Modeling the Risk in Mortality Projections

Nan Zhu and Daniel Bauer*

Abstract

This paper presents and applies models for the valuation and management of mortality-contingent exposures. Such exposures include insurance and pension benefits as well as novel mortality-linked securities traded in financial markets. Unlike conventional approaches to modeling mortality, we consider the stochastic evolution of *mortality projections* rather than realized *mortality rates*. Relying on a time series of age-specific mortality forecasts, we develop a set of stochastic models that—unlike conventional mortality models—capture the evolution of mortality forecasts over the past fifty years. In particular, the dynamics of our models reflect the substantial observed variability of long-term projections, and are therefore particularly well-suited for financial applications where long-term demographic uncertainty is relevant.

JEL classification: C14; C53; J11

Keywords: Longevity Risk, Mortality Risk Management, Forward Mortality Models, Affine Models.

*Nan Zhu is at the Department of Risk Management, Smeal College of Business, Pennsylvania State University, University Park, PA 16802. Zhu can be contacted via email: nanzhu@psu.edu. Daniel Bauer (corresponding author) is at the Department of Risk and Insurance and the Center for the Demography of Health and Aging, University of Wisconsin-Madison, Madison, WI 53706. Bauer can be contacted via email: daniel.bauer@wisc.edu.

1   Introduction

Aggregate mortality risk refers to the risk that the future evolution of aggregate mortality rates deviates from current projections, e.g. due to medical advances or unforeseen events such as the current Coronavirus pandemic affecting human longevity. Unlike individual mortality risk—i.e., the risk of dying young or old—aggregate mortality risk cannot be diversified by pooling across individuals. Hence, it presents a substantial exposure for insurance companies, pension funds, and retirement systems. One possibility to manage this risk is via so-called *longevity derivatives* that tap diversification possibilities of financial markets.

   This paper presents, estimates, and applies novel models for measuring, pricing, and managing aggregate mortality risk. While many stochastic mortality forecasting models have been presented and in principle any of them can be used, our models exhibit a number of key advantages in these contexts. First, we present *forward mortality* or *mortality surface models*, in analogy to forward interest or yield curve models from financial engineering. These models take the current age/term-*surface* of mortality as a (non-parametric) input and describe its stochastic evolution going forward. While a number of papers in the mathematical finance literature have discussed frameworks for life-contingent forward rate modeling emphasizing advantages in this arena (Bauer et al., 2012; Tappe and Weber, 2014; Christiansen and Niemeyer, 2015; Buchardt et al., 2019), we are the first to present an *empirically derived mortality surface model*.

   The forward modeling perspective is not only convenient for certain applications, but it also leads to fundamentally different *dynamics* that are more apt in financial applications. More specifically, existing papers on forecasting mortality do not focus attention on the uncertainty associated with the resulting *mortality projections*. Rather, the ubiquitous approach following the seminal work by Lee and Carter (1992) relies on past *mortality rates* to furnish an estimate of the mortality experience for subsequent years that matches past observations in some optimal sense. And while resulting models present excellent tools for applications where the uncertainty in mortality rates is central, they are not built for depicting the risk situation in mortality projections that is more relevant for measuring and managing long-term mortality exposures in pension and insurance.

   To illustrate, Figure 1 plots cohort life expectancies for 70-year old US females—which can also be interpreted as the present values of an immediate life annuity under a zero interest rate— as forecast via the Lee and Carter (1992) model until 2016 (blue solid line).[1] In addition, the figure shows mean life expectancy forecasts from 2017 until 2036 (blue solid line) as well as corresponding point-wise 95% prediction intervals (red dashed line) based on Lee-Carter and years 1987-2016 for the estimation. We observe significant variation in the time series but the prediction

---

[1]Here and in what follows, we curtail the age range because of data quality in very high ages. Hence, "life expectancy" technically should be interpreted as expected number of years lived until age 95. Generalizations are possible (Thatcher et al., 2002, e.g.) but not in our focus.

(a) Mortality Surface Model
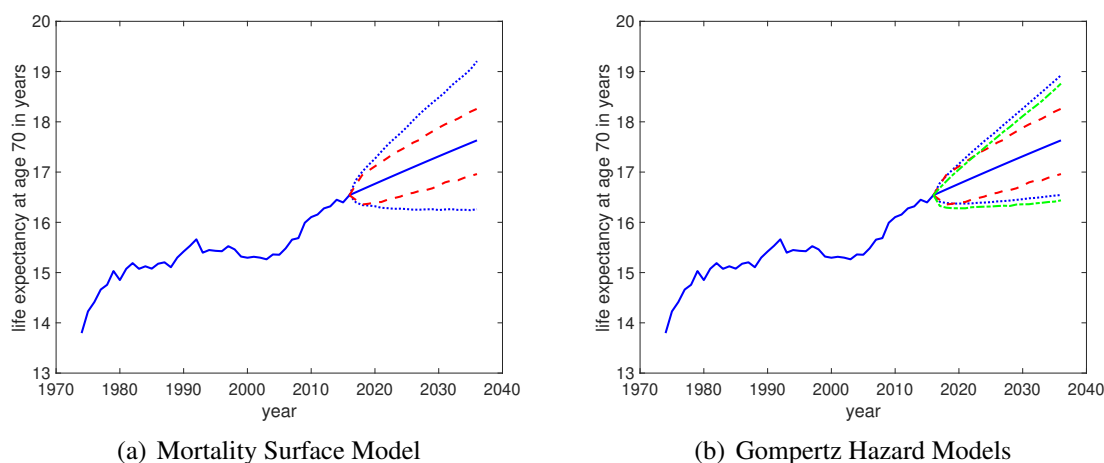
(b) Gompertz Hazard Models

Figure 1: Showing cohort life expectancies for 70-year old U.S. females based on the Lee-Carter model until 2016 and corresponding future twenty-year (2017-2036) forecasts (blue solid), including 95% prediction intervals based on Lee-Carter (red dashed) and mortality models with long-term risk. (a): Mortality surface model (blue dotted). (b): Gaussian (blue dotted) and Gamma (green dotdashed) hazard models.

intervals are rather narrow—too narrow from a historical perspective. For instance, the historical path veers outside of past prediction bands in backtests, and univariate time series forecasts of life expectancies break the intervals far more often than predicted (see Online Appendix E for details). This observation echoes findings in the literature. For instance, Cocco and Gomes (2012) assume a higher dispersion than suggested by Lee-Carter estimates, pointing to statements by the Government Actuary's Department (2001) "that uncertainty over long-term mortality levels is higher" than suggested by statistical techniques. In fact, even in their original work, Lee and Carter (1992) pose the question of why "the confidence bands on life expectancy [based on their procedure] are so narrow"?

As an alternative approach, in this paper, we base our analysis on a time series of age-specific mortality forecasts as they are tabulated within cohort (or generational) life tables that are common in business and governmental operations, such as pricing life annuities in the insurance industry or estimating future cash flows in the *U.S. Old-Age, Survivors, and Disability Insurance* (OASDI) program (Bell and Miller, 2005). We note that resulting models are theoretically equivalent: A stochastic model for mortality experience in every period implies a stochastic mortality forecast and vice versa. In other words, one can reformulate a spot model into a forward model and vice versa. However, the distinction is relevant for the statistical approach and, therefore, is important for the specification and estimation process. In particular, including the current forecast in the cross-sectional view is important to identify the persistence and transiency of random shocks.

We develop our models by analyzing different time series of mortality forecasts in the form of

expected survival probabilities as objects on some space of functions in two variables (current age and forecasting term). We find that one factor explains the vast majority of the variation in the data, and the shape as a function of term and age is highly systematic and very similar across different populations and data sources. While statistical (factor) models are easily derived, we demonstrate that a five-parameter affine forward mortality model—three parameters for capturing the age pattern and two for capturing the dynamic evolution—is able to depict the relevant relationships and results in very similar predictions as the (non- or high-parametric) statistical models. In addition, the surface model is *self-consistent*, i.e. the expected value when successively forecasting single year mortality will line up with the expected value ingrained in current long-term forecasts—which is analogous to absence of arbitrage in forward interest rate models. Importantly, in contrast to conventional approaches, the model captures the empirical variability in mortality projections. To illustrate, in Panel (a) of Figure 1 we include prediction intervals for future life expectancies based on our mortality surface model (blue dotted line), which are are substantially wider.

The intuition for this observation becomes clear when reformulating our forward model as a spot (hazard) model. The spot dynamics—in contrast to the surface dynamics—are non-Markovian and depend on *two* state variables that are driven by the same source of randomness. One of the state variables captures the contemporaneous impact of a random shock, similarly as in conventional mortality models. However, the second state variable puts a persistent effect on the evolution of mortality rates, and thus generates substantial variability in the long run. Both aspects are relevant to modeling mortality risk in financial applications. We propose fully parametric (Gaussian and non-negative Gamma) hazard models as stochastic extensions to the Gompertz mortality law that capture both aspects and perform similarly to our forward model for the evolution of the full age/term-surface of mortality—but are easy-to-use, parsimonious (six/seven parameters in total), transparent, and tractable. Panel (b) of Figure 1 depicts resulting prediction intervals.

We showcase these models with a focus on their difference in performance to conventional models in two financial applications: Pricing a so-called *guaranteed annuity option* in insurance savings contracts; and a (mortality) call spread option, which we present in context with some background on the longevity risk transfer market. Results from both applications reiterate the key point illustrated in Figure 1, namely that conventional models fail to adequately describe the risk situation. We discuss potential consequences of this observation for related financial markets.

Related Literature and Organization of the Paper

This paper relates to several literatures. We present models that are suited for the valuation and risk management of life-contingent and mortality derivative securities. Broadie and Detemple (2004) highlight the contributions of the OR/MS literature to the study of derivative securities,

although the focus has been on options for financial indices. In particular, a number of affine pricing models have been presented due to their analytical tractability (Kou, 2002; Broadie and Kaya, 2006; Christoffersen et al., 2009; Kang et al., 2017, e.g.). We contribute by proposing novel, empirically motivated, tractable (affine) models suited for the analysis of these alternate contracts.

The statistical, demographic, and insurance literatures have produced a variety of statistical mortality forecasting models, such as the popular Lee and Carter (1992) and its extensions (Wilmoth, 1993; Booth et al., 2002; Cairns et al., 2006b, e.g.). In contrast to these approaches and as a major innovation, we do not rely on a time series of age-specific mortality rates as a starting point, but the basis for our approach is a time series of age-specific mortality forecasts. This allows us to identify the persistence and transiency of random shocks affecting the age/term structure of mortality. This point is widely recognized for interest rate models: Models for short rate give models for the entire term structure of interest rates; yet any meaningful empirical approach for forecasting the yield curve requires the consideration of all cross-sectional data and not only observations on the short end, and one usually distinguishes *level*, *slope*, and *curvature* factors describing the shocks' impact across the term structure (Diebold and Li, 2006; Piazzesi, 2010, e.g.). We provide parallel insights for the age/term structure of mortality.

