Gompertz Law Revisited: Forecasting Mortality with a Multi-factor Exponential Model

Hong Li[*], Ken Seng Tan[†], Shripad Tuljapurkar[‡] and Wenjun Zhu[§]

First Version: Dec 2017; *Last Version:* Jan 2020

Abstract

This paper provides a flexible way to address some ongoing challenges in mortality modeling, with a special focus on the mortality curvature and possible mortality plateau for extremely old ages. In particular, we extend the Gompertz law (1825) by proposing a multi-factor exponential model, a framework that is able to capture flexible mortality patterns, and allows for a convenient estimation and prediction algorithm. An extensive empirical analysis is conducted using the proposed framework with a merged mortality database containing a large number of countries and regions with credible old-age mortality data. We find that the proposed exponential model leads to superior goodness-of-fit to historical data, and better out-of-sample forecasting performance. Moreover, the exponential model predicts more balanced mortality improvements across ages, and thus leads to higher projected remaining life expectancy for the old ages than existing Gompertz-based mortality models. Finally, the modeling capacity of the proposed exponential model is further demonstrated by a multi-population extension, and an illustrative example of estimation and forecast is provided.

Keywords: Old age mortality forecasting; Gompertz law; Factor model; Mortality curvature; Mortality plateau.

[*]Corresponding author. li.hong@umanitoba.ca. Warren Centre for Actuarial Studies and Research, University of Manitoba, Canada.

[†]kenseng.tan@ntu.edu.sg. Division of Banking & Finance and Insurance Risk & Finance Research Centre, Nanyang Business School, Nanyang Technological University, Singapore.

[‡]tulja@stanford.edu. Department of Biology, Stanford University, USA.

[§]wjzhu@ntu.edu.sg. Division of Banking & Finance, Nanyang Business School, Nanyang Technological University, Singapore.

1 Introduction

Human life expectancy has been increasing around the globe. Over the last two centuries, life expectancy at birth has more than doubled worldwide, and improved even more in developed regions (Riley, 2001). For example, according to the World Bank, life expectancy increased from a global average of 52.57 years in 1960 to 72.23 years in 2017 (World Bank, 2019). A main driver of this increase in life expectancy is the reduction in mortality at older ages since 1950s (Rau et al., 2008; Vaupel, 1997; Vaupel et al., 1998). Such old age mortality deceleration leads to an accumulation of the number of nonagenarians, centenarians, and supercentenarians, especially in developed regions (see, among others, Thatcher, 1999b; Desjardins and Bourbeau, 2010; Kestenbaum and Ferguson, 2010). In fact, as shown by World Bank, the old age dependency ratio, defined as the ratio of the population above 65 over the working age population, has increased from 20.37% to 27.48% for the high income regions, compared with a worldwide contemporary increase from 10.94% to 13.67%.[1] As a result, forecasting mortality of senior populations has been a crucial task in demographic analysis.

One of the earliest attempts to model human mortality was made by the English actuary Benjamin Gompertz in 1825. In his seminal work, Gompertz (1825) argued that, starting from a certain age, the logarithm of force of mortality is a linear function of age. This assumption, later referred to as the Gompertz law, has been used widely in demographic and actuarial projections of mortality in the past two centuries. However, starting from the intriguing finding by Greenwood and Irwin (1939) that Gompertz's formula may over-state senile mortality, a growing number of studies suggest that mortality at old ages tend to deviate from the Gompertz law (Horiuchi and Wilmoth, 1998; Thatcher, 1999a; Gavrilov and Gavrilova, 1991; Olshansky, 1998). They argue that mortality rates at high ages tend to exhibit curvature, rather than being linear. This phenomenon is also documented in Cairns et al. (2009), Cairns et al. (2011a), and Currie (2011), among others. Moreover, mortality increases may slow down, or even be stagnated, at very high ages - a phenomenon known as the "mortality leveling-off" or the old age "mortality plateau" (Curtsinger et al., 2006). For example, Barbi et al. (2018) provide evidence for the existence of extreme-age mortality plateaus in humans. By analyzing the International Database on Longevity, Gampe (2010, p. 227) concludes that "*human mortality after age 110 is flat*".

These features and findings at old ages have become increasingly important issues for accurately estimating and forecasting mortality. Ignoring them would lead to underestimation of life expectancy and longevity risk. In fact, official population projections have been persistently overestimating mortality rates and underestimating life expectancy, particularly in developed counties, including the UK (Shaw, 2007), Australia, Canada, Japan, New Zealand, and the U.S, (National Research Council, 2000).

The objective of this paper is to propose a flexible multi-factor exponential model ca-

[1] Data source: https://data.worldbank.org.

pable of modeling different types of mortality curves: increasing, decelerating, or flat. In particular, the proposed model is based on the Laguerre approximation, captures a range of mortality patterns including curvature and an old age plateau, and permits a convenient estimation and prediction algorithm. Empirical analysis in this paper demonstrates that the proposed exponential model has better in-sample fitness and out-of-sample prediction power. In terms of the long-term forecasting, the proposed model produces mortality projections with patterns more similar to historical data, and predicts higher remaining life expectancy at the very advanced ages. Finally, we show that our proposed model can easily be extended to modeling multi-population mortality rates. We note an earlier study by Haldrup and Rosenskjold (2019) who used the first order Laguerre function to capture the infant and the accident hump loading of mortality.

It is worth noting that one difficulty which hinders mortality modeling at very old ages is data credibility. First, the number of survivors at these ages has been very small for many countries, making the estimated mortality rates unreliable (Gavrilov and Gavrilova, 2011, 2019a). Second, age misreporting by the very old and data aggregation cause additional methodological problems to the analysis of old age mortality. More seriously, researches conducted by, e.g., Gavrilov and Gavrilova (2019a); Newman (2018a,b), question the presence of late-life mortality deceleration and plateau and argue that these phenomena can be attributed to age misreporting. See also the discussion in Gavrilov and Gavrilova (2019b).Finally, smoothing and extrapolation techniques are often applied on old age mortality data, and some of which may affect the raw mortality patterns. To cope with these issues, part of the empirical analysis of this paper is based on a merged data set that combines the Human Mortality Database (HMD) and the Kannisto-Thatcher Database (KTD). In particular, mortality data for ages below 80 are obtained from the HMD and those older than 80 are obtained from the KTD. The KTD is an assembled data set of old age mortality that has been tested for quality. Hence using KTD for very old ages would improve the credibility of estimating mortality at high ages (say above age 80).

The remainder of this paper is organized as follows. Section 2 reviews the Gompertz law and its variations. Section 3 describes our proposed multi-factor exponential model, including its mathematical background and estimation algorithm. Section 4 provides a detailed description on the database and presents some extensive empirical results on the merged data. To provide additional evidence on the robustness of our proposed multi-factor exponential model, Section 5 extends the empirical analysis by considering more populations which are available in HMD. Section 6 discusses the multi-population extension of the proposed model. Section 7 concludes the paper.

2 Gompertz law and its variations

Gompertz (1825) stipulates that the mortality pattern, after a certain age x_0, is given by

$$\log \mu_x = a + bx, \tag{2.1}$$

where μ_x is the force of mortality at age x, a is the level parameter, and b is the slope parameter. The above Gompertz law of mortality assumes that the logarithm of the force of mortality increases linearly with age. In other words, mortality rate increases exponentially with age after age x_0. As pointed out in the preceding section that the mortality pattern at high ages remains a controversial issue. Some argue that the mortality rate will continue to increase at high ages while others justify the presence of late-life mortality deceleration and plateau. In anticipation of the potentially more complex mortality pattern at high ages, several generalizations of the Gompertz law have been proposed. One possible generalization is to consider other transformations of mortality instead of the logarithm (Horiuchi and Wilmoth, 1998). In particular, a popular choice is the logistic extension of mortality, known as the Kannisto formula (Kannisto, 1994; Thatcher et al., 1998):

$$\text{logit } \mu_x = a + xb,$$

where $\text{logit } \mu_x = \dfrac{\mu_x}{1-\mu_x}$. The Kannisto formula is later generalized to the dynamic context by Cairns et al. (2006). The resulting model, which is denoted as the CBD model, admits the following representation:

$$\text{logit } q_{x,t} = \kappa_{1,t} + \kappa_{2,t}(x - \bar{x}), \tag{2.2}$$

where $\bar{x}$ is the mean age, and $\kappa_{1,t}$ and $\kappa_{2,t}$ are two latent factors. $\kappa_{1,t}$ represents the average mortality level, while $\kappa_{2,t}$ is the slope of $\text{logit}\,(q_{x,t})$ over age in year t. By making $\kappa_{1,t}$ and $\kappa_{2,t}$ time-varying, the CBD model allows for changes in both the aggregate mortality level and the mortality slope over time. In the CBD model, the left hand side is a logit transformation of the death probability $q_{x,t}$ for age x in year t. In actuarial science literature, $q_{x,t}$ is commonly approximated by the force of mortality, i.e., $q_{x,t} \approx 1 - \exp(-\mu_{x,t})$.[2]

In order to address the issue of curvature, an extension to the CBD model with a quadratic term has been proposed (Cairns et al., 2009, 2011a):

$$\text{logit } q_{x,t} = \kappa_{1,t} + \kappa_{2,t}(x - \bar{x}) + \kappa_{3,t}((x - \bar{x})^2 - \sigma_x^2), \tag{2.3}$$

where σ_x^2 is the mean of $(x - \bar{x})^2$. For ease of referencing, we refer the original CBD model

[2]This approximation implicitly assumes a stationary population and that the force of mortality over each year of integer age and over each calendar year is constant (see, for example, Pitacco et al. (2009) and Cairns et al. (2009)). Although the constant force of mortality does not normally hold exactly, the resulting relationship between $\mu_{x,t}$ and $q_{x,t}$ generally provides an accurate approximation.

(2.2) as the CBD-L model and the CBD model (2.3) as the CBD-Q model.[3] However, the quadratic specification still imposes strong restrictions on the mortality shape, and hence the improvement is somewhat limited. There are some other researches that attempt to address the issue of curvature. For example, Currie (2011) replaces the linear term $(x - \bar{x})$ in the original CBD model by a non-parametrically smoothed term, and the model obtains similar fitting results as the CBD-Q model. As will be discussed in the next section, we propose to capture the mortality curvature via an exponential model.

3 The multi-factor exponential model

In this section, we propose a multi-factor exponential model based on the class of Laguerre functions as a viable alternative. Subsection 3.1 introduces the model specification, including the motivation and the decomposition framework, Subsection 3.2 discusses the estimation procedure of the proposed model, and Subsection 3.3 illustrates the benefit of the proposed model with an estimation example based on the female mortality data of the U.S.

