marsili, M., Vaupel, J.W., and Wachter, K.W. (2018). The plateau of human mortality: Demography of longevity pioneers. *Science* 360(6396): 1459–1461. doi:10.1126/science.aat3119.

Basellini, U., Camarda, C.G., and Booth, H. (2023). Thirty years on: A review of the Lee–Carter method for forecasting mortality. *International Journal of Forecasting* 39(3): 1033–1049. doi:10.1016/j.ijforecast.2022.11.002.

Bergeron-Boucher, M.P., Canudas-Romo, V., Oeppen, J., and Vaupel, J.W. (2017). Coherent forecasts of mortality with compositional data analysis. *Demographic Research* 37(17): 527–566. doi:10.4054/DemRes.2017.37.17.

Booth, H., Hyndman, R., Tickle, L., and De Jong, P. (2006). Lee–Carter mortality forecasting: A multi-country comparison of variants and extensions. *Demographic Research* 15(9): 289–310. doi:10.4054/DemRes.2006.15.9.

Booth, H., Maindonald, J., and Smith, L. (2002). Applying Lee–Carter under conditions of variable mortality decline. *Population Studies* 56(3): 325–336. doi:10.1080/00324720215935.

Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science* 3(1–2): 3–43. doi:10.1017/S1748499500000440.

Brouhns, N., Denuit, M., and Vermunt, J.K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics* 31(3): 373–393. doi:10.1016/S0167-6687(02)00185-3.

Cairns, A.J.G., 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. doi:10.1111/j.1539-6975.2006.00195.x.

Cairns, A.J.G., 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. doi:10.1080/10920277.2009.10597538.

Chan, W.S., Li, J.S.H., and Li, J. (2014). The CBD mortality indexes: Modeling and applications. *North American Actuarial Journal* 18(1): 38–58. doi:10.1080/10920277.2013.854161.

Human Mortality Database (2020). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). http://www.mortality.org.

Hunt, A. and Villegas, A.M. (2015). Robustness and convergence in the Lee–Carter model with cohort effects. *Insurance: Mathematics and Economics* 64: 186–202. doi:10.1016/j.insmatheco.2015.05.004.

Hyndman, R.J. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software* 27(3): 1–22. doi:10.18637/jss.v027.i03.

Kjærgaard, S., Ergemen, Y.E., Bergeron-Boucher, M.P., Oeppen, J., and Kallestrup-Lamb, M. (2020). Longevity forecasting by socio-economic groups using compositional data analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society* 183(3): 1167–1187. doi:10.1111/rssa.12555.

Lee, R. and Miller, T. (2001). Evaluating the performance of the Lee–Carter method for forecasting mortality. *Demography* 38(4): 537–549. doi:10.1353/dem.2001.0036.

Lee, R.D. and Carter, L.R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association* 87(419): 659–671. doi:10.1080/01621459.1992.10475265.

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee–Carter method. *Demography* 42(3): 575–594. doi:10.1353/dem.2005.0021.

Li, N., Lee, R., and Gerland, P. (2013). Extending the Lee–Carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography* 50(6): 2037–2051. doi:10.1007/s13524-013-0232-2.

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. doi:10.1080/10920277.2022.2050260.

Renshaw, A.E. and Haberman, S. (2003). Lee–Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics* 33(2): 255–272. doi:10.1016/S0167-6687(03)00138-0.

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. doi:10.1016/j.insmatheco.2005.12.001.

Tan, C.I., Li, J., Li, J.S.H., and Balasooriya, U. (2014). Parametric mortality indexes: From index construction to hedging strategies. *Insurance: Mathematics and Economics* 59: 285–299. doi:10.1016/j.insmatheco.2014.10.005.

Thinggaard, M., Jeune, B., Osler, M., Vaupel, J.W., McGue, M., and Christensen, K. (2020). Are advances in survival among the oldest old seen across the spectrum of

health and functioning? *The Journals of Gerontology: Series A* 75(12): 2354–2360. doi:10.1093/gerona/glaa009.

Vaupel, J.W., Villavicencio, F., and Bergeron-Boucher, M.P. (2021). Demographic perspectives on the rise of longevity. *Proceedings of the National Academy of Sciences* 118(9): e2019536118. doi:10.1073/pnas.2019536118.