A number of contributions in the actuarial literature discuss pricing and managing mortality-contingent claims (Milevsky and Promislow, 2001; Dahl and Møller, 2006; Cairns et al., 2006a, e.g.), yet they rely on simple models that are not able to capture relevant aspects of long-term mortality risk in financial applications. We focus on specifying and estimating suitable models by an extensive analysis of suitably compiled mortality data for these applications. We draw on approaches and models from the fixed income literature (Hull and White, 1990; Litterman and Scheinkman, 1991; Björk and Gombani, 1999; Duffee, 2011, among others), and we compare and contrast our results to those produced by alternate popular mortality models.

Section 2 discusses the modeling of mortality forecasts. Here we draw on the forward mortality modeling frameworks from prior literature (Bauer et al., 2012; Tappe and Weber, 2014) but present some new results that are conducive to deriving our models. Section 3 uses a variety of different data sources to specify and estimate a parametric forward mortality model. Section 4 discusses the resulting model dynamics and presents a set of simple, transparent, and tractable hazard rate models with long-term risk. Sections 5 and 6 apply our models for pricing insurance options and derivatives for longevity risk transfer, respectively. The Online Appendix collects proofs, other technical material, model extensions, supporting analyses, and rich robustness analyses for various datasets.

2   Modeling Mortality Forecasts

The typical starting point in mortality modeling are *realized* "annual age-specific death rates" (Lee and Carter, 1992, p. 659) leading to *models for mortality experience.* In contrast, we base our analysis on a time series of age-specific mortality forecasts as expectations to survive for a certain period of time. More specifically, we assume that at time t, we are given $\{_\tau p_x(t)|\,(\tau, x) \in \mathcal{C}\}$, where $_\tau p_x(t)$ denotes the probability for an x-year old to survive for τ periods until time $t + \tau$ based on the information at time t, and $\mathcal{C}$ denotes a (large) collection of term/age combinations.[2] Our goal is to propose dynamic stochastic models for *mortality forecasts* $(\{_\tau p_x(t)|\,(\tau, x) \in \mathcal{C}\})_{t \geq 0}$.

We note that once the forecast at time t is available, various quantities follow immediately. For instance, for the expected future lifetime (cohort life expectancy, $\overset{\circ}{e}_x(t)$) and the expected discounted payoff of an annual life annuity of \$1 ($a_x(t)$) for an x-year old at time t, we have:

$$\overset{\circ}{e}_x(t) = \int_0^\infty {}_\tau p_x(t)\, d\tau \text{ and } a_x(t) = \sum_{k=1}^\infty B(t, k)\, {}_k p_x(t),$$

where $B(t, \tau)$ is the time-t price of a zero-coupon bond maturing at $t + \tau$ (τ-year discount factor).

A (continuous-time) model for $(\{_\tau p_x(t)|\,(\tau, x) \in \mathcal{C}\})_{t \geq 0}$ can be formulated via a stochastic (differential) equation on a suitable function space. However, since the specific monotonicity and boundedness requirements of the forecasted probabilities lead to complications in their modeling, assuming sufficient regularity, it is easier to work with the transformed objects:

$$\mu_t(\tau, x) = -\frac{\partial}{\partial \tau} \log \{_\tau p_x(t)\} \text{ with } {}_\tau p_x(t) = \exp\left\{-\int_0^\tau \mu_t(s, x)\, ds\right\}. \tag{1}$$

Here, $\mu_t(\tau, x)$ is called the *forward force of mortality* (Cairns et al., 2006a), and we interpret $\mu_t(\cdot, \cdot)$ as an element of some space $\mathcal{H}$ of continuous functions (we refer to Bauer et al. (2012) and Tappe and Weber (2014) for details on these models).

2.1   Gaussian Forward Models

We commence by considering time-homogeneous, Gaussian models of the type:

$$d\mu_t = (A\,\mu_t + \alpha)\, dt + \boldsymbol{\sigma}\, dW_t, \tag{2}$$

where (W_t) is a D-dimensional Brownian motion, $\alpha \in \mathcal{H}$, $\boldsymbol{\sigma} \in L(\mathbb{R}^D, \mathcal{H})$ (linear mappings from $\mathbb{R}^D$ to $\mathcal{H}$), and A is the infinitesimal generator of a strongly continuous semigroup (S_t) that

---

[2]Our notation is based on *International Actuarial Notation*, where 'p' generally denotes survival probabilities, but we extend it to introduce time dependence via $\cdot(t)$.

coincides with the translation semigroup of left shifts in the first and right shifts in the second variable:

$$(S_t\, f)\,(\tau, x) = f(\tau + t, x - t),\ 0 \le t \le x.$$

To provide intuition for the latter, for a particular future rate, age decreases and forecasting term increases as we move *backward* in time t. We obtain $A = \partial/\partial\tau - \partial/\partial x$ on the domain of A, dom(A).

While the focus on Gaussian models limits generality and is associated with the possibility of survival probabilities exceeding unity, these shortcomings are countervailed by statistical tractability that we exploit below. We present non-negative variants of our models in Section 4. In addition, Equation (2) entails that μ_t is stationary as well as Markovian. However, we note that this is not as restrictive as it may seem, since the stationarity/Markov properties are on the forward force of mortality (the object $\mu_t(\cdot, \cdot)$) instead of the age-specific *spot* force of mortality (hazard) $\mu_t(0, x)$. In particular, we can—and will—include time trends within the initial surface. And while the assumption of a Markovian spot force model leads to highly restrictive conditions on the forward force model (e.g., Musiela and Rutkowski, 1998, Prop. 13.3.2), the Markovian assumption directly on the surface is much less restrictive. In particular, the equivalent spot force model corresponding to Equation (2) generally is neither stationary nor Markovian (see Proposition 2.2 in Bauer et al. (2012))—and, indeed, the models we propose below are neither Markovian nor time-homogeneous when formulated as a model for the hazard (see Proposition 4.1 below).

Assume cohort mortality data $_\tau p_x(t_j)$ is provided for different dates t_j, $j = 0, 1, \ldots, N$. In addition, denote by l the lag time, and choose a sub-collection $\tilde{\mathcal{C}} \subset \mathcal{C}$, $|\tilde{\mathcal{C}}| = K$, such that for $(\tau, x) \in \tilde{\mathcal{C}}$, $(\tau + l, x)$, $(\tau + t_{j+1} - t_j, x - t_{j+1} + t_j)$, $(\tau + l + t_{j+1} - t_j, x - t_{j+1} + t_j) \in \mathcal{C}$, $\forall j \in \{0, 1, \ldots, N - 1\}$. For each $(\tau, x) \in \tilde{\mathcal{C}}$, $j \in \{0, 1, \ldots, N - 1\}$, define:

$$F_l\,(t_j, t_{j+1}, (\tau, x)) = -\log\left\{ \frac{_{\tau+l}p_x(t_{j+1})}{_\tau p_x(t_{j+1})} \bigg/ \frac{_{\tau+l+t_{j+1}-t_j}p_{x-t_{j+1}+t_j}(t_j)}{_{\tau+t_{j+1}-t_j}p_{x-t_{j+1}+t_j}(t_j)} \right\}. \tag{3}$$

Conceptually, $F_l\,(t_j, t_{j+1}, (\tau, x))$ measures the log change of the l-year *marginal* survival probability for the period $[t_{j+1} + \tau, t_{j+1} + \tau + l)$ from projection at time t_{j+1} relative to time t_j, for an—at time t_{j+1}—x-year old individual. We obtain:

Proposition 2.1. *Under the dynamics* (2) *the vectors:*

$$\boldsymbol{F}_l(t_j, t_{j+1}) \;=\; \left(\frac{F_l(t_j, t_{j+1}, (\tau, x))}{\sqrt{t_{j+1} - t_j}} \right)_{(\tau, x) \in \tilde{\mathcal{C}}} = \left(\frac{F_l(t_j, t_{j+1}, (\tau_1, x_1))}{\sqrt{t_{j+1} - t_j}}, \ldots, \frac{F_l(t_j, t_{j+1}, (\tau_K, x_K))}{\sqrt{t_{j+1} - t_j}} \right),$$

$j = 0, 1, \ldots, N - 1$, *are independent and Gaussian distributed.*[3] *If the data is equidistant with*

---

[3]By scaling the data points by $1/\sqrt{t_{j+1} - t_j}$, we ascertain that the vectors $\boldsymbol{F}_l(t_j, t_{j+1})$, $j = 0, 1, \ldots, N - 1$, are

$t_{j+1} - t_j = \Delta$, $\boldsymbol{F}_l(t_j, t_{j+1})$ *are also identically distributed.*

In what follows, for simplicity of exposition and without loss of generality, unless stated otherwise we assume equidistant annual data ($t_{j+1} - t_j \equiv 1, j \geq 0$) and the lag time $l = 1$. We write:

$$\boldsymbol{F}_1(t_j, t_j + 1) = \boldsymbol{F}^{j+1} = \left(F^{j+1}(\tau, x)\right)_{(\tau, x) \in \tilde{\mathcal{C}}}.$$

Modifications in the examples where these assumptions do not hold are straightforward but notationally tedious. With equidistant data, the i.i.d. property can now be immediately exploited in devising non-parametric and factor models for mortality forecasting. For instance, to devise non-parametric forecasts we can sample with replacement (non-parametric bootstrap) from the (i.i.d.) vectors $\{\boldsymbol{F}^1, \boldsymbol{F}^2, \ldots, \boldsymbol{F}^N\}$ and use Equation (3) to forecast $\{_\tau p_x(T+1)\}$ based on $\{_\tau p_x(T)\}$, $\{_\tau p_x(T+2)\}$ based on $\{_\tau p_x(T+1)\}$, and so on. We refer to Online Appendix B for details.

However, for similar ages and terms, of course the components of the Gaussian vectors $\boldsymbol{F}^j$, $j = 1, 2, \ldots, N$, are highly correlated (which motivates the assumption of a finite-dimensional process in the dynamics (2)). Hence, in analogy to basic approaches for forecasting yield curves (Litterman and Scheinkman, 1991; Diebold and Li, 2006, e.g.) or period age-dependent mortality rates (Lee and Carter, 1992, e.g.), we can obtain a more parsimonious and tractable class of models from conventional factor analysis.

The procedure is standard: We decompose the empirical covariance matrix of $\{\boldsymbol{F}^1, \boldsymbol{F}^2, \ldots, \boldsymbol{F}^N\}$ as $\hat{\boldsymbol{\Sigma}} = \sum_{\nu=1}^{K} \lambda_\nu \boldsymbol{u}_\nu \boldsymbol{u}_\nu'$, where $\lambda_\nu, \nu = 1, 2, \ldots, K$, are the eigenvalues of $\hat{\boldsymbol{\Sigma}}$ in decreasing order, and $\boldsymbol{u}_\nu, \nu = 1, 2, \ldots, K$, are the corresponding eigenvectors. We then pick the D greatest eigenvalues that explain the majority of the variation in the data. Therefore, the resulting approximate covariance matrix is $\sum_{\nu=1}^{D} \lambda_\nu \boldsymbol{u}_\nu \boldsymbol{u}_\nu' = \text{Cov}\left(\sum_{\nu=1}^{D} \boldsymbol{u}_\nu \sqrt{\lambda_\nu} \varepsilon_\nu\right)$, where ε_ν are i.i.d. (scalar) standard Normal random variables, $\nu \in \{1, \ldots, D\}$. Isolating the first D eigenvalues suggests the following *factor model*:

$$\boldsymbol{F} = \bar{\boldsymbol{F}}^\cdot + \sum_{\nu=1}^{D} \boldsymbol{u}_\nu \sqrt{\lambda_\nu} \varepsilon_\nu, \tag{4}$$

so that mortality forecasts can be generated as described above by sampling D univariate Normal random variables rather than bootstrapping $\{\boldsymbol{F}^1, \ldots, \boldsymbol{F}^N\}$ (see also Online Appendix B).