3.1 Model specification

In this paper, we propose using the following decomposition of logit $q_{x,t}$ to capture various characteristics of mortality curve; i.e. the level, the slope, the curvature and the mortality plateau:

$$\text{logit } q_{x,t} = \kappa_{1,t} + \kappa_{2,t}(x - \delta_{\boldsymbol{x}}) + \kappa_{3,t}\left(\frac{1 - e^{-\lambda_1(x-\delta_{\boldsymbol{x}})}}{\lambda_1(x - \delta_{\boldsymbol{x}})} - \sigma_3\right) + \kappa_{4,t}\left(\frac{1 - e^{-\lambda_2(x-\delta_{\boldsymbol{x}})}}{\lambda_2(x - \delta_{\boldsymbol{x}})} - e^{-\lambda_2(x-\delta_{\boldsymbol{x}})} - \sigma_4\right). \tag{3.1}$$

Here $\delta_{\boldsymbol{x}}$ is a parameter as a function of the sample age, $\boldsymbol{x} = \{x_0, x_1, \ldots, x_{\max}\}$, parameter λ_1 controls the exponential growth rate, and parameter λ_2 reflects the exponential decay rate. The combination of λ_1 and λ_2 provides a flexible way of capturing the curvature in the mortality curve and describing the possible old age mortality plateau. Finally, σ_3 and σ_4 are the means of $\frac{1-e^{-\lambda_1(x-\delta_{\boldsymbol{x}})}}{\lambda_1(x-\delta_{\boldsymbol{x}})}$ and $\frac{1-e^{-\lambda_2(x-\delta_{\boldsymbol{x}})}}{\lambda_2(x-\delta_{\boldsymbol{x}})} - e^{-\lambda_2(x-\delta_{\boldsymbol{x}})}$ over age, respectively.

Mathematical motivation of the exponential model

Our proposed exponential model (3.1) is motivated from approximating mortality measures with Laguerre functions. Let us denote by $f(x,t)$ the instantaneous infinitesimal increment of logit $q_{x,t}$ at age x in year t. A meticulous choice of $f(x,t)$ could model $q_{x,t}$ (via logit transformation) with desirable features. Here we propose to model $f(x,t)$ through Laguerre

[3] CBD-Q model (2.3) is a special case of model M7 in Cairns et al. (2009) without a cohort factor. A cohort effect is not considered here because the focus of this paper is on the curvature of logit $(q_{x,t})$. The cohort effect can be included whenever necessary.

functions as follows:

$$f(x,t) = \eta_{1,t} + \eta_{2,t}x + \eta_{3,t}e^{-\lambda_1 x} + \eta_{4,t}\lambda_2 x e^{-\lambda_2 x}, \tag{3.2}$$

where, to be shown below, $\eta_{1,t}, \eta_{2,t}, \eta_{3,t}$, and $\eta_{4,t}$ are related to $\kappa_{1,t}, \kappa_{2,t}, \kappa_{3,t}$, and $\kappa_{4,t}$.

Equation (3.2) can also be viewed as a linear term plus the solution to a third-order differential equation with one unique and two repeated real roots. Integrating both sides of Equation (3.2) from 0 to x, and normalizing the integrals by x gives:

$$F(x,t) = \frac{1}{x}\int_0^x f(s,t)ds = \eta_{1,t} + \frac{1}{2}\eta_{2,t}x + \eta_{3,t}\left(\frac{1-e^{-\lambda_1 x}}{\lambda_1 x}\right) + \eta_{4,t}\left(\frac{1-e^{-\lambda_2 x}}{\lambda_2 x} - e^{-\lambda_2 x}\right). \tag{3.3}$$

Reparameterizing the time-varying factors then yields:

$$F(x,t) = \kappa_{1,t} + \kappa_{2,t}x + \kappa_{3,t}\left(\frac{1-e^{-\lambda_1 x}}{\lambda_1 x} - \sigma_3\right) + \kappa_{4,t}\left(\frac{1-e^{-\lambda_2 x}}{\lambda_2 x} - e^{-\lambda_2 x} - \sigma_4\right), \tag{3.4}$$

where $\kappa_{1,t} = \eta_{1,t} + \sigma_3\eta_{3,t} + \sigma_4\eta_{4,t}$, $\kappa_{2,t} = \frac{1}{2}\eta_{2,t}, \kappa_{3,t} = \eta_{3,t}$, and $\kappa_{4,t} = \eta_{4,t}$. Replacing x with a more general linear term $x - \delta_{\boldsymbol{x}}$ in Equation (3.4) leads us to a mortality model in Equation (3.1).

Our proposed exponential model has several theoretical advantages. First, as a class of complete approximating functions, linear combinations of Laguerre functions can approximate any piecewise continuous function $f(x)$ (for which the integral $\int_0^\infty f^2(x)dx$ exists) arbitrarily well in the mean.[4] Second, the proposed exponential model can be treated as an infinite sum of polynomial terms with increasing order of age, making it a more natural generalization of the existing models than adding higher order terms. Third, in situations where the first two Laguerre functions cannot reasonably capture the underlying mortality patterns, one can easily improve the model by adding higher order Laguerre functions.

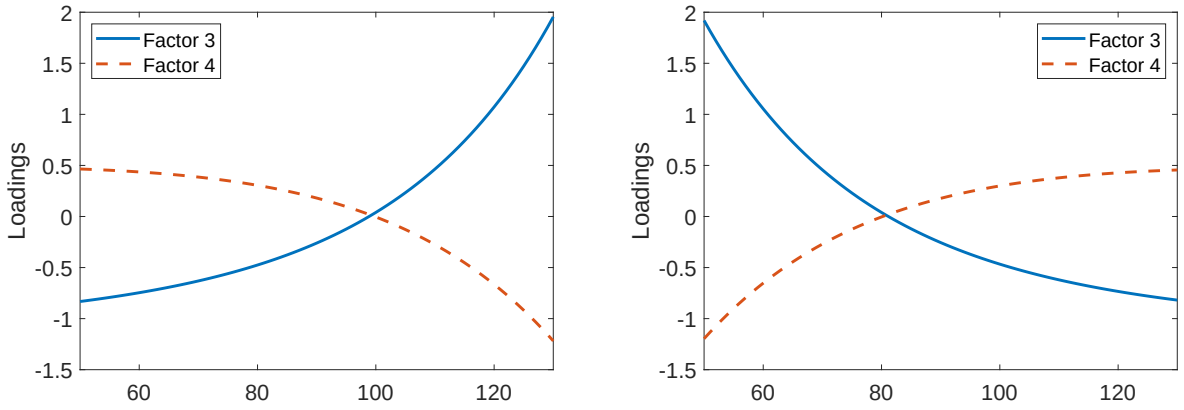
Demographic interpretations of the exponential model

The proposed exponential model has appealing demographic interpretations. Under the framework (3.1), the mortality curve is decomposed into four factors. The first two factors, $\kappa_{1,t}$ and $\kappa_{2,t}$, have similar interpretation as in the original CBD model in that they represent, respectively, the level and the slope. The two additional factors, $\kappa_{3,t}$ and $\kappa_{4,t}$, capture the possible mortality curvature and plateau. For illustration, we plot sample functions $L_3(x) = \frac{1-e^{-\lambda_1(x-\delta_{\boldsymbol{x}})}}{\lambda_1(x-\delta_{\boldsymbol{x}})} - \sigma_3$ (the loading on $\kappa_{3,t}$) and $L_4(x) = \frac{1-e^{-\lambda_2(x-\delta_{\boldsymbol{x}})}}{\lambda_2(x-\delta_{\boldsymbol{x}})} - e^{-\lambda_2(x-\delta_{\boldsymbol{x}})} - \sigma_4$ (the loading on $\kappa_{4,t}$) in Figure 1 for two pairs of (λ_1, λ_2); i.e. $(-0.05, -0.03)$ and $(0.05, 0.03)$, and with $\delta_{\boldsymbol{x}} = \bar{x} + 0.25$. We observe that $L_3(x)$ is convex but it can increase or decrease with age depending on the sign of λ_1. Similarly for $L_4(x)$ except that it behaves in the opposite direction to $L_3(x)$. Therefore, $L_3(x)$ or $L_4(x)$ captures the curvature of the mortality curve

[4]For more discussions on Laguerre functions, see, for example, Courant and Hilbert (1953).

while both functions $L_3(x)$ and $L_4(x)$ combined capture the possible mortality plateau.

While the presence of late-life mortality deceleration and plateau is controversial, here we point out that several theories have been hypothesized to account for the observed old age phenomenon. In particular, Beard in 1959 postulated the concept of population heterogeneity. Population heterogeneity causes a decreased slope of the mortality curve because *"sub-populations with the higher injury levels die out more rapidly, resulting in progressive selection for vigour in the surviving populations"* (Sacher, 1968). See also Horiuchi and Wilmoth (1998). The proposed exponential model provides a data-driven framework for modelling the possible population heterogeneity.



(a) $\lambda_1 = -0.05, \lambda_2 = -0.03$ (b) $\lambda_1 = 0.05, \lambda_2 = 0.03$

Figure 1: **Sample loadings on Factor 3 and Factor 4 for two different pairs of λ_1 and λ_2. x-axis is age.**

Trade-off between flexibility and model parsimony

In term of empirical application, care needs to be taken for the trade-off between flexibility and parsimony. On the one hand, a flexible model with many parameters tends to provide better in-sample goodness of fit, but is prone to overfitting. On the other hand, a parsimonious model with fewer parameters could achieve more robust out-of-sample forecasting but at the expense of in-sample fit. For our proposed framework (3.1), the model can be adjusted appropriately to reflect the trade-off. In particular, we consider the following two specifications for our empirical analysis:

- One-Exponential-Factor Model (EM1f)

$$\text{logit } q_{x,t} = \kappa_{1,t} + \kappa_{2,t}(x - \bar{x}) + \kappa_{3,t}\left(\frac{1 - e^{-\lambda_2(x-\delta_{\boldsymbol{x}})}}{\lambda_2(x - \delta_{\boldsymbol{x}})} - e^{-\lambda_2(x-\delta_{\boldsymbol{x}})} - \sigma_4\right), \tag{3.5}$$

- Two-Exponential-Factor Model (EM2f)

$$\text{logit } q_{x,t} = \kappa_{1,t} + \kappa_{2,t}(x-\bar{x}) + \kappa_{3,t}\left(\frac{1-e^{-\lambda_1(x-\delta_{\boldsymbol{x}})}}{\lambda_1(x-\delta_{\boldsymbol{x}})} - \sigma_3\right) + \kappa_{4,t}\left(\frac{1-e^{-\lambda_2(x-\delta_{\boldsymbol{x}})}}{\lambda_2(x-\delta_{\boldsymbol{x}})} - e^{-\lambda_2(x-\delta_{\boldsymbol{x}})} - \sigma_4\right), \tag{3.6}$$

where $(x-\delta_{\boldsymbol{x}})$ is set to be $x-\bar{x}+0.25$. The value of 0.25 is imposed simply to avoid possible zero value, as $(x-\delta_{\boldsymbol{x}})$ appears in the denominator and ages are integers. Using other values of k leads to very similar empirical results. Note that EM1f is a three-factor model involving one additional dynamic factor, $\kappa_{3,t}$, for the exponential terms, while EM2f is a four-factor model with two additional dynamic factors, $\kappa_{3,t}$ and $\kappa_{4,t}$, for the exponential terms. Hence EM2f is more flexible than EM1f. The validity of these models will be tested statistically by empirically applying to various populations. In our empirical analysis,we set the linear term (i.e., the slope) to be $x-\bar{x}$, as in the CBD models, for ease of comparison.

3.2 Estimation of the exponential models

One way to estimate the proposed exponential model is by nonlinear least squares. However, since there is a large number of parameters to be estimated in this model for T years of mortality experience, there are $3T+2$ parameters for EM1f and $4T+2$ parameters for EM2f), and λ_1 and λ_2 interacts with $\kappa_{3,t}$'s and $\kappa_{4,t}$'s in a highly nonlinear way, the nonlinear least squares may be time-consuming and produce inflated standard errors.

Alternatively, we can estimate EM1f and EM2f using the method of maximum likelihood estimation (MLE). Assume that the number of deaths $D_{x,t}$ in a population at age x last birthday during calendar year t follows a Poisson distribution: $D_{x,t} \sim Poi(E_{x,t}\mu_{x,t})$, where $E_{x,t}$ is the average size of the population aged x last birthday during year t. Let $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\kappa}_t)$ be a vector that collects all the parameters to be estimated in the EM models. More specifically, we have $\boldsymbol{\theta} = (\lambda_1, \lambda_2, \kappa_{1,t}, \kappa_{2,t}, \kappa_{3,t})_{t=t_0,\ldots,t_{\max}}$ for EM1f and $\boldsymbol{\theta} = (\lambda_1, \lambda_2, \kappa_{1,t}, \kappa_{2,t}, \kappa_{3,t}, \kappa_{4,t})_{t=t_0,\ldots,t_{\max}}$ for EM2f. The corresponding log-likelihood function can be expressed, in general, as:

$$\boldsymbol{\ell}(D_{x,t}, E_{x,t}; \boldsymbol{\theta}) = \sum_{t=t_0}^{t_{\max}} \ell_t(D_{x,t}, E_{x,t}; \boldsymbol{\theta}_t), \tag{3.7}$$

where $\ell_t(D_{x,t}, E_{x,t}; \boldsymbol{\theta}_t)$ is the log-likelihood function in year t, i.e.,

$$\begin{aligned}\ell_t(D_{x,t}, E_{x,t}; \boldsymbol{\theta}_t) = \sum_{x=x_0}^{x_{\max}} D_{x,t} \log\left[E_{x,t}\log(1+\exp(\text{logit } q_{x,t}))\right] \\ - E_{x,t}\log\left[1+\exp(\text{logit } q_{x,t})\right] - \log(D_{x,t}!),\end{aligned}$$

where $\boldsymbol{\theta}_t = (\lambda_1, \lambda_2, \kappa_{1,t}, \kappa_{2,t}, \kappa_{3,t})$ for EM1f and $\boldsymbol{\theta}_t = (\lambda_1, \lambda_2, \kappa_{1,t}, \kappa_{2,t}, \kappa_{3,t}, \kappa_{4,t})$ for EM2f, and logit $q_{x,t}$ follows the model specification (3.5) or (3.6), depending on which model we are estimating. The method of MLE is implemented via a two-stage procedure, involving an initialization stage and an iteration stage. In the initialization stage, we use ordinary least square to determine the starting values of the parameters, $\boldsymbol{\theta}^0$. In the iteration stage, the

parameters are re-estimated by maximizing the log-likelihood (3.7) with $\boldsymbol{\theta}^0$ as initial inputs. The estimation procedure is summarized in Algorithm 1.

Input • Set $X = x_{\max} - x_0 + 1$ and $T = t_{\max} - t_0 + 1$
• A $X \times T$ matrix of mortality data: $(\text{logit } q_{x,t})_{(X\times T)}$
• Tolerance level ϵ

Output • Estimated parameters $\boldsymbol{\theta}^* = (\boldsymbol{\lambda}^*, \boldsymbol{\kappa}_t^*)$

(1) Initialization:

(1.1) Specify N candidates of λ: $\lambda^{(1)}$, $\lambda^{(2)}$, ..., $\lambda^{(N)}$;

(1.2) **for** $n \leftarrow 1$ **to** N **do**

Set $\lambda_1 = \lambda_2 = \lambda^{(n)}$, and estimate $\boldsymbol{\kappa}_t^{(n)}$ using ordinary least square for each t;

end

(1.3) Choose λ^{0*} that minimizes the sum of squared error:
$\sum_{t=t_0}^{t_{\max}} \sum_{x=x_0}^{x_{\max}} (\text{logit } q_{x,t} - \widehat{\text{logit } q_{x,t}})^2$, and record the corresponding $\boldsymbol{\kappa}_t^{0*}$;

(2) Iteration of MLE:

(2.1) Set $\boldsymbol{\theta}^0 = (\lambda^{0*}, \lambda^{0*}, \boldsymbol{\kappa}_t^{0*})_{t=t_0,\ldots,t_{\max}}$, $\boldsymbol{\ell}^0 = \boldsymbol{\ell}(D_{x,t}, E_{x,t}; \boldsymbol{\theta}^0)$, $\boldsymbol{\ell}^1 = \boldsymbol{\ell}^0 + 1$;

(2.2) **while** $|\boldsymbol{\ell}^1 - \boldsymbol{\ell}^0| > \epsilon$ **do**

Set $\boldsymbol{\ell}^0 = \boldsymbol{\ell}^1$;

for $t \leftarrow t_0$ **to** $t_{\max}$ **do**

Seek $\boldsymbol{\kappa}_t$ that maximizes, ℓ_t, the log-likelihood in year t;

end

Estimate $\boldsymbol{\lambda}$ that maximizes, $\boldsymbol{\ell}$, the total log-likelihood;

Update $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\kappa}_t)_{t=t_0,\ldots,t_{\max}}$, $\boldsymbol{\ell}^1 = \boldsymbol{\ell}(D_{x,t}, E_{x,t}; \boldsymbol{\theta})$;

end

Return $\boldsymbol{\theta}^* = \boldsymbol{\theta}$.

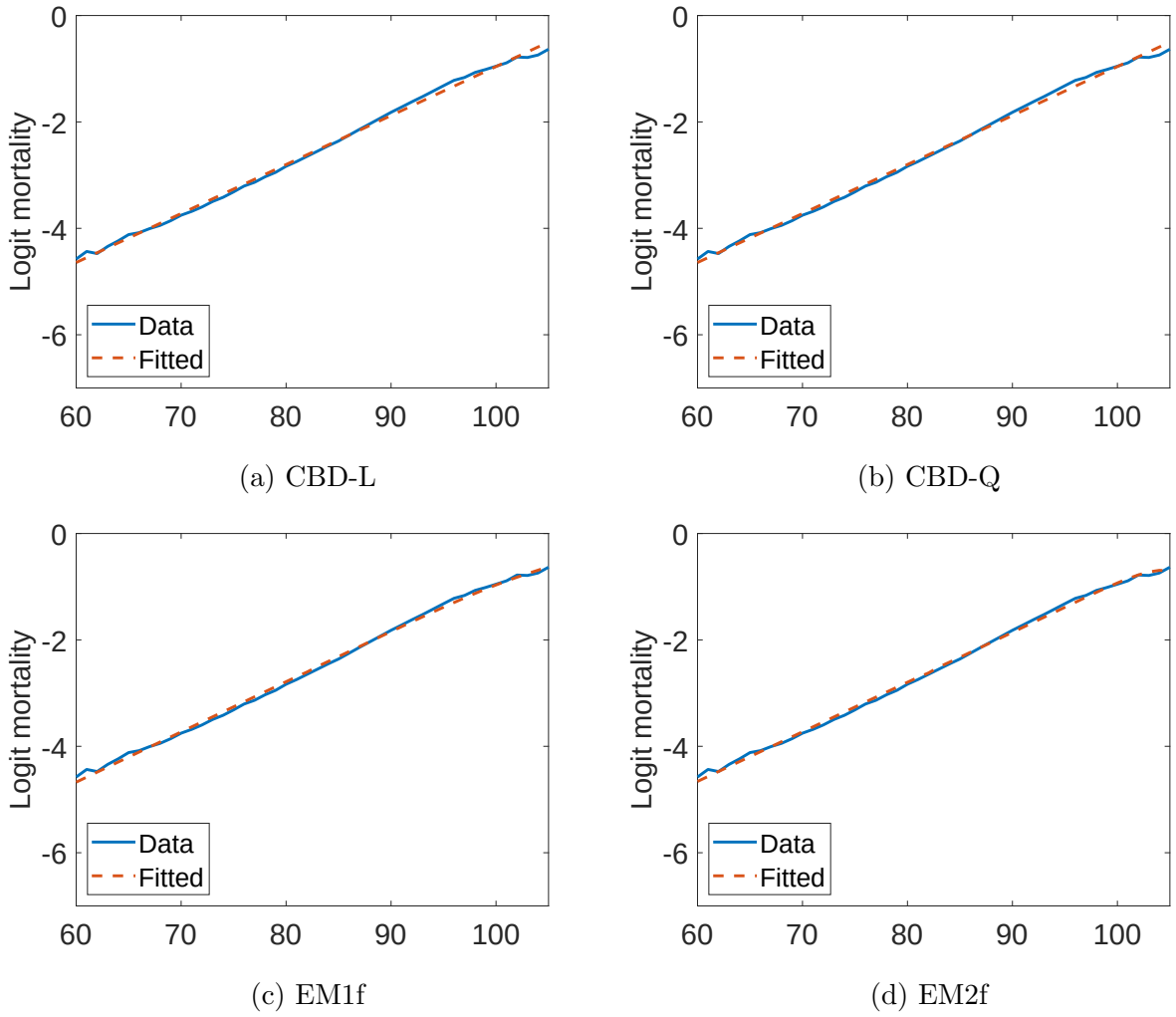
Algorithm 1: Estimating parameters of the exponential model based on the method of MLE.

3.3 Application of the exponential model: illustration

In this subsection, we illustrate the goodness-of-fit of the exponential models using the U.S. female mortality data between age 50 and 105 and 1959 to 2006.[5] Figure 2 displays the observed and the fitted curves of logit $q_{x,t}$ based on different models. We see that the logit transformation of mortality exhibit curvatures, and the late-life mortality plateau is preserved. For CBD-L model (Figure 2a), while it is able to capture general mortality

[5]Mortality data used in this paper are compiled from the Human Mortality Database and the Kannisto-Thatcher database. 2006 is the most recent year in the merged data set for the U.S. See Section 4.1 for data description.

increase over age, it misses the nonlinear mortality patterns, due to its linear assumption. In CBD-Q model (Figure 2b), we observe slight improvement of the data fitness with the quadratic extension. However, the quadratic specification still exerts strong restrictions on the mortality shape, and hence the improvement is somewhat limited. Comparing the fitted curves from EM1f (Figure 2c) and EM2f (Figure 2d), the fitness improvement compared to CBD-L and CBD-Q models is substantial, with the nonlinear mortality patterns better captured.