2.2 Mortality Surface Models

While non-parametric and factor forecasting models for mortality projections are simple to implement, the expected value of forecasts does not necessarily align with the projection ingrained in the current cross-section. So-called *mortality surface models* enforce this *self-consistency* relationship

---

also approximately i.i.d. when using non-equidistant data.

(Bauer et al., 2012). To elaborate, *self-consistency* of a dynamic model for mortality forecasts takes the form of a martingale property—expected realized mortality rates should align with the given forecasts ($\mathbb{E}_t$ denotes the expectation operator based on information up to time t):

$$\mathbb{E}_t\left[\exp\left\{-\int_0^T \mu_s(0, x_0 + s)\, ds\right\}\right] = \exp\left\{-\int_0^t \mu_s(0, x_0 + s)\, ds\right\}{}_{T-t}p_{x_0+t}(t), \qquad (5)$$

that is, $(\exp\{-\int_0^t \mu_s(0, x_0 + s)\, ds\}{}_{T-t}p_{x_0+t}(t))_{t\geq 0}$ are martingales. This yields a condition akin to the well-known HJM drift condition for models of the yield curve (see Equation (3.5) in Cairns et al. (2006a), Corollary 3.1 in Bauer et al. (2012), or Theorem 5.3 in Tappe and Weber (2014)):

$$\alpha(\tau, x) = \boldsymbol{\sigma}(\tau, x) \times \int_0^\tau \boldsymbol{\sigma}'(s, x)\, ds. \qquad (6)$$

Similarly to Equation (4) above, we are interested in factor models $\mu_t(\tau, x) = G(\tau, x; Z_t)$, where G is a known deterministic function, and $(Z_t)_{t\geq 0}$ is a finite-dimensional stochastic process. Proposition 4.1 in Bauer et al. (2012) shows that for the time-homogeneous Gaussian models we consider here, in order for such a factor model to exist, the volatility structure must necessarily be of the form:

$$\boldsymbol{\sigma}(\tau, x) = \boldsymbol{C}(x + \tau) \times \exp\{\boldsymbol{H}\,\tau\} \times \boldsymbol{J}, \qquad (7)$$

where $\boldsymbol{J} \in \mathbb{R}^{m\times D}$, $\boldsymbol{H} \in \mathbb{R}^{m\times m}$, and $\boldsymbol{C}' \in C^1\left([0, \infty), \mathbb{R}^m\right)$; the *factor model* is then given by:

$$
\begin{aligned}
\mu_t(\tau, x) \;=\; & \mu_0(\tau + t, x - t) && (8)\\
& + \int_0^t \boldsymbol{C}(x + \tau)\exp\{\boldsymbol{H}\,(\tau + t - s)\}\,\boldsymbol{J}\,\boldsymbol{J}' \int_0^{\tau+t-s} \exp\{\boldsymbol{H}\,u\}\,\boldsymbol{C}'(x - t + s + u)du\,ds \\
& + \boldsymbol{C}(x + \tau)\exp\{\boldsymbol{H}\,\tau\}\underbrace{\int_0^t \exp\{\boldsymbol{H}\,(t - s)\}\,\boldsymbol{J}\,dW_s}_{=Z_t}.
\end{aligned}
$$

The intuition is that for any specification of $\boldsymbol{\sigma}(\cdot, \cdot)$ other than (7), the innovations of the Brownian path (W_t) over $[0, t]$ will be convoluted dissimilarly for different ages x and forecasting terms τ, so that the resulting model will not be finite-dimensionally realized. That is, the form in (7) is necessary and sufficient for the existence of a *finite-dimensional realization* of the model (2) (see e.g. Björk and Gombani (1999) or Filipović and Teichmann (2004) for interest rate modeling).

Equations (7) and (8) now present a semi-parametric representation, and the task for specifying a *self-consistent forward mortality* or *mortality surface model* is to find appropriate choices for D, m, $\boldsymbol{H}$, $\boldsymbol{J}$, and $\boldsymbol{C}(\cdot)$ in the context of given mortality data.

3 Forward Mortality Model: Specification and Estimation

As indicated, instead of relying on realized age-specific death rates, we base our statistical analysis on a time series of age-specific mortality forecasts. We first describe our data sources, and then detail how we use these data to specify and estimate different models for mortality forecasts.

3.1 Data

There are at least three potential data sources for age-specific mortality forecasts: (i) cohort mortality tables as employed by insurance companies, pension funds, or government agencies; (ii) mortality forecasts based on different windows of past mortality experience for a given population, using a fixed projection methodology; and (iii) mortality forecasts stripped from prices of pension annuities or other life-contingent insurance contracts.

Since insurance prices in (iii) are obscured by insurance expenses, idiosyncrasies of the particular insured population, and credit risk, we limit our analyses to (i) and (ii), with an emphasis on (ii). More precisely, for (i) we rely on U.K. life tables and projections for pension annuities as published by the *Institute and Faculty of Actuaries*. And for (ii), we rely on gender- and region-specific mortality data from the *Human Mortality Database* in concert with three popular mortality models to generate forecasts: The Lee-Carter model, the two-factor CBD model (Cairns et al., 2006b), and the P-spline method (Currie et al., 2004).[4]

However, we note that when utilizing these mortality forecasting models in (ii), we do not adopt the assumptions on the forecasting errors but view them merely as methodologies to generate *deterministic* extrapolations, which is common actuarial practice to generate trend-adjusted life tables; and, indeed, "Mortality improvement modeling" via the Lee-Carter and the CBD models that we rely on here is common (Dickson et al., 2020, Chapter 19). Our approach may be interpreted as an attempt to devise a suitable "stochastic wrap" around resulting deterministic forecasts. Thus, it can also be used in situations where generating different paths of age-specific mortality forecasts— rather than just a central projection—is not straightforward, such as within the P-spline method also considered here or within non-stochastic/statistical models such as demographic accounting methods that Land (1986) highlights for long-term mortality forecasting.

To keep the presentation concise, if not mentioned otherwise, in the main text results and statements refer to our primary data set relying on approach (ii) using mortality for the U.S. female population and the Lee-Carter approach[5] for generating forecasts. Due to changes in demographic

---

[4]Our approach is not limited to these three forecasting models, and can be applied to more complex models (incl. cohort effects, other factors, e.g.). We choose these as popular representatives.

[5]For the Lee-Carter parameter estimation, we use the modified weighted least-squares algorithm from Wilmoth (1993) and further adjust the driving process by fitting a Poisson regression model to the annual number of deaths at each age (Booth et al., 2002).

patterns around the middle of the twentieth century (Wilmoth, 2005), we rely on post World War II data only. Specifically, we have cohort mortality tables compiled for forty-three consecutive years (1974-2016) each using historical mortality data of the past thirty years (that is, for the table for year 1974, data from years 1945-1974 is employed; for the table for year 1975, data from years 1946-1975 is employed; etc.). Hence, the resulting time series comprises 43 mortality forecasts. As suggested above, the lag time l is chosen at 1. We use ages ranging from 30 to 95, and choose $\tilde{\mathcal{C}}$ as large as possible, which yields $K = 2145$. We also considered different choices for the time windows and/or choosing the lag $l = 5$, obtaining analogous results.

For some analyses, we also show results based on (i) the U.K. pensioner life tables, as well as for similar data sets based on (ii) and using other combinations of gender (male, female); region (England & Wales, France, Japan, and the U.S.); and mortality forecasting methods (Lee-Carter, CBD, and P-spline). Online Appendix F shows results for all relevant analyses. We also refer to the description there for more details on all the data sets.

3.2 Factor Analysis and Model

Having available a time series of mortality forecasts, as outlined above we calculate the $\boldsymbol{F}^j$, $j = 1, \ldots, N$, based on Equation (3). Following the standard procedure to obtain principal components, we calculate the empirical mean $\bar{\boldsymbol{F}}^{\cdot}$ and covariance matrix $\hat{\boldsymbol{\Sigma}}$, and then determine the eigenvalues and eigenvectors of $\hat{\boldsymbol{\Sigma}}$. Note that each K-dimensional column in the data matrix $(\boldsymbol{F}^1, \ldots, \boldsymbol{F}^N)$ is a function of age x and term τ, and therefore so are the eigenvectors. In particular, we can illustrate the eigenvectors—which in the factor model (4) govern how different age/term improvement factors, $F^j(\tau, x)$, are affected by the random shock ε—as a function in the two variables.

Table 1 and Figure 2 provide key empirical results. Table 1 shows the four largest eigenvalues (factor loadings) for a selection of data sets, and Figure 2 plots first corresponding scaled eigenvectors (factors) $\boldsymbol{u}_1\sqrt{\lambda_1}$ as a function of x and τ. The leading factor explains the vast majority of the variation in the data for all considered sets (minimum: 71.73%, mean: 93.01%, median: 96.03%), and the corresponding factor is highly systematic. In particular, for all data sets the factor entries increase exponentially in both the age and the term dimension, whereas the diagonals (where age at maturity, $x + \tau$, is constant) are increasing and concave.

The exponentially increasing pattern in age of course is familiar for mortality data, and indeed this is part of the reason why the first factor dominates. However, when considering the eigen-decomposition of the correlation matrix instead of the covariance matrix, we still find that the first factor accounts for roughly 75% of total variation. Furthermore, when determining the appropriate number of factors relying on the framework from Bai and Ng (2002), a single-factor model results

Sample	Value (Weight)			
	λ_1	λ_2	λ_3	λ_4
U.S. female, Lee-Carter	0.0241 (94.47%)	0.0009 (3.44%)	0.0002 (0.94%)	0.0001 (0.45%)
U.S. female, CBD	0.0297 (98.54%)	0.0004 (1.19%)	0.0001 (0.24%)	0.0000 (0.02%)
U.S. female, P-spline	0.1481 (98.22%)	0.0023 (1.50%)	0.0003 (0.22%)	0.0001 (0.04%)
U.S. male, Lee-Carter	0.0401 (90.36%)	0.0023 (5.17%)	0.0011 (2.37%)	0.0005 (1.20%)
French female, Lee-Carter	0.0383 (96.03%)	0.0008 (2.09%)	0.0003 (0.77%)	0.0002 (0.55%)
Japanese male, Lee-Carter	0.0384 (91.31%)	0.0017 (4.08%)	0.0009 (2.18%)	0.0005 (1.23%)
England & Wales female, CBD	0.1410 (97.20%)	0.0038 (2.65%)	0.0002 (0.12%)	0.0000 (0.03%)
U.K. pensioner	0.0229 (71.73%)	0.0076 (23.86%)	0.0009 (2.84%)	0.0004 (1.21%)

Table 1: Showing the absolute values and relative weights of the four largest eigenvalues based on principal component analysis of various data sets. The data sets are chosen as representative combinations of forecasting method and gender/region-specific mortality data as available from the *Human Mortality Database* (see Online Appendix F for all data sets). We also consider U.K. life tables and projections for pension annuities.