(a) CBD-L

(b) CBD-Q

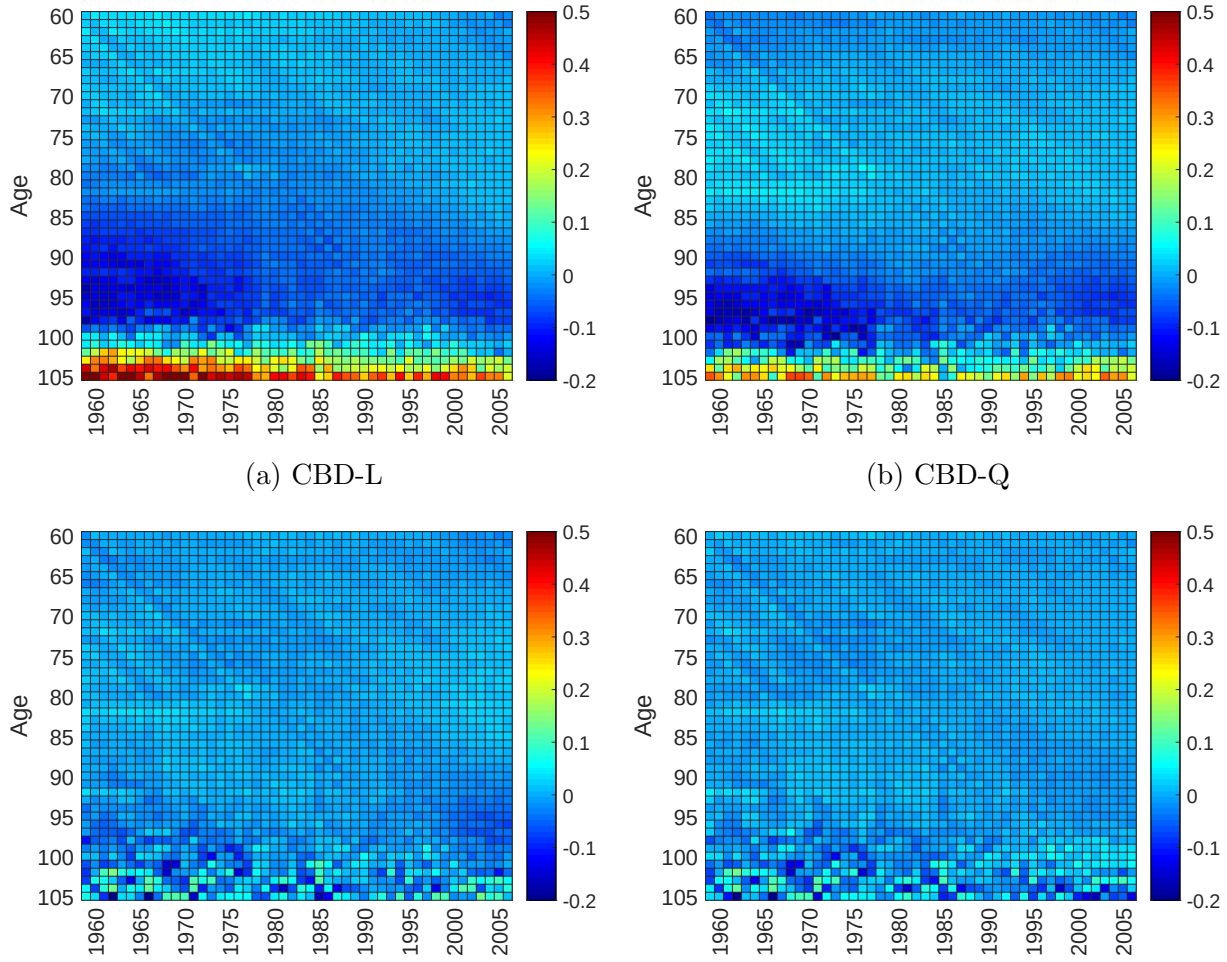
(c) EM1f

(d) EM2f

Figure 2: **Observed and fitted mortality curves over age.** The figures display observed (solid lines) and fitted (dashed lines) logit $(q_{x,t})$ for CBD-L, CBD-Q, EM1f, EM2f models fitted to the U.S. female population in 2006. x-axis is age.

To further evaluate their goodness-of-fit, we investigate the residual of the four models. Figure 3 displays the heat maps of the standardized residuals $(\tilde{\varepsilon}_{x,t} = \frac{\text{logit } q_{x,t} - \widehat{\text{logit } q_{x,t}}}{\text{logit } q_{x,t}})$ for CBD-L, CBD-Q, EM1f, EM2f models fitted to the U.S. female populations. We can see that the standardized residuals from the CBD-L and CBD-Q models exhibit systematic pattern of

over/under estimation. In particular, these two models tend to mildly overestimate the logit of death probabilities for the young and old ages in all years (with much more pronounced deviations at the oldest ages), but tend to underestimate the death probabilities for the ages in middle. In contrast, the distributions of standardized residuals from both exponential models appear to be much more concentrated and random. This observation indicates that the proposed exponential terms in the EM models are effective at capturing the observed curvature and plateau of the mortality data.



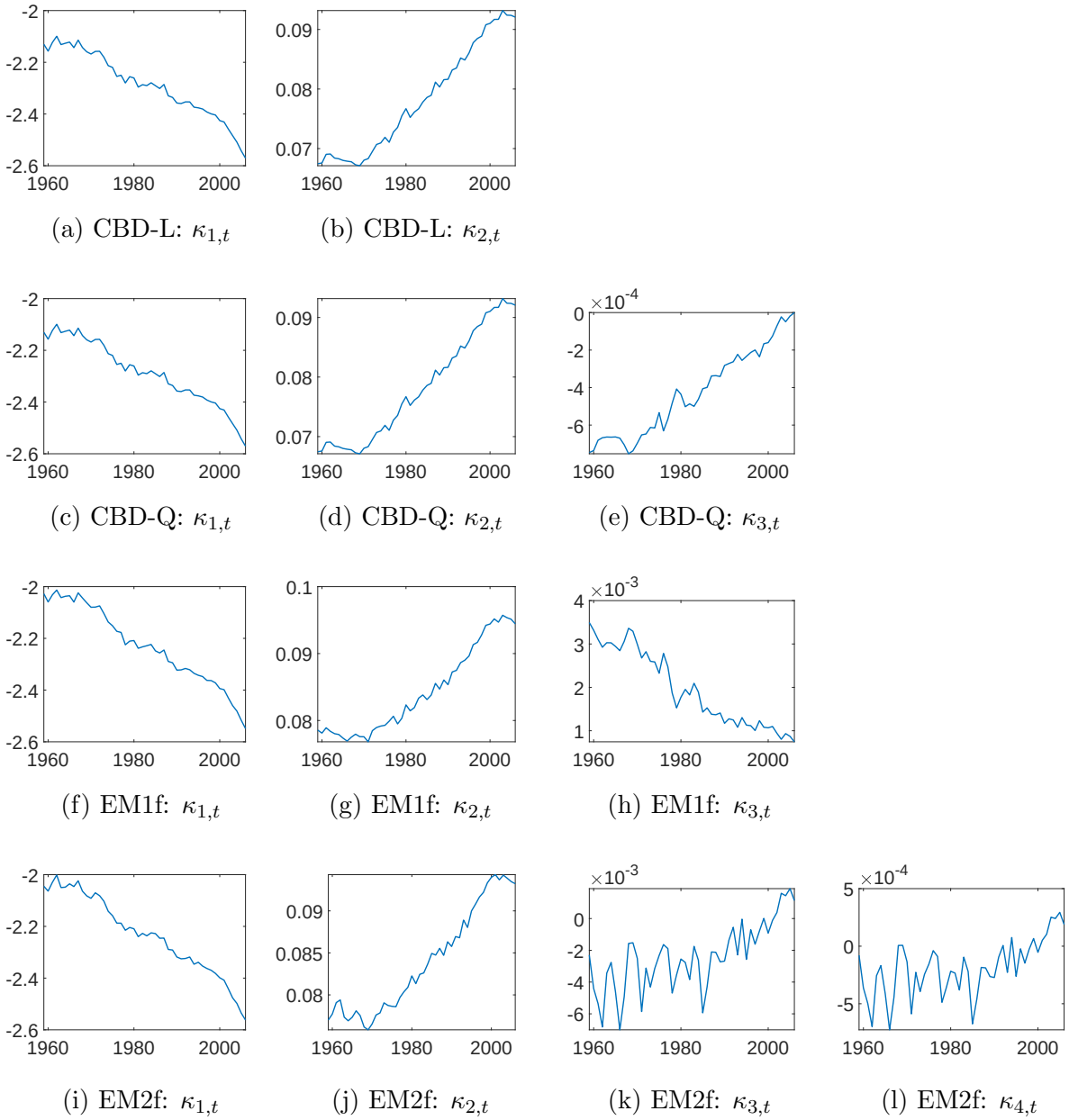
(a) CBD-L (b) CBD-Q

(c) EM1f (d) EM2f

Figure 3: **Standardized residuals heat maps.** The figures show the heat maps of standardized residuals for the CBD-L, CBD-Q, EM1f, EM2f models fitted to the U.S. female population, plotted against ages and years.

Compare the estimated time-varying parameters, $\boldsymbol{\kappa}_t$, as shown in Figure 4, we observe that $\kappa_{3,t}$ and $\kappa_{4,t}$ from EM2f produce very similar patterns, except that $\kappa_{4,t}$ has smaller magnitude than $\kappa_{3,t}$. This leads to the concern of possible over-parameterization issue discussed earlier. Although EM2f preserves some advantages of better goodness-of-fit, as shown from

the fitted curves in Figure 2 and its residuals in Figure 3, such better fitness may simply be induced by having more parameters, and may have no benefit for forecasting. Therefore, in the empirical application, it is important to assess the trade-off between flexibility and parsimony in deciding which model to use. We will revisit this issue in the following sections.



(a) CBD-L: $\kappa_{1,t}$ (b) CBD-L: $\kappa_{2,t}$

(c) CBD-Q: $\kappa_{1,t}$ (d) CBD-Q: $\kappa_{2,t}$ (e) CBD-Q: $\kappa_{3,t}$

(f) EM1f: $\kappa_{1,t}$ (g) EM1f: $\kappa_{2,t}$ (h) EM1f: $\kappa_{3,t}$

(i) EM2f: $\kappa_{1,t}$ (j) EM2f: $\kappa_{2,t}$ (k) EM2f: $\kappa_{3,t}$ (l) EM2f: $\kappa_{4,t}$

Figure 4: **Estimated time-varying factors, $\boldsymbol{\kappa}_t$, for the U.S. data over year.** The figures display $\boldsymbol{\kappa}_t$ estimated from the CBD-L, CBD-Q, EM1f, EM2f models, fitted to the U.S. female population. X-axis is year.

4 Empirical analysis

In this section, we extend the empirical analysis by considering mortality data of 22 population. The data sets used in the empirical analysis are described in Subsection 4.1. The in-sample fit and the out-of-sample forecast performance of each model are compared in Subsections 4.2 and 4.3, respectively. Finally, the long-term forecasts produced by different models are presented in Subsection 4.4.

4.1 Data description

In our empirical analysis, the data come from two sources: the Human Mortality Database (HMD) and the Kannisto-Thatcher database (KTD)[6]. The HMD collects a rich set of mortality data in 41 countries and regions, with sample length typically longer than 50 years, while The KTD, with sample length typically around 50 years, contains mortality data at old ages of 35 countries and regions. Moreover, the KTD is an assembled data set and has been tested for quality. Given that our focus is on old age mortality modeling, the KTD is particularly suited for our empirical studies in view of its data credibility. For this reason we consider a merged data set, with the mortality data for ages younger (resp. older) than 80 obtained from the HMD (resp. KTD). We use 80 as the splitting age as it is the youngest age for which the mortality data is available in the KTD.

The empirical analysis in this paper performs estimation for the merged data set that covers 22 countries and regions whose life expectancy at birth was higher than 80 in 2018. Abbreviations of the counties/regions used in this analysis are collected in Appendix A. For countries and regions with a small population, there are not enough observed deaths at the very old ages, especially in the early years. With the above limitations in mind, we use uni-sex mortality data between ages 60 and 105 to fit the models.[7]

4.2 In-sample fitness

This subsection focuses on in-sample goodness of fit. To this end, information criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are commonly used. Formally, the AIC and BIC are defined respectively as:

$$\begin{aligned} \text{AIC} &= -2\log\left(\widehat{\boldsymbol{\ell}}\right) + 2k, \\ \text{BIC} &= -2\log\left(\widehat{\boldsymbol{\ell}}\right) + k \cdot \log(N), \end{aligned}$$

[6]HMD source: http://www.mortality.org. KTD source: http://www.demogr.mpg.de/cgi-bin/databases/ktdb/datamap.plx.

[7]The estimated results on male and female populations are qualitatively similar to that based on the uni-sex mortality data. These results are available upon request.

where $\hat{\boldsymbol{\ell}}$ is the estimated value of the likelihood function, k is the number of parameters in the model, and N is the total sample size. Models with lower values of AIC and BIC are preferred.

The resulting AIC and BIC for each population and all four models are displayed in Table 1. The better in-sample goodness of fit of the exponential models, relative to the CDB models, is clearly demonstrated, as can be seen from their lower values of AIC's and BIC's. In particular, using the criterion of AIC, three populations select EM1f as their optimal model while the remaining 19 populations are best modeled by EM2f. Under the BIC, additional penalty is imposed by the size of the sample data. Due to the stronger penalty of additional parameters, the BIC criterion suggests that 7 populations selected EM1f and 14 populations selected EM2f. One population; i.e., Iceland, selects the most parsimonious CBD-L as the optimal model under BIC.

4.3 Out-of-sample forecast

The preceding subsection shown that, statistically, the exponential models have better goodness of fit than the CBD models for all the 22 populations. It is, however, important to keep in mind that better in-sample fitness does not necessarily translate into better out-of-sample predictive power. The objective of this subsection is to delve into the out-of-sample prediction performance of the models. We first split the data set into two subsets, using the first 80% of each population data as training set to fit the models, and then use the remaining 20% of data as test set to evaluate the predictive power. To perform the out-of-sample forecast, following the mortality forecasting literature (see, e.g., Li and Lee, 2005; Cairns et al., 2006), we perform the Augmented Dickey-Fuller (ADF) test (Said and Dickey, 1984) to test for stationarity of each estimated time-varying factor, $\boldsymbol{\kappa}_t$. Then we fit a random walk with drift process for non-stationary factors and an $AR(1)$ model for stationary factors. After generating out-of-sample forecast, we compare the forecasting performance for each population by calculating the following mean absolute percentage error (MAPE):

$$\text{MAPE} = \frac{100\%}{(x_{\max} - x_0 + 1) \times (t_{\max} - t_0 + 1)} \sum_{x=x_0}^{x_{\max}} \sum_{t=t_0}^{t=t_{\max}} \left| \frac{q_{x,t} - \hat{q}_{x,t}}{q_{x,t}} \right|. \tag{4.1}$$

The relative performance (in terms of MAPE) of the mortality models and over 22 populations is presented in Table 2. Apart from four populations (i.e., Denmark (DNK), Finland (FIN), New Zealand (NZL), and Sweden (SWE)), the remaining 18 populations are in favour of the exponential models as gauged by their lowest out-of-sample MAPE. Among the exponential models, 14 of them are optimal with EM2f while the remaining 4 prefer EM1f. Moreover, 6 of the exponential models produce no more than 6% MAPE, with the lowest MAPE being 3.43% for U.S. arising from EM2f. The superior out-of-sample forecast perfor-

Table 1: **In-sample fitness AIC and BIC of merged mortality data.** This table summarizes AIC and BIC of CBD-L, CBD-Q, EM1f, and EM2f models for uni-sex populations of the merged mortality data set. For each population, the best-performed model (i.e., lowest AIC/BIC) is highlighted in green.

AIC of different models

	CBD-L	CBD-Q	EM1f	EM2f		CBD-L	CBD-Q	EM1f	EM2f
AUS	31965	27145	23913	22695	ITA	138208	85672	49607	39309
AUT	111961	86650	74331	65017	JPN	312752	146459	87201	45036
BEL	52187	39745	33833	32404	LUX	13961	13068	12795	12701
CAN	63278	52838	48360	46588	NLD	45990	33899	28265	28179
DNK	48788	39915	36853	36665	NOR	73460	66308	61802	61724
DUETE	108633	48629	35375	32468	NZL	28967	25663	23857	24202
DUETW	199190	103786	67948	45925	PRT	111199	59576	53032	50531
FIN	76894	59697	52895	51492	ESP	963536	719978	614528	563600
FRA	186355	136617	85722	45327	SWE	146498	110375	81117	80829
ISL	12610	12559	12473	12501	CHE	91659	65272	58613	56370
IRL	40069	32674	32066	32075	USA	203247	195035	107626	56452

BIC of different models

	CBD-L	CBD-Q	EM1f	EM2f		CBD-L	CBD-Q	EM1f	EM2f
AUS	32473	27906	24685	23721	ITA	138823	86594	50541	40550
AUT	112713	87778	75471	66533	JPN	313449	147504	88259	46442
BEL	52967	40915	35015	33977	LUX	14429	13769	13507	13648
CAN	64359	54459	49993	48762	NLD	46687	34945	29322	29585
DNK	49898	41580	38530	38898	NOR	75717	69693	65201	66251
DUETE	109248	49551	36309	33709	NZL	29664	26709	24914	25607
DUETW	199805	104708	68882	47166	PRT	112035	60831	54299	52216
FIN	78650	62331	55543	55018	ESP	964837	721930	616493	566216
FRA	187108	137745	86862	46844	SWE	148534	113429	84185	84914
ISL	13171	13400	13325	13633	CHE	93447	67952	61307	59958
IRL	40766	33720	33123	33480	USA	203795	195855	108458	57558

mance, coupling with their in-sample goodness of fit efficiency, reaffirm the effectiveness of the exponential models.

Table 2: **Out-of-sample MAPE for uni-sex populations.** This table compares out-of-sample MAPE of CBD-L, CBD-Q, EM1f, and EM2f models for uni-sex populations of the merged mortality data set. For each population, the best-performed model is highlighted in green.

	CBD-L	CBD-Q	EM1f	EM2f		CBD-L	CBD-Q	EM1f	EM2f
AUS	5.19%	5.49%	4.34%	3.65%	ITA	7.88%	7.78%	7.12%	15.00%
AUT	15.05%	14.99%	14.65%	13.17%	JPN	5.94%	6.01%	6.43%	4.76%
BEL	8.10%	8.60%	9.22%	5.69%	LUX	12.72%	12.63%	12.63%	12.62%
CAN	7.52%	7.48%	7.35%	6.22%	NLD	9.88%	9.49%	8.71%	10.86%
DNK	8.14%	8.32%	8.58%	17.91%	NOR	13.76%	13.67%	12.40%	12.97%
DUETE	12.48%	12.64%	13.29%	11.89%	NZL	13.46%	13.49%	13.51%	16.69%
DUETW	7.39%	7.29%	6.85%	4.90%	PRT	10.47%	8.57%	7.50%	6.37%
FIN	21.43%	21.70%	34.73%	21.66%	ESP	14.78%	13.40%	10.82%	8.38%
FRA	8.49%	7.76%	8.02%	5.16%	SWE	17.78%	17.84%	19.17%	19.78%
ISL	18.32%	18.30%	18.32%	18.10%	CHE	15.40%	16.44%	16.22%	14.65%
IRL	15.30%	14.47%	12.05%	12.29%	USA	4.66%	4.48%	4.53%	3.43%

4.4 Long-term forecast

In this subsection, we continue to evaluate the four mortality models by comparing their long-term forecasting performance. We continue to rely on the U.S. female population as an illustration example. Figure 5 displays the fan charts of the confidence intervals of the remaining curtate life expectancy at age 90 produced by the four mortality models. Each plot displays the 5% - 95% intervals with 10% increment. We note that uncertainties produced by the four models appear to be quite similar. Figure 6 displays the projected remaining life expectancy at ages 70, 80, 90, and 100, based on CBD-L, CBD-Q, and the optimal exponential model (i.e., EM2f which has the lowest BIC and out-of-sample MAPE for the U.S. data). First, we observe that the CBD-L model produces the lowest projected remaining life expectancy for all ages, followed by the CBD-Q model. The exponential model generates the highest life expectancy for all ages. This observation is consistent with the findings that the exponential model could better capture the mortality curvature and old-age mortality plateau, while the CBD models tend to overestimate death probabilities, especially for the very old ages.

5 Robustness analysis

The empirical analysis in the previous sections utilizes 22 populations of the merged data set. However, due to limited sample size of the KTD, the most recent year of data is typically

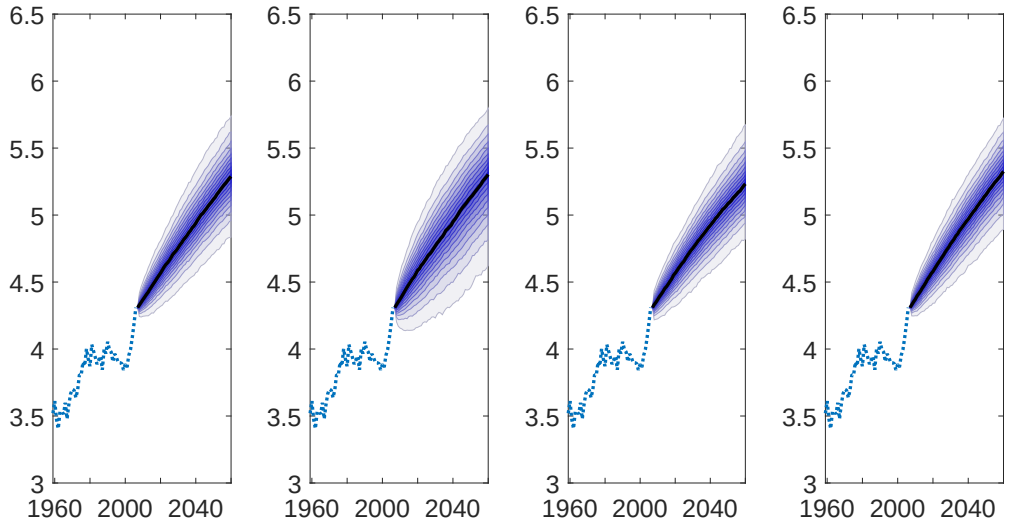
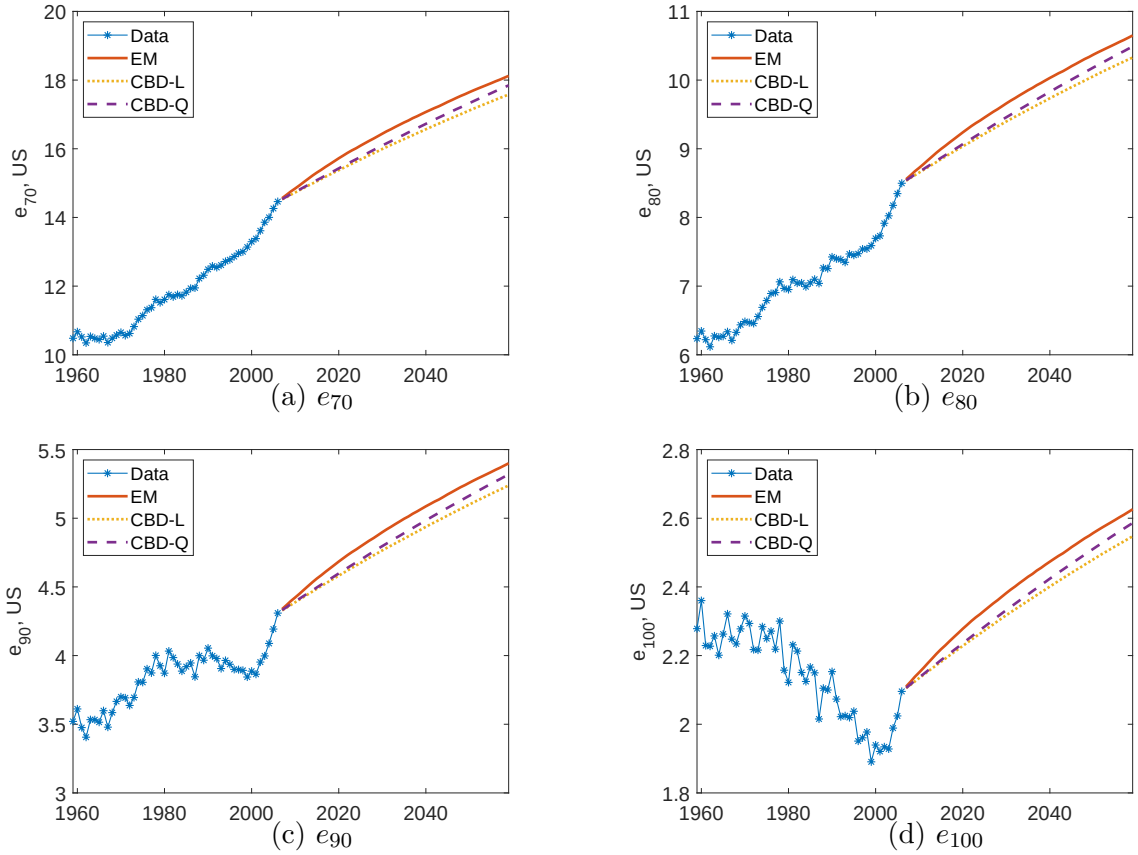