(a) U.S. female, Lee-Carter   (b) U.S. female, CBD   (c) U.S. female, P-spline   (d) U.S. male, Lee-Carter

(e) E & W female, CBD   (f) French fem., Lee-Carter   (g) Jap. male, Lee-Carter   (h) U.K. pensioner
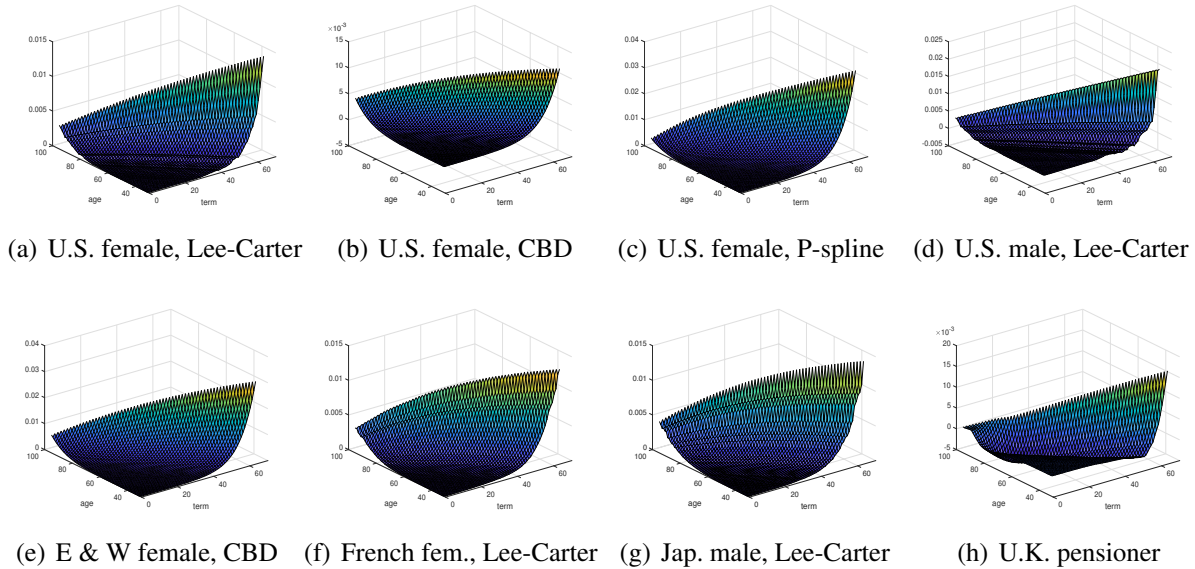
Figure 2: Showing the first eigenvectors (*slope factor*) that correspond to the data sets considered in Table 1 (see Online Appendix F for all data sets). Each eigenvector is plotted as a function of age (x) and term (τ), with $x \in \{31, \ldots, 95\}$, $\tau \in \{0, \ldots, 64\}$, and $x + \tau \in \{31, \ldots, 95\}$.

in the smallest value of the penalty function (Bai and Ng, 2002, p. 201). So again we have evidence that the first factor dominates and captures the most important patterns. The corresponding factors based on maximum likelihood look highly similar to those from the conventional factor analysis.

The increasing pattern on the diagonal implies that a random shock (ε in Equation (4)) affects mortality in the future more severely, i.e. mortality in the future is *more volatile* than near-term mortality. This contrasts implicit patterns for long-term mortality in popular mortality models for the hazard—that were conceived based on time series of (period) mortality *rates*. For instance, mean-reverting mortality models that are used for evaluating life contingencies in the actuarial literature imply a *decreasing* diagonal pattern, i.e. lower volatility for mortality in the future (see e.g. Milevsky and Promislow (2001); Dahl and Møller (2006), among others, for mortality models based on mean-reverting processes). A unit-root AR(1) process driving (log)-mortality as within Lee-Carter and similar models implies a flat diagonal pattern, so rates are affected similarly across the term structure. In analogy to the literature on forecasting the yield curve, we refer to this leading factor as the *slope factor* as the influence varies with the term. The shape for the second factor (eigenvector) $\boldsymbol{u}_2$ is drastically different than that of the leading factor—but again it is similar/systematic across all considered sets. We omit the discussion here, referring to Online Appendix F for illustrations and discussion (as explained there, we refer to it as the *curvature factor*).

The factor model (4) as well as the non-parametric forecasting approach from Algorithm B.1 in B rely on the assumption that the $\boldsymbol{F}^j$, $j = 1, \ldots, N$, are i.i.d. (which is implicit from the dynamics (2), cf. Prop. 2.1). To validate whether the assumption is satisfied in the data, we run (univariate) tests for independence (Ljung-Box test) and Normality (Jarque-Bera test) for the components of $\boldsymbol{F}$ (combinations of x and τ). We find that at the 95% level, only 78 of the 2145 components (3.64%) reject the Normality test. However, the Ljung-Box test for independence rejects for 406 out of the 2145 components (18.93%). In line with this observation, independence is rejected for the first principal component $(\boldsymbol{u}_1 \sqrt{\lambda_1})' \boldsymbol{F}$ (p-value 0.01), whereas Normality is not rejected.

It turns out that the key driver for the lack of independence are the components for high values of τ. Indeed, when considering the weights corresponding to the second eigenvector, $(\boldsymbol{u}_2 \sqrt{\lambda_2})' \boldsymbol{F}$, the tests neither reject Normality nor independence (p-value 0.29)—since the second principal component puts less weight on high-τ combinations. Similarly, when restricting the consideration to components with $\tau \leq 20$, only 100 of the 1155 (8.66%) univariate independence tests are rejected, and the independence test for $(\boldsymbol{u}_1 \sqrt{\lambda_1})' \boldsymbol{F}$ does not reject (p-value 0.11). Hence, it appears that violations are associated with forecasts in the far future, which may be linked to "irregularities" for long-term Lee-Carter forecasts (McNown, 1992). We note that for corresponding approaches for yield curve forecasting, the independence assumption for the component vectors is typically not satisfied, so the evolution of the factors is generally modeled via (vector)-auto-regressions (see

Piazzesi (2010) and references therein). Of course, a similar approach is feasible here.

Whether a one-factor or a higher/non-parametric model is suitable depends on the application. The observation that the first factor by far dominates points to a single-factor model. Indeed, e.g. for forecasting future life expectancies as within Figure 1, a single-factor model:

$$\boldsymbol{F} = \bar{\boldsymbol{F}}^{\cdot} + \phi \, \boldsymbol{u}_1 \, \varepsilon, \ \varepsilon \sim \mathrm{N}(0, 1), \tag{9}$$

delivers similar prediction intervals as the non-parametric model. Again in analogy to the literature on yield curve forecasting (Duffee, 2002, e.g.), it is straightforward to (re)-estimate the parameter for the factor loading ϕ: Simply multiply the (ortho-normal) vector $\boldsymbol{u}_1'$ from the left to the centered data vectors $\boldsymbol{F}^j - \bar{\boldsymbol{F}}^{\cdot}$, $j = 1, \ldots, N$, and determine ϕ as the sample standard deviation (similarly we can determine the factor loadings in multi-factor models).

3.3 Mortality Surface Model

As outlined in Section 2.2, the task for specifying a *self-consistent* model for mortality forecasts is to find appropriate choices for $D, m \in \mathbb{N}$, $\boldsymbol{J} \in \mathbb{R}^{m \times D}$, $\boldsymbol{H} \in \mathbb{R}^{m \times m}$, and $\boldsymbol{C} \in C^1([0, \infty), \mathbb{R}^m)$ in Equation (8) given our data sets. Based on our previous discussion, we limit our consideration to one-factor models so that we set $D = 1$. We describe extensions to multi-factor models in Online Appendix D.1.

Plugging μ_t into the basic Equation (1) for $_\tau p_x(t)$, and evaluating $F(\tau, x)$ in Equation (3) yields:

$$
\begin{aligned}
F^j(\tau, x) &= \mathbb{E}[F(\tau, x)] + \int_\tau^{\tau+1} \boldsymbol{C}(x + \nu) \, \exp\{\boldsymbol{H} \, \nu\} \, d\nu \int_{t_j}^{t_{j+1}} \exp\{\boldsymbol{H} \, (t_{j+1} - s)\} \, \boldsymbol{J} \, dW_s \quad (10) \\
&\approx \mathbb{E}[F(\tau, x)] + \int_\tau^{\tau+1} \boldsymbol{C}(x + \nu) \, \exp\{\boldsymbol{H} \, (\nu + 1)\} \, d\nu \, \boldsymbol{J} \, \underbrace{(W_{t_{j+1}} - W_{t_j})}_{=\epsilon_j \sim N(0,1)},
\end{aligned}
$$

where the approximation originates from the standard Euler-Maruyama method. This expression is directly analogous to the single-factor model (9) above. In particular, equating $F^j(\tau, x)$ with the factor model yields:

$$
(\phi \, \boldsymbol{u}_1)_{(\tau, x)} \approx \int_\tau^{\tau+1} \boldsymbol{C}(x + \nu) \, \exp\{\boldsymbol{H} \, (\nu + 1)\} \, d\nu \, \boldsymbol{J} \approx \underbrace{\boldsymbol{C} \left(x + \tau + {}^1/_2\right) \exp\left\{\boldsymbol{H} \left(\tau + {}^3/_2\right)\right\} \, \boldsymbol{J}}_{=\boldsymbol{\sigma}(\tau + {}^3/_2, x-1)},
$$
$$\tag{11}$$

where the second approximation originates from a standard midpoint rule. Hence, the principal components are directly related to the volatility structure in the dynamics (2).

In order to determine a functional structure that is able to capture the shape of the first eigenvector, and particularly the increasing and concave shape of the diagonals (Fig. 2), we rely on a modification of Example 6.2 in Björk and Gombani (1999) for interest models and set:

$$\boldsymbol{C}(x+\tau) = f(x+\tau) \times \begin{bmatrix} 0 & 1 \end{bmatrix}, \boldsymbol{H} = \begin{bmatrix} -2b & -b^2 \\ 1 & 0 \end{bmatrix}, \text{ and } \boldsymbol{J} = \begin{bmatrix} 1-ab \\ a \end{bmatrix}, \text{ which yields:}$$

$$\sigma(\tau, x) = f(x+\tau)(a+\tau)\exp(-b\tau). \tag{12}$$

Using this functional form in Equation (11), we rely on least-squares regression to determine a non-parametric estimate for $f(\cdot)$. The solid blue curves in Panels (b) and (c) of Figure 3 illustrate the fit. We also include a parametric fit based on a simple Gompertz-Makeham function $f(z) = k + c \times \exp(r \times z)$ (red dotted curves), one of the most common analytical mortality laws (Bowers et al., 1997, Chap. 3.7). As is usual in the mortality context, it provides a good match, especially for higher ages. In Panel (d) we plot the predicted parametric approximation for the scaled leading eigenvector $\boldsymbol{u}_1\sqrt{\lambda_1}$, which we re-plot in Panel (a) for a direct comparison (identical to Panel (a) in Figure 2). As is evident, the functional form is able to accurately capture the relationship.