Figure 5: **Fan charts of the remaining life expectancy projections.** This figure displays the confidence intervals of the remaining life expectancy projections at age 90. From left to right: EM1f, EM2f, CBD-L, and CBD-Q. Models are fitted to the U.S. female populations in the merged data. x-axis is year.

2007 or 2008. To examine the robustness of the proposed model, in this section we perform the same in-sample and out-of-sample analysis but based on a larger number of populations and a longer calibration window. Specifically, we use mortality data from the HMD, which includes 41 countries and regions, most of which have the most recent year of data later than 2015. The advantages of the HMD data include its long periods and larger number of populations, although the data quality can be an issue, especially at old ages (say older than 95). Nevertheless, we rely on HMD to conduct robustness analysis. We use uni-sex mortality data between age 60 and age 95 to fit the models. For the starting year, we use 1950, or the earliest year available for each population, whichever is later. The abbreviations of the countries and regions included in the analysis are collected in Appendix A.

For the in-sample fitness, the AIC's and BIC's for each population are reported in Table 3. According to the AIC, the optimal models are EM2f for 40 out of 41 populations, and EM1f for the remaining one population. Even under the more stringent BIC criterion, there are still 30 out of 41 populations that choose EM2f as their optimal models and 10 populations remain with EM1f model. These results confirm the superior in-sample fitness of the exponential models.

Table 4 reproduces out-of-sample MAPE using the methodology as described in Subsection 4.3 but for the 41 populations extracted from the HMD. By comparing the MAPE, the exponential models still produce better out-of-sample prediction than the other two CBD models for 36 out of 41 populations, with the lowest MAPE being 2.84% (for Korea (KOR) under EM1f). These results confirm the out-of-sample prediction power of the exponential



(a) e_{70} (b) e_{80} (c) e_{90} (d) e_{100}

Figure 6: **Projected remaining life expectancy.** This figure shows the projected remaining life expectancy at 70, 80, 90, and 100 by the CBD-L, CBD-Q, and the optimal exponential model, fitted to the U.S. female populations in the merged data. x-axis is year.

Table 3: **In-sample fitness of different mortality models for HMD data.** This table reports AIC and BIC of CBD-L, CBD-Q, EM1f, and EM2f models for uni-sex populations of the HMD data. For each population, the best-performed model (i.e., lowest AIC/BIC) is highlighted in green.

AIC of different models

Population	CBD-L	CBD-Q	EM1f	EM2f	Population	CBD-L	CBD-Q	EM1f	EM2f
AUS	38569	30262	29951	29807	JPN	182151	65803	54922	42573
AUT	43907	29638	28425	26952	KOR	12809	10163	7906	7471
BLR	56387	44251	43360	38330	LVA	26950	20739	20520	20165
BEL	39427	28627	27390	26578	LTU	33628	24443	23827	22231
BGR	44968	34902	33195	29710	LUX	14927	14637	14545	14546
CAN	44412	30134	29958	29469	NLD	34330	26708	26245	25718
CHL	7468	7192	7155	7117	NZL	23552	22416	22351	22275
HRV	9155	7337	6705	6394	NOR	24616	22477	22285	22177
CZE	39245	29165	28474	27638	POL	82670	36680	34410	31248
DNK	27902	24787	24669	24302	PRT	42975	35417	34786	31483
EST	22152	19107	18958	18840	RUS	404973	199596	187339	134301
FIN	30260	23913	23483	23249	SVK	30472	26274	25695	25413
FRATNP	207724	53750	41840	32964	SVN	14642	12112	11838	11678
DEUTE	61545	38376	35242	32240	ESP	110412	62632	59662	50160
DEUTW	164882	64587	57730	49439	SWE	34411	27029	26654	25932
GRC	23275	20414	18160	15912	CHE	36653	25668	25152	24464
HUN	51487	33565	31660	30512	TWN	21366	19266	18848	17370
ISL	14693	14625	14596	14573	UK	60154	44477	42652	40652
IRL	33220	32322	32009	31618	USA	331829	76636	74388	67621
ISR	14356	13565	13483	13296	UKR	152315	98204	94016	74526
ITA	81916	45385	39248	33325					

BIC of different models

Population	CBD-L	CBD-Q	EM1f	EM2f	Population	CBD-L	CBD-Q	EM1f	EM2f
AUS	39345	31426	31126	31369	JPN	182940	66987	56117	44163
AUT	44696	30821	29621	28542	KOR	12927	10340	8092	7716
BLR	57041	45233	44353	39651	LVA	27618	21741	21534	21512
BEL	40189	29771	28544	28114	LTU	34296	25445	24840	23578
BGR	45663	35944	34248	31111	LUX	15542	15559	15479	15787
CAN	45188	31298	31133	31032	NLD	35105	27871	27420	27281
CHL	7618	7418	7389	7426	NZL	24288	23518	23465	23757
HRV	9295	7546	6923	6682	NOR	25365	23600	23420	23686
CZE	40034	30349	29670	29228	POL	83338	37682	35423	32596
DNK	28678	25950	25844	25865	PRT	43737	36560	35940	33018
EST	22820	20109	19972	20188	RUS	405601	200539	188292	135569
FIN	31022	25057	24638	24785	SVK	31262	27458	26890	27003
FRATNP	208500	54914	43015	34527	SVN	15002	12652	12388	12408
DEUTE	62253	39438	36316	33667	ESP	111188	63796	60837	51723
DEUTW	165590	65649	58803	50866	SWE	35200	28213	27850	27522
GRC	23610	20917	18673	16593	CHE	37429	26831	26327	26027
HUN	52276	34749	32856	32102	TWN	21851	19994	19586	18351
ISL	15468	15789	15771	16136	UK	60930	45641	43827	42214
IRL	33969	33445	33143	33127	USA	332619	77820	75584	69211
ISR	14704	14086	14015	14001	UKR	152930	99126	94950	75767
ITA	82665	46508	40383	34833					

models.

Table 4: **Out-of-sample MAPE of different mortality models for HMD data.** This table summarizes out-of-sample MAPE of CBD-L, CBD-Q, EM1f, and EM2f models for uni-sex populations of the HMD data. For each population, the best-performed model is highlighted in green.

Population	CBD-L	CBD-Q	EM1f	EM2f	Population	CBD-L	CBD-Q	EM1f	EM2f
AUS	9.77%	7.89%	7.87%	13.48%	JPN	7.41%	4.62%	4.49%	4.51%
AUT	12.28%	7.14%	7.32%	7.11%	KOR	6.01%	4.53%	2.84%	5.82%
BLR	20.08%	18.51%	18.57%	18.49%	LVA	17.99%	15.07%	15.04%	15.30%
BEL	9.58%	6.05%	5.68%	5.67%	LTU	15.02%	10.14%	10.07%	9.96%
BGR	18.90%	18.12%	18.23%	18.02%	LUX	21.67%	20.81%	20.97%	22.78%
CAN	9.41%	7.92%	7.90%	7.87%	NLD	15.10%	14.55%	14.55%	14.55%
CHL	4.58%	4.19%	4.14%	4.19%	NZL	10.77%	9.80%	9.75%	8.57%
HRV	8.84%	5.71%	5.42%	5.13%	NOR	11.96%	10.89%	10.89%	14.20%
CZE	15.66%	14.28%	14.27%	14.26%	POL	10.94%	6.70%	6.59%	6.50%
DNK	13.81%	13.41%	13.43%	13.44%	PRT	12.54%	11.52%	11.54%	15.98%
EST	19.38%	16.78%	16.82%	16.79%	RUS	25.24%	23.74%	23.77%	23.73%
FIN	11.08%	7.67%	7.52%	7.54%	SVK	13.02%	11.53%	15.81%	11.50%
FRA	13.59%	6.72%	6.58%	6.43%	SVN	10.46%	5.85%	5.46%	5.84%
DEUTE	9.65%	5.37%	5.27%	5.71%	ESP	10.78%	6.89%	6.78%	6.44%
DEUTW	9.29%	4.37%	4.09%	3.91%	SWE	8.49%	5.32%	5.30%	5.23%
GRC	9.10%	6.81%	5.84%	4.93%	CHE	11.02%	4.97%	4.92%	4.71%
HUN	12.72%	9.59%	9.52%	9.49%	TWN	7.44%	6.43%	6.49%	6.33%
ISL	14.81%	14.43%	14.65%	14.48%	UK	10.90%	10.54%	10.55%	10.53%
IRL	19.45%	19.61%	19.39%	19.31%	USA	10.20%	7.20%	7.20%	7.22%
ISR	5.69%	5.06%	4.91%	5.02%	UKR	13.24%	11.23%	11.38%	11.20%
ITA	8.87%	6.82%	6.86%	6.78%					

6 Multi-population extension

In this section, we extend the proposed exponential mortality model to multi-population framework. The simultaneous mortality modeling of multiple populations has become increasingly popular in recent academic literature (see, among many others, Cairns et al., 2011b; Dowd et al., 2011; Zhou et al., 2013; Li and Lu, 2017; Zhu et al., 2017; Li and Shi, 2020). In contrast to the single-population counterparts, multi-population mortality models, such as those based on Li and Lee (2005), Zhou et al. (2014), Li et al. (2019), and many others, have the desirable property that the long-term mortality projections among different populations are non-diverging. Moreover, there have been an abundant literature focused on the potential of using longevity-linked securities to mitigate the longevity risk exposure of pension plans and life insurers since early 2000s. Many of the securities discussed are based on projection of mortality rates of multiple populations (see, e.g., Blake et al., 2013; Li et al., 2017; Li, 2018; MacMinn and Richter, 2018). The development of such innovative

securities calls for more sophisticated multi-population mortality models that are able to provide better pricing and capture population basis risk (Blake et al., 2013).

The multi-population extension of the proposed exponential models is similar to Li and Lee (2005), who extend the single-population model by (Lee and Carter, 1992). Specifically, consider a group of J populations, we first calculate the cross-section (weighted) average of the population-specific logit $q_{j,x,t}$'s:

$$\overline{\text{logit } q_{x,t}} = \sum_{j=1}^{J} w_{x,t}^{(j)} \text{ logit } q_{x,t}^{(j)},$$

where $w_{x,t}^{(j)}$ is the weight for population j, age x, and year t. Then, we can apply the estimation Algorithm 1 introduced in Section 3 to estimate the common mortality factors. Finally, given the estimated common factors, population-specific residuals are calculated, to which the population-specific factors are estimated. The optimal order of common and population-specific factors can be determined by cross-validation or information criteria such as AIC or BIC. The multi-population extension introduced here can be quit flexible, in terms of the number of common and population-specific factors, and number of exponential terms included, etc. Trade-off between goodness-of-fit and model parsimony should be taken into consideration in model construction. Finally, coherence can be guarantee if all population-specific factors are stationary. As an illustration, we consider the following two multi-population model specifications:

- EM2f common factors with a linear population-specific factor:

$$\begin{aligned}\text{logit } q_{x,t}^{(j)} =& K_{1,t} + K_{2,t}(x-\bar{x}) + K_{3,t}\left(\frac{1-e^{-\lambda_1(x-\delta_{\boldsymbol{x}})}}{\lambda_1(x-\delta_{\boldsymbol{x}})} - \sigma_3\right) \\ &+ K_{4,t}\left(\frac{1-e^{-\lambda_2(x-\delta_{\boldsymbol{x}})}}{\lambda_2(x-\delta_{\boldsymbol{x}})} - e^{-\lambda_2(x-\delta_{\boldsymbol{x}})} - \sigma_4\right) + \kappa_{1,t}^{(j)} + \kappa_{2,t}^{(j)}(x-\bar{x}); \end{aligned} \tag{6.1}$$

- EM2f common factors with a linear and an exponential population-specific factors:

$$\begin{aligned}\text{logit } q_{x,t}^{(j)} =& K_{1,t} + K_{2,t}(x-\bar{x}) + K_{3,t}\left(\frac{1-e^{-\lambda_1(x-\delta_{\boldsymbol{x}})}}{\lambda_1(x-\delta_{\boldsymbol{x}})} - \sigma_3\right) \\ &+ K_{4,t}\left(\frac{1-e^{-\lambda_2(x-\delta_{\boldsymbol{x}})}}{\lambda_2(x-\delta_{\boldsymbol{x}})} - e^{-\lambda_2(x-\delta_{\boldsymbol{x}})} - \sigma_4\right)\kappa_{1,t}^{(j)} + \kappa_{2,t}^{(j)}(x-\bar{x}) \\ &+ \kappa_{3,t}^{(j)}\left(\frac{1-e^{-\lambda^{(j)}(x-\delta_{\boldsymbol{x}})}}{\lambda^{(j)}(x-\delta_{\boldsymbol{x}})} - e^{-\lambda^{(j)}(x-\delta_{\boldsymbol{x}})} - \sigma_3^{(j)}\right). \end{aligned} \tag{6.2}$$

The factors with superscript "(j)" are to emphasize that these are population specific factors. Similarly, $\sigma_{j,3}$ is the mean of $\frac{1-e^{-\lambda^{(j)}(x-\delta_{\boldsymbol{x}})}}{\lambda^{(j)}(x-\delta_{\boldsymbol{x}})} - e^{-\lambda^{(j)}(x-\delta_{\boldsymbol{x}})}$ over age. The above two multi-population exponential models are calibrated to the merged mortality data from the U.S. and

Japanese female populations. The estimated time-varying common factors and population specific factors are displayed in Figures 7 and 8 for models (6.1) and (6.2), respectively. We see that $K_{1,t}$ has a clear downward linear trend, representing the aggregate mortality improvement. $K_{2,t}$ has a pattern that first increases and then decreases, while $K_{3,t}$ and $K_{4,t}$ both decrease initially and then steeply increase after 1970. The population-specific factors estimated for the two populations are drastically different. For example, in Model (6.2), $\kappa_{1,t}^{(\mathrm{US})}$ has an increasing linear trend, while $\kappa_{1,t}^{(\mathrm{JP})}$ has a decreasing linear trend.

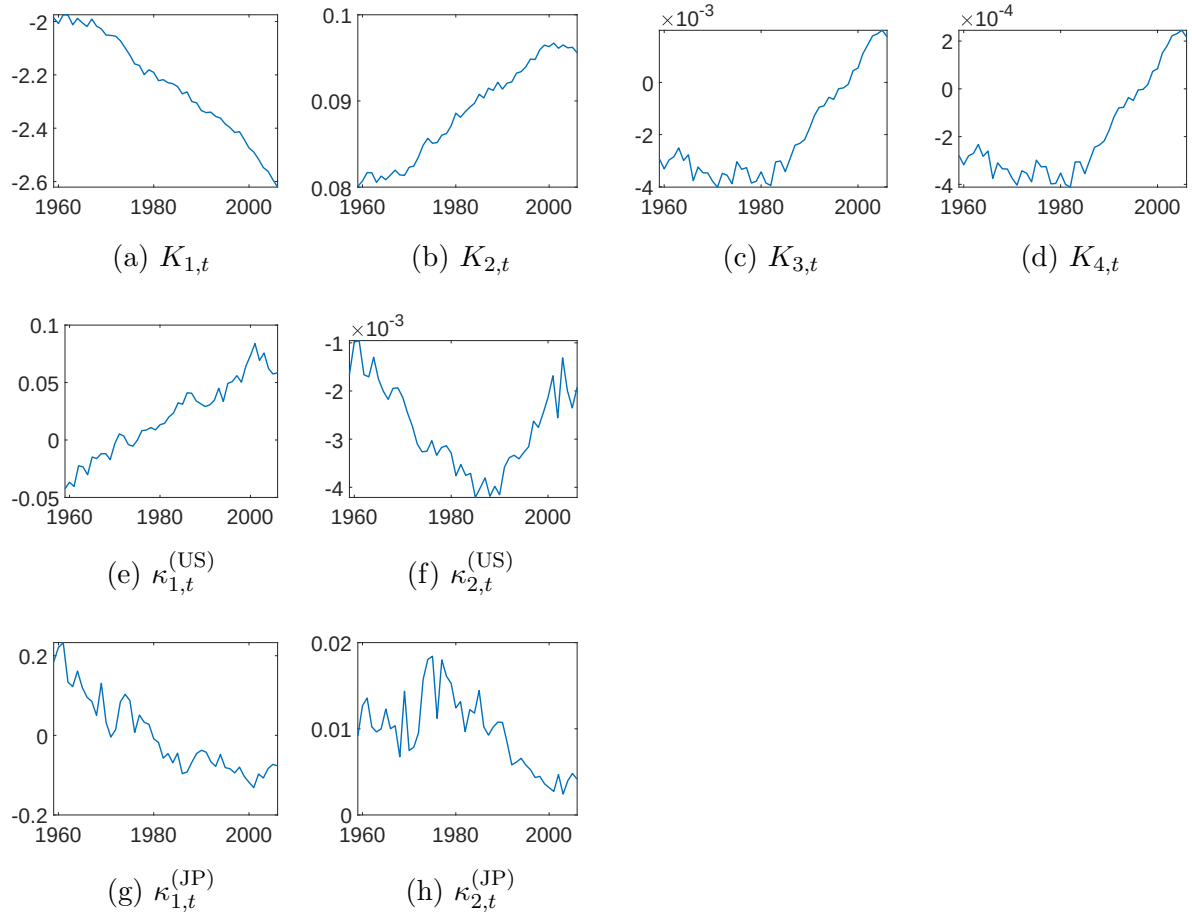


Figure 7: **Estimated time-varying factors for the two-population exponential model with one linear population-specific factor** (6.1)**.** This figure displays the estimated time-varying common factors, $\boldsymbol{K}_t = (K_{1,t}, K_{2,t}, K_{3,t}, K_{4,t})$, and population-specific factors, $\boldsymbol{\kappa}_t = (\kappa_{1,t}^{(j)}, \kappa_{2,t}^{(j)})$ for the two-population exponential model (6.1), fitted to the U.S. and Japanese female populations. x-axis is year.

The out-of-sample forecasting for the multi-population models is conducted as well. We perform the out-of-sample forecast for the merged data (female, male, and uni-sex) following the procedure described in Subsection 4.3. The results are summarized in Table 5. We see that except for the U.S. male population, the multi-population exponential models lead to better predictive accuracy. One possible reason that the multi-population exponential model

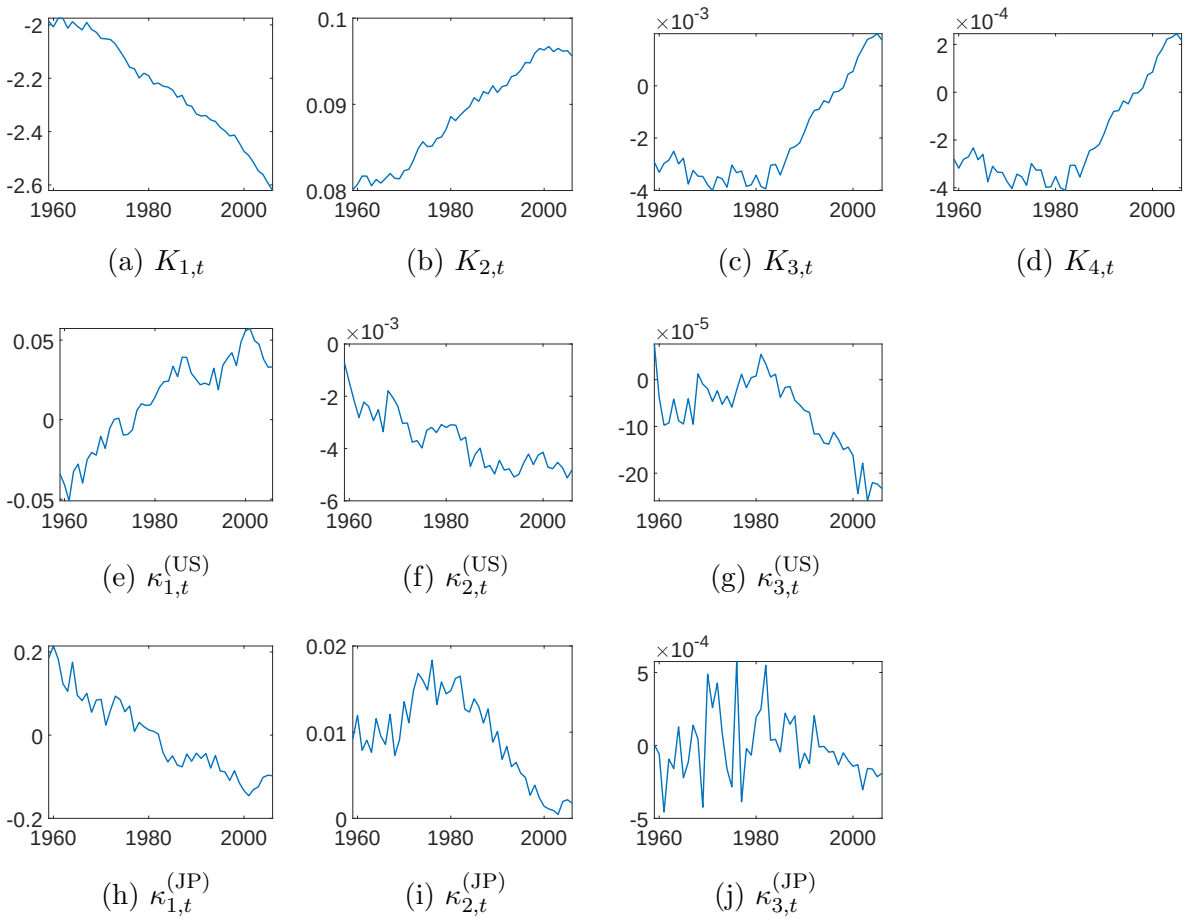


Figure 8: **Estimated time-varying factors for the two-population exponential model with one linear and one exponential population-specific factors** (6.2)**.** This figure displays the estimated time-varying common factors, $\boldsymbol{K}_t = (K_{1,t}, K_{2,t}, K_{3,t}, K_{4,t})$, and population-specific factors, $\boldsymbol{\kappa}_t = (\kappa_{1,t}^{(j)}, \kappa_{2,t}^{(j)}, \kappa_{3,t}^{(j)})$ for the two-population exponential model (6.2), fitted to the U.S. and Japanese female populations. x-axis is year.