Using the Gompertz-Makeham functional form, we thus have a parametric model for mortality forecasts with parameters k, c, and r (Gompertz-Makeham form); plus a and b (dynamics). Since the $\boldsymbol{F}^j$s are i.i.d. Normal (Prop. 2.1), we can obtain maximum likelihood estimates by deriving the mean and the covariance matrix as functions of the parameters from expressions (8) and (10).

Proposition 3.1. *We have for the mean $\boldsymbol{M} = (\boldsymbol{M}_{(\tau_i,x_i)})_{(\tau_i,x_i)\in\mathcal{C}}$ and the covariance matrix $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_{(\tau_i,x_i),(\tau_j,x_j)})_{(\tau_i,x_i),(\tau_j,x_j)\in\mathcal{C}}$ of $\boldsymbol{F}$:*

$$\boldsymbol{\Sigma}_{(\tau_i,x_i),(\tau_j,x_j)} = \int_0^1 \int_{\tau_i+1-s}^{\tau_i+2-s} \boldsymbol{\sigma}(u, x_i-1+s)\,du \int_{\tau_j+1-s}^{\tau_j+2-s} \boldsymbol{\sigma}'(u, x_j-1+s)\,du\,ds$$

$$\boldsymbol{M}_{(\tau_i,x_i)} = \frac{1}{2}\boldsymbol{\Sigma}_{(\tau_i,x_i),(\tau_i,x_i)} + \int_0^1 \int_{\tau_i+1-s}^{\tau_i+2-s} \boldsymbol{\sigma}(u, x_i-1+s)\,du \int_0^{\tau_i+1-s} \boldsymbol{\sigma}'(u, x_i-1+s)\,du\,ds.$$

We note the different integration limits in the second term for $\boldsymbol{M}_{(\tau_i,x_i)}$, which arise since in the survival probabilities (that are subject to the martingale constraints (5)) the innovations in the different components of $\boldsymbol{F}$ aggregate. Plugging in the volatility specification (12) with the Gompertz-Makeham form for $f(z)$, we can solve the integrals in closed form. We omit the presentation here since *Mathematica* outputs more than 2,000 terms, although simplifications may be possible.
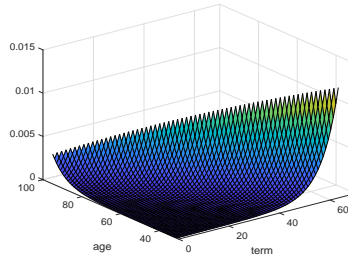
We derive parameter estimates by (numerically) maximizing the log-likelihood function. To avoid numerical instabilities, we set $\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \zeta \times \boldsymbol{I}_K$ similar as in ridge regression or as for the measurement equation in state-space models. Intuitively, we allow for non-systematic deviations to the model-implicit vectors accounting for noisy mortality observations as well as variation not
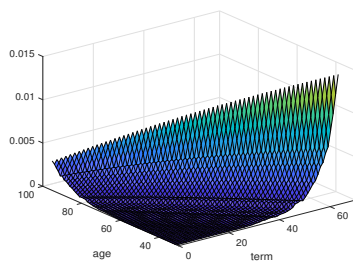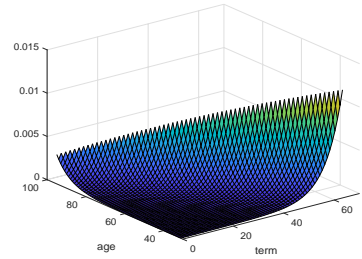
(a) First eigenvector (PCA)   (b) Fit $f(x+\tau)$, $x+\tau \in \{31, \ldots, 62\}$ (c) Fit $f(x+\tau)$, $x+\tau \in \{63, \ldots, 95\}$
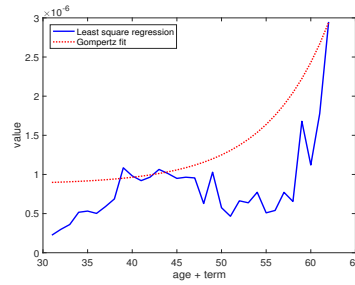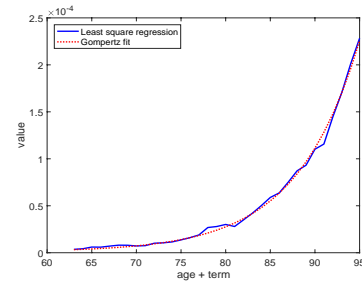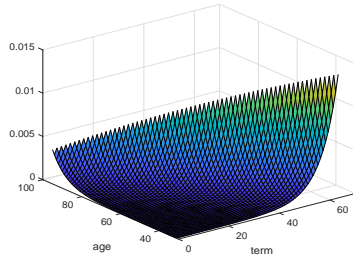
(d) Predicted first ev. (least-squares)   (e) Pred. first ev. (MLE, wo self-con.)   (f) Pred. first ev. (MLE, w\ self-con.)

Figure 3: Showing fits of the first eigenvector. Panel (a) replots Figure 2(a) for comparison. Panels (b) and (c) show the non-parametric estimate for $f(x+\tau)$ (blue solid line) and Gompertz-Makeham fit (red dotted line). Panels (d) to (f) show the parametric prediction under the least-squares regression, MLE without self-consistency condition, and MLE with self-consistency condition, respectively.

	Parameter point estimate (st. dev.)				
	k	c	r	a	b
w\ self-consistency	$1.57 \times 10^{-6}\,(1.0 \times 10^{-6})$	$1.95 \times 10^{-10}\,(3.6 \times 10^{-10})$	$0.146\,(0.01)$	$16.54\,(2.1)$	$0.0042\,(0.001)$
wo\ self-consistency	$1.96 \times 10^{-6}\,(1.2 \times 10^{-6})$	$2.00 \times 10^{-10}\,(3.3 \times 10^{-10})$	$0.145\,(0.01)$	$16.65\,(2.2)$	$0.0041\,(0.001)$

Table 2: Showing maximum likelihood point estimates for the parameters k, c, r, a, and b, both with and without the self-consistency constraint. Standard deviations (in parentheses) based on 100 bootstrap samples.

accounted for by only considering the leading factor, so we set the parameter ζ as the sum of all remaining eigenvalues. We note that while the results are not very sensitive to the choice of ζ, alternative specifications of the non-systematic deviations, such as adding noise *relative* to the variance of each observation ($\zeta \times \text{diag}\{\boldsymbol{\Sigma}\}$), will change the estimates substantially. We work with $\tilde{\boldsymbol{\Sigma}}$ since for forecasts of life expectancy or of the population age-distribution, what is relevant are absolute mortalities so that higher relative errors for low ages have little impact—as was already pointed out by Lee and Carter (1992, p. 662).

Table 2 presents our estimation results. We provide two sets of estimates that differ by whether they incorporate cross-sectional self-consistency constraints associated with the interpretation as forecasts. More precisely, note that within Equation (8), the model parameters not only affect the volatility term but also the expected values (this originates from the drift condition (6)). In particular, unlike the mortality factor models (4)/(9) where the expected value is simply taken to be the sample mean, Proposition 3.1 demonstrates that $\mathbb{E}[F(\tau, x)]$ in (10) is also given in terms of the model parameters. In the first set of estimates ("w\ self-consistency"), we enforce these constraints, whereas in the second set of estimates ("wo\ self-consistency"), we simply plug in the sample mean for $\mathbb{E}[F(\tau, x)]$. Hence, one can interpret the latter model/estimates as a version of the single-factor model (9), where we replace the vector $\phi\,\boldsymbol{u}_1$ by the parametric form (10).

Table 2 gives point estimates as well as bootstrapped standard errors (based on one hundred non-parametric bootstrap samples of the $\boldsymbol{F}$ realizations). The Gompertz parameter r is similar to but on the high end relative to other studies, possibly because of the forward-looking nature of our analysis. In particular, note that here r does not directly impact mortality, but solely the mortality volatility. The dynamic parameters a and b suggest an increasing, concave diagonal as observed in Figure 2. For comparison, we plot the factor surfaces (first eigenvector) based on the parameters without and with enforcing the self-consistency constraint in Figure 3, Panels (e) and (f), respectively. The shape is similar to the factor from the PCA (Panel (a)) and the least-squares approximation (Panel (d)), suggesting that the parametric model captures the relevant patterns and that the maximum-likelihood estimation is in line with our previous results. The bootstrap error of the mortality level c is large relative to the estimate, which is due to a few large estimates for c

for certain bootstrap samples. However, in these samples the other parameters, particularly r and b in the exponents, also show relatively large deviations to their point estimates, and in concert the volatility function (12) still closely resembles the shape in Figure 3.

The most blatant result from Table 2, however, is that the two sets of estimates are highly similar. To interpret this finding, we lean on the interest rate modeling literature, where the necessity—and also the conduciveness—of imposing similar cross-sectional constraints that arise from no-arbitrage is heavily debated (Duffee, 2002; Diebold and Li, 2006; Piazzesi, 2010; Duffee, 2011, among others). As pointed out by Piazzesi (2010), the constraints should increase the efficiency of estimates, particularly since in our setting there are no risk premiums segregating the dynamic (t) and cross-sectional (τ) perspective. However, our bootstrapped errors do not clearly support this. Possibly the efficiency gain may become apparent as the data availability (length of series) increases.

An important insight from Duffee (2011), that imposing cross-sectional constraints should not invalidate estimates resulting in their absence as long as these constraints are satisfied in the data, is relevant in our context. In other words, if the estimates significantly change after the self-consistency condition is imposed, this is an indication that the cross-sectional constraints do not hold true in the data—or that current forecasts deviate from future realizations in some *non-random* fashion. Indeed, it is possible to formalize this intuition via a misspecification test, e.g. by comparing the empirical distribution of $Y = (\boldsymbol{u}_1 \sqrt{\lambda_1})^T \boldsymbol{F}$ to the (known) distribution when enforcing self-consistency. Given the similarity of the "w\" and "wo\" estimates, clearly here such a test does not reject—lending support that the forecasts and the model are not misspecified.

We illustrate the model in Figures 1 (a) and 4. Figure 4 presents box-whisker plots of future cohort life expectancy from the terminal point of the underlying mortality time series t_0, for 50-year olds in 20 years $\mathring{e}_{50}(t_0 + 20)$ ($\mathring{e}_{50}(t_0) = 34.59$) and for 70-year olds in 10 years $\mathring{e}_{70}(t_0 + 10)$ ($\mathring{e}_{70}(t_0) = 16.54$). We show results for different approaches, from left to right: (i) the Lee-Carter model forecasts with error terms; (ii) the single-factor model (9); and (iii) the mortality surface model (8) with enforcing the cross-sectional constraint (we return to the remaining two models later). Panel (a) in Figure 1 shows projected cohort life expectancies for 70-year old US females $\mathring{e}_{70}(t)$ over time with corresponding prediction intervals based on Lee-Carter and our mortality surface model.