delivers the best forecasting results is that Japan and the U.S. exhibit similar mortality patterns over sample size considered. As a result, combining the two populations together may reduce the impact of sampling errors, and hence improve model estimation. However, it is worth noting that the optimal model specification could depend on the populations considered, and thus the model selection procedure should be conducted when different mortality data are used.

Table 5: **Out-of-sample MAPE of two-population exponential models for merged data.** This table summarizes out-of-sample MAPE of CBD-L, CBD-Q, EM1f, EM2f, two-population EM model with one linear population-specific factor, and two-population EM model with one linear and one exponential population-specific factors fitted to the U.S. and Japanese populations in the merged mortality data set. The best-performed model is highlighted in green.

Population	Gender	CBD-L	CBD-Q	EM1f	EM2f	Two-Pop. Linear	Two-Pop. Liear+Exp
JP	Female	5.06%	5.36%	5.28%	4.40%	3.61%	3.47%
	Male	8.26%	8.52%	8.71%	5.61%	3.97%	6.45%
	Uni-sex	5.94%	6.01%	6.17%	4.76%	3.32%	3.83%
US	Female	6.45%	6.40%	5.87%	5.18%	5.20%	5.02%
	Male	5.22%	4.57%	5.17%	3.01%	3.02%	3.16%
	Uni-sex	4.66%	4.48%	4.50%	3.43%	3.01%	3.08%

7 Conclusion

An improved mortality model is an essential and fundamental issue for various areas, including demographic analysis and life insurance. In addition, it is important for governments to adapt retirement benefit policies to population aging. This paper proposes a multi-factor exponential model for mortality modeling, with a special focus on the mortality curvature and old age mortality plateau. In particular, the proposed model is able to generate a variety of mortality patterns, and is parsimonious, allowing for convenient estimation and forecasting algorithms. In the empirical study, we find that the proposed exponential model produces better in-sample goodness-of-fit, as well as out-of-sample forecasting performance, than existing Gompertz-based mortality models. Moreover, in long-term forecast, the proposed model projects higher remaining life expectancy, especially at the very old ages. We also illustrate a multi-population extension of the proposed exponential framework.

Since the focus of this research is on the curvature and possible mortality plateau in the mortality curve, the cohort effect is not considered. As a result, an interesting future research is to include the cohort effect in the model. In addition, it is interesting to investigate how the proposed model could be extrapolated to the extremely old ages beyond the highest age

in sample. In this case, a carefully selected appropriate ultimate mortality rate could be used as an input, and additional factors may be needed to ensure that the extrapolated mortality rates converge smoothly. Finally, exogenous factors, such as GDP and medical expenditure, could be used to improve the fitness and forecasting ability of the model (see, for example, Boonen and Li, 2017).

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.

Beard, R. E. (1959). Appendix: Note on some mathematical mortality models. In Wolstenholme, G. E. W. and O'Conner, M., editors, *Ciba Foundation Symposium - The Lifespan of Animals (Colloquia on Ageing), Volume 5*, pages 302–311. John Wiley & Sons, Ltd, Chichester, UK.

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

Boonen, T. J. and Li, H. (2017). Modeling and forecasting mortality with economic growth: a multipopulation approach. *Demography*, 54(5):1921–1946.

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

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

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

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

Courant, E. and Hilbert, D. (1953). *Methods of Mathematical Physics*. John Wiley, New York.

Currie, I. D. (2011). Modelling and forecasting the mortality of the very old. *ASTIN Bulletin: The Journal of the IAA*, 41(2):419–427.

Curtsinger, J. W., Gavrilova, N. S., and Gavrilov, L. A. (2006). Biodemography of aging and age-specific mortality in Drosophila melanogaster. In *Handbook of the Biology of Aging*, pages 261–288. Academic Press San Diego.

Desjardins, B. and Bourbeau, R. (2010). The emergence of supercentenarians in Canada. In *Supercentenarians*, pages 59–74. Springer-Verlag Berlin Heidelberg.

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

Gampe, J. (2010). Human mortality beyond age 110. In *Supercentenarians*, pages 219–230. Springer-Verlag Berlin Heidelberg.

Gavrilov, L. A. and Gavrilova, N. S. (1991). *The biology of life span: a quantitative approach*. New York New York/Chur Switzerland Harwood Academic 1991.

Gavrilov, L. A. and Gavrilova, N. S. (2011). Mortality measurement at advanced ages: a study of the social security administration death master file. *North American Actuarial Journal*, 15(3):432–447.

Gavrilov, L. A. and Gavrilova, N. S. (2019a). Late-life mortality is underestimated because of data errors. *PLoS biology*, 17(2):e3000148.

Gavrilov, L. A. and Gavrilova, N. S. (2019b). New trend in old-age mortality: Gompertzialization of mortality trajectory. *Gerontology*, 65:451–457.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical transactions of the Royal Society of London*, 115:513–583.

Greenwood, M. and Irwin, J. O. (1939). The biostatistics of senility. *Human Biology*, 11(1):1–23.

Haldrup, N. and Rosenskjold, C. (2019). A parametric factor model of the term structure of mortality. *Econometrics*, 7(1):9.

Horiuchi, S. and Wilmoth, J. R. (1998). Deceleration in the age pattern of mortality at older ages. *Demography*, 35(4):391–412.

Kannisto, V. (1994). *Development of oldest-old mortality, 1950-1990: Evidence from 28 developed countries*. University Press of Southern Denmark.

Kestenbaum, B. and Ferguson, B. R. (2010). Supercentenarians in the United States. In *Supercentenarians*, pages 43–58. Springer-Verlag Berlin Heidelberg.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American statistical association*, 87(419):659–671.

Li, H. (2018). Dynamic hedging of longevity risk: The effect of trading frequency. *ASTIN Bulletin: The Journal of the IAA*, 48(1):197–232.

Li, H., De Waegenaere, A., and Melenberg, B. (2017). Robust mean–variance hedging of longevity risk. *Journal of Risk and Insurance*, 84(S1):459–475.

Li, H., Li, H., Lu, Y., and Panagiotelis, A. (2019). A forecast reconciliation approach to cause-of-death mortality modeling. *Insurance: Mathematics and Economics*, 86:122–133.

Li, H. and Lu, Y. (2017). Coherent forecasting of mortality rates: A sparse vector-autoregression approach. *ASTIN Bulletin: The Journal of the IAA*, 47(2):563–600.

Li, H. and Shi, Y. (2020). Forecasting mortality with international linkages: A global vector-autoregression approach. *Available at SSRN 3700586*.

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.

MacMinn, R. and Richter, A. (2018). The choice of trigger in an insurance linked security: The mortality risk case. *Insurance: Mathematics and Economics*, 78:174–182.

National Research Council (2000). *Beyond Six Billion: Forecasting the World's Population*. National Academies Press.

Newman, S. J. (2018a). Errors as a primary cause of late-life mortality deceleration and plateaus. *PLoS biology*, 16(12).

Newman, S. J. (2018b). Plane inclinations: A critique of hypothesis and model choice in Barbi et al. *PLoS biology*, 16(12).

Olshansky, S. J. (1998). On the biodemography of aging: a review essay. *Population and Development Review*, 24(2):381–393.

Pitacco, E., Denuit, M., Haberman, S., and Olivieri, A. (2009). *Modelling longevity dynamics for pensions and annuity business*. Oxford University Press.

Rau, R., Soroko, E., Jasilionis, D., and Vaupel, J. W. (2008). Continued reductions in mortality at advanced ages. *Population and Development Review*, 34(4):747–768.

Riley, J. C. (2001). *Rising life expectancy: a global history*. Cambridge University Press.

Sacher, G. A. (1968). The Gompertz transformation in the study of the injury-mortality relationship: Application to late radiation effects and ageing. In Lindop, P. J. and Sacher, G. A., editors, *Radiation and Aging*, pages 411–441, London. Taylor & Francis.

Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607.

Shaw, C. (2007). Fifty years of United Kingdom national population projections: how accurate have they been? *Population trends*, 128:8–23.

Thatcher, A. R. (1999a). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(1):5–43.

Thatcher, A. R., Kannisto, V., and Vaupel, J. W. (1998). *The force of mortality at ages 80 to 120: Monographs on Population Aging Series v5*. University Press of Southern Denmark.

Thatcher, R. (1999b). The demography of centenarians in England and Wales. *Population trends*, 96:5–12.

Vaupel, J. W. (1997). The remarkable improvements in survival at older ages. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1363):1799–1804.

Vaupel, J. W., Carey, J. R., Christensen, K., Johnson, T. E., Yashin, A. I., Holm, N. V., Iachine, I. A., Kannisto, V., Khazaeli, A. A., Liedo, P., et al. (1998). Biodemographic trajectories of longevity. *Science*, 280(5365):855–860.

World Bank (2019). Life expectancy at birth, total (years). https://data.worldbank.org/indicator/SP.DYN.LE00.IN.

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

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

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

A Appendix: Abbreviations of the populations used in this paper

Table A.1: **The countries/regions in HMD and their abbreviations.**

Population	Abbrev.	Population	Abbrev.	Population	Abbrev.	Population	Abbrev.
Australia	AUS	Finland	FIN	Korea	KOR	Slovenia	SVN
Austria	AUT	France	FRA	Latvia	LVA	Spain	ESP
Belarus	BLR	East Germany	DEUTE	Lithuania	LTU	Sweden	SWE
Belgium	BEL	West Germany	DEUTW	Luxembourg	LUX	Switzerland	CHE
Bulgaria	BGR	Greece	GRC	Netherlands	NLD	Taiwan	TWN
Canada	CAN	Hungary	HUN	New Zealand	NZL	UK	UK
Chile	CHL	Iceland	ISL	Norway	NOR	USA	USA
Croatia	HRV	Ireland	IRL	Poland	POL	Ukraine	UKR
Czechia	CZE	Israel	ISR	Portugal	PRT		
Denmark	DNK	Italy	ITA	Russia	RUS		
Estonia	EST	Japan	JPN	Slovakia	SVK		