There are two main observations. First, we observe that while the median projections are very close in all cases, the variation is far greater for the approaches proposed here relative to the Lee-Carter intervals. The similarity in the median projections arises since the (deterministic) Lee-Carter forecasts are the basis for all projections. The increased variation is in line with previous assessments (Cocco and Gomes, 2012; Government Actuary's Department, 2001, e.g.), and the key driver is the high uncertainty in mortality projections for the far future, as demonstrated by

(a) $\overset{\circ}{e}_{50}\,(t_0 + 20)$
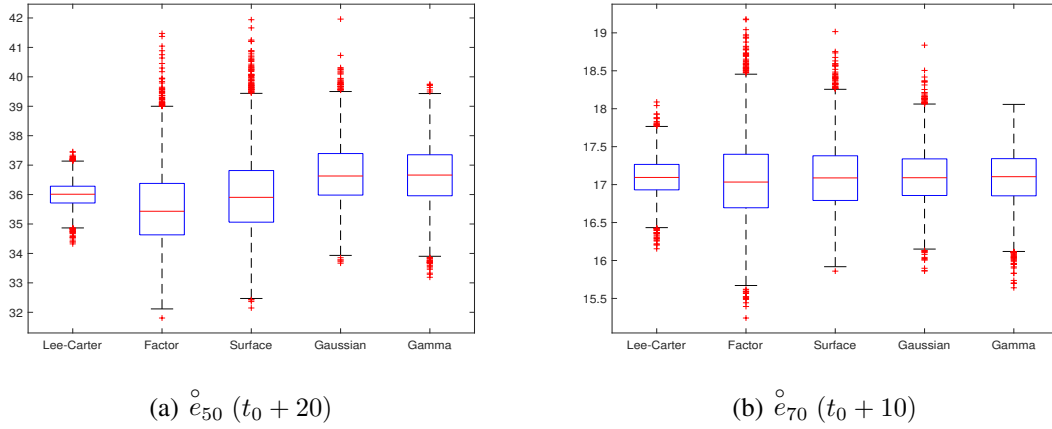
(b) $\overset{\circ}{e}_{70}\,(t_0 + 10)$

Figure 4: Showing realizations of expected future life-time for age-50 in 20 years (left panel) and age-70 in 10 years (right panel), $t_0 = 2016$. Left-to-right: Lee-Carter predictions; one-factor model (9); mortality-surface model (w\ self-consistency); Gaussian Gompertz-hazard model; Gamma Gompertz-hazard model.

the shape of the leading principal component. Second, the factor and the mortality surface model produce highly similar results. This validates the parametric model, as it performs very similar to the (purely) statistical factor models. These observations carry through across all the considered data sets. We refer to Online Appendix F for details, where we also demonstrate that the surface model without self-consistency constraints and the non-parametric bootstrapping forecasts perform very similarly to the two models showcased here.

4 Models for the Hazard

4.1 Dynamics

Models for mortality forecasts and for mortality rates are theoretically equivalent, as a stochastic model for mortality experience in every period implies a stochastic mortality forecast and vice versa. In particular, we can determine the model for the hazard $\mu_t(0, x)$—or the *spot force of mortality*—for our mortality surface model (8) going forward (we denote t_0 as the current time):

Proposition 4.1. *The spot force of mortality (hazard) for the model* (8)*,* $\mu_t(0, x)$*, with* $\boldsymbol{C}$*,* $\boldsymbol{H}$*, and* $\boldsymbol{J}$ *as specified in the previous section (Equation (12)), and* $f(z) = k + c \times \exp(r \times z)$*, is:*

$$\mu_t(0, x) = \mu_{t_0}(t - t_0, x - t + t_0) + \int_{t_0}^{t} \alpha(t - s, x - t + s)\, ds + (k + c\exp(rx)) \times Z_t^{(2)}, \ t \geq t_0, \ (13)$$

where $Z_t^{(2)}$ is the second component of Z_t with dynamics evolving according to:

$$dZ_t = \boldsymbol{H} Z_t \, dt + \boldsymbol{J} \, dW_t = \left\{ \begin{bmatrix} -2b & -b^2 \\ 1 & 0 \end{bmatrix} \times Z_t \right\} dt + \begin{bmatrix} 1 - ab \\ a \end{bmatrix} dW_t, \ t \geq t_0. \qquad (14)$$

The more familiar hazard formulation of the model reveals various characteristics that are not obvious from the forward model (1) with volatility given in (12). First, while the forward model is stationary/time-homogeneous, this is not the case for the *spot* model (13)/(14) as time trends are ingrained by the current surface $\mu_{t_0}(\cdot, \cdot)$. Second, while the underlying forward model is itself Markovian, this is again not the case in (13)/(14): The evolution of mortality depends on the two-dimensional Markov process (Z_t), although it is driven by the single source of randomness (W_t). More precisely, $\mu_t(0, x)$ is (only) a function of its second component $Z_t^{(2)}$, which changes with "shocks" dW_t via the parameter a. This contemporaneous impact of a random shock on mortality is also present in conventional models such as Lee-Carter. In contrast to conventional models, the local behavior of (W_t) also has a *persistent* impact on the evolution of mortality rates via the other component $Z_t^{(1)}$—which appears as the drift term of $Z_t^{(2)}$ in the dynamics of (Z_t). This persistent impact is exactly what generates variability in the long run. However, it is important to emphasize that this structure emanates from the data. In particular, our analyses show that both uncertainty channels reflected in the dynamics (14) are relevant to modeling the evolution of mortality risk.

To illustrate the dynamics, in Panel (a) of Figure 5 we plot the historical paths ($t \leq t_0$) of the processes $(Z_t^{(1)})$ and $(Z_t^{(2)})$ as well as the underlying Brownian motion (W_t) (random walk) based on our estimation process, and in Panels (b) and (c) we show two simulated future paths ($t \geq t_0$). The first component mostly traces the underlying random walk, whereas (W_t) affects the second component that directly determines the hazard in two ways: A shock dW_t translates into a shock $dZ_t^{(2)}$ on mortality as governed by the parameter a, *and* via changing its drift term $Z_t^{(1)}$. In particular, via the second channel, large changes in (W_t) have *lasting* impact on the path of $(Z_t^{(2)})$.

4.2 Gaussian Hazard Model

The forward model takes the current mortality surface $\{\mu_{t_0}(\tau, x)\}$ as an input and models its dynamics, so that μ_{t_0} enters the dynamics of the spot model (13)—similarly to how the current yield curve enters the Hull and White (1990) interest rate model. This is a feature in that it allows the model to reflect nuances in the original surface. However, it can also be a drawback as the model is less user-friendly with regards to the necessary ingredients—one requires $\{\mu_{t_0}(\tau, x)\}$ to use it—and as the model is less transparent. In contrast, authors in the actuarial and financial engineering literatures have proposed stochastic versions of classical mortality laws that present transparent models for pricing or hedging mortality-contingent exposures (Milevsky and Promis-
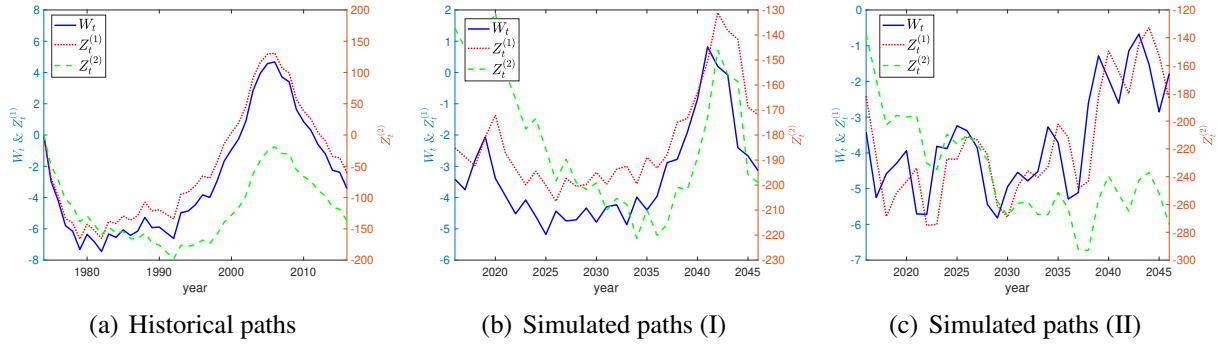
(a) Historical paths   (b) Simulated paths (I)   (c) Simulated paths (II)

Figure 5: Showing path of the processes W_t, $Z_t^{(1)}$, and $Z_t^{(2)}$. Panel (a) is based on the MLE parameters (w\ self-consistency) in conjunction with Equations (10) and Proposition 4.1. Future path of W_t are simulated and then $Z_t^{(1)}$ and $Z_t^{(2)}$ are obtained via Proposition 4.1. Given our data, $t_0 = 2016$.

low, 2001; Dahl and Møller, 2006; Tappe and Weber, 2014, among many others). Yet, as indicated above, those mortality models rely on familiar processes from financial engineering for modeling the stochastic component. As a consequence, they are not suitable for capturing the large and increasing uncertainty in long-term mortality forecasts.

We can proceed analogously. The most parsimonious approach is to approximate the age pattern in the deterministic component $[\mu_{t_0}(t - t_0, x - t + t_0) + \int_{t_0}^{t} \alpha(t_0 - s, x - t_0 + s)\, ds]$ in (13) by the Gompertz-Makeham form $\xi \times (k + c\exp(rx))$. The mortality model then takes the simple form:

$$\mu_t(0, x) = (k + c\exp(rx)) \times \tilde{Z}_t^{(2)}, \ t \geq t_0, \ \tilde{Z}_{t_0} = (Z_{t_0}^{(1)}, \xi). \tag{15}$$

However, the deterministic component ingrains the baseline future trend of mortality, which is not captured by this model in concert with dynamics (14). A straightforward way is to include the trend via a drift term d in $Z_{t_0}^{(2)}$, the component that directly governs $\mu_t(0, x)$:

$$d\tilde{Z}_t = \left\{ \begin{bmatrix} -2b & -b^2 \\ 1 & d \end{bmatrix} \times \tilde{Z}_t \right\} dt + \begin{bmatrix} 1 - ab \\ a \end{bmatrix} dW_t \equiv \tilde{\boldsymbol{H}} \tilde{Z}_t \, dt + \boldsymbol{J} \, dW_t, \ \tilde{Z}_{t_0} = (Z_{t_0}^{(1)}, \xi), \ t \geq t_0. \tag{16}$$

The model (15)/(16) performs well. We use the same parameters from our forward model but calibrate the deterministic trend d and the initial state ξ to match (average) cohort life expectancy for 70-year-old U.S. females at $t_0 = 2016$ and as projected in 20 years ($t_0 + 20$), resulting in $\xi = 1,675.01$ and $d = -0.0078$. Again, Figure 4 presents resulting box-whisker plots for future life expectancies for the two age/term combinations (labeled "Gaussian") and Panel (b) of Figure 1 presents resulting 95% prediction intervals for $\overset{\circ}{e}_{70}(t)$, $t \geq t_0$ (blue dotted line).

In Figure 4(b), the resulting realizations range very similar to the forward/surface model for

70-year olds. Moreover, the prediction intervals in Figure 1(b) are similar to Panel (a) though slightly tighter, which may be an artifact of our calibration procedure—since we did not consider adjustments in the dynamics parameters a and b that may interact with our choices for ξ and d. The box-whisker plot for 50-year olds in Figure 4(a) is also broadly similar to the surface model—and certainly show more variation than Lee-Carter. However, the results are a little offset, which is due to our reliance on 70-year olds' life expectancies in the calibration process. If we chose life expectancies for 50-year olds for calibration, we would have alignment here but at the same time an offset for the 70-year olds. This demonstrates the tradeoff between simplicity of the hazard model (15)/(16) and the forward model that allows for distinctive trends via the initial surface μ_{t_0}.

Thus, the resulting *Gaussian Gompertz model* (15)/(16) captures the dynamics; with its six parameters ($\{a, b, d\}$ for dynamics, $\{k, c, r\}$ for the cross-sectional pattern of mortality) plus the initial states ($\tilde{Z}_{t_0}$) it is parsimonious and transparent; and as an affine model it is highly tractable. We refer to Online Appendix C.1 for derivations of survival probabilities and other quantities.

However, since $\tilde{Z}_t$ follows a Gaussian law, negative realizations for $\mu_t(0, x)$—or, equivalently, survival probabilities exceeding unity—are possible. The extent to which this happens primarily depends on the trend parameter d. For $d = 0$ as in the dynamics (14), negative outcomes are exceedingly rare so that they can be ignored ($< 3 \times 10^{-7}$ over the next 70 years). However, when calibrating d to match future life expectancies as above, while survival probabilities above one are still rare ($< 1 \times 10^{-4}$ across all experiments), they do occur with large numbers of simulated paths.

4.3 Gamma Hazard Model

This section introduces a version of the model that performs very similarly but can rule out negative hazard rates. Accomplishing this in the class of continuous-time affine models that preserve tractability proves delicate, which according to Gourieroux et al. (2006) is "likely due to a lack of flexibility of these models [...] due to the continuous-time assumption." However, it is relatively straightforward to specify a non-negative discrete-time analogue of our Gaussian model.

Discretizing our Gaussian hazard model (15)/(16) (Euler-Maruyama method), we obtain:

$$\tilde{Z}_{t+1} = (\tilde{\boldsymbol{H}} + \boldsymbol{I})\,\tilde{Z}_t + \boldsymbol{J}\,\varepsilon_{t+1} = (\tilde{\boldsymbol{H}} + \boldsymbol{I})\,\tilde{Z}_t - \theta\,\boldsymbol{J} + \boldsymbol{J}\,\gamma_{t+1},\; t \geq t_0,\; \tilde{Z}_{t_0} = (Z_{t_0}^{(1)}, \xi), \quad (17)$$

where $\boldsymbol{I}$ is the identity matrix, ε_t are i.i.d. standard Normal, and $\gamma_t = \varepsilon_t + \theta$ are i.i.d. θ-shifted Normal with parameter θ. Following ideas from Gourieroux and Jasiak (2006) for positively valued time series, we simply replace the assumption on γ_t by an i.i.d. Gamma distribution with shape parameter θ^2 and scale parameter $1/\theta$. As a consequence, the first two moments of γ_t—and, thus, those of $\tilde{Z}_{t+1}|\tilde{Z}_t$—remain the same in the *Gamma Gompertz model* (15)/(17) as in the original Gaussian model. However, since now the distribution of the innovations has positive support, it is

Initial age & policy term	$(30, 35)$	$(40, 25)$	$(50, 15)$
Gaussian hazard	1.64% / 54.72%	1.43% / 35.69%	1.24% / 18.83%
Gamma hazard	1.63% / 53.96%	1.42% / 34.73%	1.23% / 17.60%
Lee-Carter	1.19% / 12.87%	1.17% / 11.60%	1.14% / 9.14%
Dahl-Møller	1.02% / 0.63%	1.02% / 0.59%	1.02% / 0.79%

Table 3: Showing the value/increase of a Guaranteed Annuity Option. Value as a markup on the initial investment for an at-the-money guaranteed rate for different age and policy term combinations under various mortality models. Increase as compared to using deterministic mortality rates (interest rate risk only).

immediate from (17) that given a set of parameters, starting values, and forecasting horizon, the process $(\tilde{Z}_t^{(2)})$—and thus the hazard rate—will remain positive. We refer to Online Appendix C.2 for details, including closed-form approximations for survival probabilities.

Given the similarities, the results generated by the Gamma Gompertz model are very close to those from the Gaussian version. The choice of θ has a modest effect on the outcomes in the range we considered (see C.2). If θ is large, the distribution of γ_t approximates a Normal but at the same time there is a larger drag on $(\tilde{Z}_t)$ in (17), so that negative rates may be possible for longer forecasting horizons. A smaller θ rules out negative rates but the distribution becomes more dissimilar, which primarily affects near-term forecasts. We choose $\theta = 1$, which ascertains positivity across all our calculations—with the forecasting horizon being the terminal age—and generates very similar results to our other model versions. To illustrate, Figure 4 presents resulting box-whisker plots for future life expectancies for the two age/term combinations (labeled "Gamma") and Panel (b) of Figure 1 presents resulting 95% prediction intervals for $\mathring{e}_{70}(t)$, $t \geq t_0$ (green dotdashed line).

Thus, our Gamma Gompertz model (15)/(17) captures the dynamics; it is parsimonious (seven parameters), transparent, tractable, and *positive*. Other non-negative versions, for instance in the class of *compound autoregressive processes* (CaR, Gourieroux et al., 2006, e.g.), are also possible.

5 Application 1: Valuation of Guaranteed Annuity Options

Valuation and risk management of life-contingent insurance products—particularly ones with optional, non-linear features—is a key application of stochastic mortality models. As an illustrative example, we consider the valuation of Guaranteed Annuity Options (GAOs) in insurance savings policies (pure endowment). Within a GAO, when the policy expires the policyholder, if alive, can choose between either a lump-sum payment or a life annuity based on a guaranteed rate. In Zhu and Bauer (2011), we emphasize the advantages of factor or affine mortality models in this context, since future survival probabilities and, thus, annuity present values can be derived in closed form.

Our focus is on showcasing results for our mortality models with long-term risk, especially in

comparison to other, conventional mortality models. Therefore, in addition to showing results for our Gaussian and Gamma Gompertz hazard models and the Lee-Carter model, we also include results for the affine mortality model from Dahl and Møller (2006). Similarly to other papers on valuing and hedging mortality-contingent liabilities, they propose a stochastic extension of the Gompertz-Makeham model with mortality improvements evolving according to a Cox–Ingersoll–Ross process. In particular, their model features mean reversion. Since we use their parameters, we report results as a markup on the initial investment. More precisely, we determine the at-the-money guaranteed rate based on information at t_0 and then determine the markup for having the option to annuitize at poilcy expiration. Furthermore, we report the percentage increase of this markup due to stochastic mortality, i.e., we calculate the value increase relative to a situation with interest rate risk only. For the interest rate, we assume a one-factor Vasicek model, and estimate the relevant parameters based on U.S. data from January 2009 to December 2019.

Table 3 shows GAO values for different age and policy term combinations. There are two key observations. First, for the mean-reverting mortality model ("Dahl-Møller"), the value increases for these long-term guarantees due to stochastic mortality are relatively modest ($< 1\%$). In contrast, the value increases originating from stochastic mortality are more substantial for Lee-Carter (around 10%) and very substantial for our models (20%-50%). Second, the impact of stochastic mortality is smaller in the longer run (greater policy terms) for the mean-reverting model and relatively flat for the Lee-Carter model. In contrast, our models capture the long-run risk in mortality projections, so that the relevance of mortality is more pronounced for longer guarantees.

6 Application 2: Longevity Risk Transfer Market and Longevity Derivatives

The global financial exposure to longevity risk in pension and insurance liabilities is massive, with estimates ranging from USD 60 to 80 trillion (Michaelson and Mulholland, 2014). Since insurers and reinsurers do not have the capacity to fully absorb this risk, a natural approach is to tap diversification opportunities of financial markets and since the mid-2000s a longevity risk transfer market has started to form. However, in relation to the potential size, the volume of longevity-linked securities coming to market "has been disappointingly low" (Blake et al., 2019, p. 1).

Michaelson and Mulholland (2014, p. 4) point to experienced financial institutions and engineers for developing adequate solutions. They argue for "out-of-the-money" hedges featuring option-like payouts based on customized indices that reflect the hedger's exposure in present value terms, particularly so-called (call) spread options. Using the present value of an annuity cash flow

	$(AP, EP) = (101\%, 105\%) \times \mathbb{E}_{t_0}[a_x(t_0 + T)]$		$(AP, EP) = (103\%, 107\%) \times \mathbb{E}_{t_0}[a_x(t_0 + T)]$	
	value	(P_1, P_2)	value	(P_1, P_2)
Gaussian hazard	0.0597	(31.47%, 1.62%)	0.0132	(8.91%, 0.16%)
Gamma hazard	0.0553	(33.97%, 0.34%)	0.0070	(7.30%, 0.00%)
Lee-Carter	0.0258	(24.40%, 0.02%)	0.0010	(1.56%, 0.00%)
Dahl-Møller	0.0003	(1.61%, 0.00%)	0	(0.00%, 0.00%)

Table 4: Showing spread option values based on four stochastic mortality models, relative increase to the deterministic mortality case, and trigger (P_1) and exhaustion (P_2) probabilities. Expiration $T = 15$, and age $x = 70$.

for an x-year old with a fixed interest rate $r = 3\%$, $a_x(t)$, as the underlying index, a spread option expiring in T years has payoff $\max(0, \min(a_x(t_0 + T) - AP, EP - AP))$, where AP is the attachment point and EP the exhaustion point. By only protecting the exposure above the attachment point, the hedger keeps some "skin in the game" reducing potential informational frictions (Biffis and Blake, 2014); and the exhaustion point limits the investor's exposure. As a second application of our models, similar to the previous section, we present values for longevity spread options on 70-year-old U.S. females expiring in $T = 15$ years (based on 50,000 simulated paths). Again, we compare them to results from conventional models. We also record trigger (P_1) and exhaustion (P_2) probabilities.

The results in Table 4 directly track those of the previous section: Values and trigger/exhaustion probabilities are substantially greater under our models with long-term risk. The mean-reverting model ("Dahl-Møller") shows very limited risk, but the values under our models also drastically exceed those under Lee-Carter—the benchmark model in industry and policy.

Blake et al. (2019) emphasize the role of models for the development of the longevity risk transfer market, and particularly point out that conventional models fail to "capture long-term changes in the trend in mortality rates" (page 34). In this paper, we present empirically-derived stochastic mortality models with long-term risk. Potentially these models can help jump-start the longevity risk transfer market by helping bridge the gap between the demand and supply side of longevity risk protection.

References

Bai J, Ng S (2002) Determining the number of factors in approximate factor models. *Econometrica* 70: 191-221.

Bauer D, Benth FE, Kiesel R (2012) Modeling the Forward Surface of Mortality. *SIAM J. Financial Math.* 3: 639-666.

Bell FC, Miller ML (2005) Life Tables for the United States Social Security Area 1900-2100. Report, Office of the Chief Actuary, Social Security Administration, Washington, DC.

Biffis E, Blake D (2014) Keeping some skin in the game: How to start a capital market in longevity risk transfers. *N. Amer. Actuarial J.* 18: 14-21.

Björk T, Gombani A (1999) Minimal realizations of interest rate models. *Finance Stochastics* 3: 413-432.

Blake D, Cairns AJG, Dowd K, Kessler AR (2019) Still living with mortality: The longevity risk transfer market after one decade. *Brit. Actuarial J.* 24: 1-80.

Booth H, Maindonald J, Smith L (2002) Applying Lee-Carter under conditions of variable mortality decline. *Population Stud.* 56: 325-336.

Bowers L, Gerber HU, Hickman JC, Jones DA, Nesbitt CJ (1997) *Actuarial Mathematics* (Society of Actuaries, Schaumburg, IL).

Broadie M, Detemple JB (2004) Option pricing: Valuation models and applications. *Management Sci.* 50: 1145-1177.

Broadie M, Kaya Ö (2006) Exact simulation of stochastic volatility and other affine jump diffusion processes. *Oper. Res.* 54: 217-231.

Buchardt K, Furrer C, Steffensen M (2019) Forward transition rates. *Finance Stochastics* 23: 975-999.

Cairns AJG, Blake D, Dowd K (2006a) Pricing Death: Frameworks for the Valuation and Securitization of Mortality Risk. *Astin Bull.* 36: 79-120.

Cairns AJG, Blake D, Dowd K (2006b) A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *J. Risk Insurance* 73: 687-718.

Christiansen MC, Niemeyer A (2015) On the forward rate concept in multi-state life insurance. *Finance Stochastics* 19: 295-327.

Christoffersen P, Heston S, Jacobs S (2009) The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Sci.* 55: 1914-1932.

Cocco JF, Gomes FJ (2012) Longevity risk, retirement savings, and financial innovation. *J. Financial Econom.* 103: 507-529.

Currie ID, Durban M, Eilers PHC (2004) Smoothing and forecasting mortality rates. *Statist. Model.* 4: 279-298.

Dahl M, Møller T (2006) Valuation and hedging of life insurance liabilities with systematic mortality risk. *Insurance Math. Econom.* 39: 193-217.

Dickson DC, Hardy M, Waters HR (2020) *Actuarial mathematics for life contingent risks*, 3rd ed. (Cambridge university press, Cambridge, MA).

Diebold FX, Li C (2006) Forecasting the term structure of government bond yields. *J. Econometrics* 130: 337-364.

Duffee GR (2002) Term Premia and Interest Rate Forecasts in Affine Models. *J. Finance* 57: 405-443.

Duffee GR (2011) Forecasting with the term structure: The role of no-arbitrage restrictions. Working Paper, Johns Hopkins University, Baltimore.

Filipović D, Teichmann J (2004) On the geometry of the term structure of interest rates. *Proc. Roy. Soc. London Ser. A: Math., Phys. Engrg. Sci.* 460: 129-167.

Gourieroux C, Jasiak J (2006) Autoregressive gamma processes. *J. Forecast.* 25: 129–152.

Gourieroux C, Monfort A, Polimenis V (2006) Affine models for credit risk analysis. *J. Financial Econometrics* 4: 494-530.

Government Actuary's Department (2001) National Population Projections: Review of Methodology for Projecting Mortality. Report, Government Actuary's Department, London, UK.

Hull J, White A (1990) Pricing interest-rate-derivative securities. *Rev. Financ. Stud.* 3: 573-592.

Kang C, Kang W, Lee JM (2017) Exact simulation of the Wishart multidimensional stochastic volatility model. *Oper. Res.* 65: 1190-1206.

Kou SG (2002) A jump-diffusion model for option pricing *Management Sci.* 48: 1086-1101.

Land KC (1986) Methods for National Population Forecasts: A Review. *J. Amer. Statist. Assoc.* 81: 888-901.

Lee RD, Carter LR (1992) Modeling and Forecasting U.S. Mortality. *J. Amer. Statist. Assoc.* 87: 659-671.

Litterman R, Scheinkman JA (1991) Common Factors Affecting Bond Returns. *J. Fixed Income* 1: 62-74.

McNown R (1992) Comment on 'Modeling and Forecasting US Mortality' by R.D. Lee and L.R. Carter. *J. Amer. Statist. Assoc.* 87: 671-672.

Michaelson A, Mulholland J (2014) Strategy for Increasing the Global Capacity for Longevity Risk Transfer: Developing Transactions That Attract Capital Markets Investors. *J. Alternative Investments* 17: 18-27.

Milevsky MA, Promislow SD (2001) Mortality Derivatives and the Option to Annuitize. *Insurance Math. Econom.* 29: 299-318.

Musiela M, Rutkowski M (1998) *Martingale Methods in Financial Modelling* (Springer, Berlin).

Piazzesi M (2010) Affine Term Structure Models. Ait-Sahalia Y, Hansen LP eds. *Handbook of Financial Econometrics vol. 1* (Elsevier, Amsterdam), 691-765.

Tappe S, Weber S (2014) Stochastic mortality models: an infinite-dimensional approach. *Finance Stochastics* 18: 209-248.

Thatcher R, Kannisto V, Andreev K (2002) The survivor ratio method for estimating numbers at high ages. *Demographic Res.* 6: 1-18.

Wilmoth JR (1993) Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change. Report, University of California, Berkeley.

Wilmoth JR (2005) Overview and Discussions of the Social Security Mortality Projections. Report, Social Security Advisory Board, Washington, DC.

Zhu N, Bauer D (2011) Applications of Gaussian Forward Mortality Factor Models in Life Insurance Practice. *Geneva Pap. Risk Insurance: Issues Pract.* 36: 567-594.

Online Appendix to
"Modeling the Risk in Mortality Projections"

This Online Appendix collects technical material and supplemental analyses. A provides the proofs for the propositions in the main text. B details the non-parametric and factor statistical mortality forecasting approaches. C presents derivations related to our hazard rate models as well as additional details. D discusses relevant extensions of our models. E demonstrates that prediction intervals based on Lee-Carter are *too* narrow. And F shows analyses presented in the main text for various alternative data sets/sources.

A   Proofs

Proof. Proof of Proposition 2.1 Taking a mild solution to the dynamics (2) (Da Prato and Zabczyk, 2014), simple evaluation yields:

$$
\begin{aligned}
&F_l(t_j, t_{j+1}, (\tau, x)) \\
&= \int_\tau^{\tau+l} \mu_{t_{j+1}}(v, x)\, dv - \int_{\tau+t_{j+1}-t_j}^{\tau+l+t_{j+1}-t_j} \mu_{t_j}(v, x - t_{j+1} + t_j)\, dv \\
&= \int_\tau^{\tau+l} \mu_{t_j}(v + t_{j+1} - t_j, x - t_{j+1} + t_j) + \int_{t_j}^{t_{j+1}} \alpha(v + t_{j+1} - s, x - t_{j+1} + s)\, ds \\
&\qquad + \int_{t_j}^{t_{j+1}} \boldsymbol{\sigma}(v + t_{j+1} - s, x - t_{j+1} + s)\, dW_s\, dv - \int_{\tau+t_{j+1}-t_j}^{\tau+l+t_{j+1}-t_j} \mu_{t_j}(v, x - t_{j+1} + t_j)\, dv \\
&= \int_{t_j}^{t_{j+1}} \int_\tau^{\tau+l} \alpha(v + t_{j+1} - s, x - t_{j+1} + s)\, dv\, ds \\
&\qquad + \int_{t_j}^{t_{j+1}} \int_\tau^{\tau+l} \boldsymbol{\sigma}(v + t_{j+1} - s, x - t_{j+1} + s)\, dv\, dW_s \\
&\stackrel{d}{=} \int_0^{t_{j+1}-t_j} \int_\tau^{\tau+l} \alpha(v + t_{j+1} - t_j - s, x - (t_{j+1} - t_j) + s)\, dv\, ds \\
&\qquad + \int_0^{t_{j+1}-t_j} \int_\tau^{\tau+l} \boldsymbol{\sigma}(v + t_{j+1} - t_j - s, x - (t_{j+1} - t_j) + s)\, dv\, dW_s,
\end{aligned}
$$

so that $\boldsymbol{F}_l(t_j, t_{j+1})$ is independent for different j (due to the Lévy property of Brownian motion), Gaussian (since α and $\boldsymbol{\sigma}$ are deterministic functions), and clearly i.i.d. if $t_{j+1} - t_j = \Delta$.

$\square$

Proof. Proof of Proposition 3.1

Proceeding similarly as in the proof of Prop. 2.1 above,

$$
F^j(\tau, x) = \int_{t_j}^{t_{j+1}} \int_0^1 \alpha(v + \tau + t_{j+1} - s, x - t_{j+1} + s)\, dv\, ds
$$

$$
+ \int_{t_j}^{t_{j+1}} \int_0^1 \boldsymbol{\sigma}(v + \tau + t_{j+1} - s, x - t_{j+1} + s)\, dv\, dW_s.
$$

Therefore, with Equation (6), by a simple application of Itō's product formula, $F^j(\tau, x)$ is Normal distributed with expected value:

$$
\mathbb{E}\left[F^j(\tau, x)\right] = \int_{t_j}^{t_{j+1}} \int_0^1 \boldsymbol{\sigma}(v + \tau + t_{j+1} - s, x - t_{j+1} + s) \int_0^{v+\tau+t_{j+1}-s} \boldsymbol{\sigma}'(u, x - t_{j+1} + s)\, du\, dv\, ds
$$

and covariance structure:

$$
\mathrm{Cov}\left[F^j(\tau_1, x_1), F^k(\tau_2, x_2)\right] = \delta_{jk} \times
$$

$$
\int_{t_j}^{t_{j+1}} \int_0^1 \boldsymbol{\sigma}(v + \tau_1 + t_{j+1} - s, x_1 - t_{j+1} + s)\, dv \int_0^1 \boldsymbol{\sigma}'(v + \tau_2 + t_{j+1} - s, x_2 - t_{j+1} + s)\, dv\, ds,
$$

in which δ_{jk} equals 1 if $j = k$ and 0 otherwise. The covariance matrix $\Sigma_{(\tau_i, x_i), (\tau_j, x_j)}$ is then immediately obtained by changing variables in the integrals.

For the expected value, note that $t_{j+1} = t_j + 1$ and:

$$
\boldsymbol{\sigma}(v + \tau + t_{j+1} - s, x - t_{j+1} + s) \int_0^{v+\tau+t_{j+1}-s} \boldsymbol{\sigma}'(u, x - t_{j+1} + s)\, du
$$

$$
= \frac{1}{2} \frac{\partial}{\partial v} \left(\int_0^{v+\tau+t_{j+1}-s} \boldsymbol{\sigma}(u, x - t_{j+1} + s)\, du \int_0^{v+\tau+t_{j+1}-s} \boldsymbol{\sigma}'(u, x - t_{j+1} + s)\, du \right),
$$

so that:

$$
\mathbb{E}\left[F^j(\tau, x)\right] = \int_{t_j}^{t_{j+1}} \int_0^1 \boldsymbol{\sigma}(v + \tau + t_{j+1} - s, x - t_{j+1} + s) \int_0^{v+\tau+t_{j+1}-s} \boldsymbol{\sigma}'(u, x - t_{j+1} + s)\, du\, dv\, ds
$$

$$
= \frac{1}{2} \left(\int_{t_j}^{t_{j+1}} \left(\int_0^{1+\tau+t_{j+1}-s} \boldsymbol{\sigma}(u, x - t_{j+1} + s)\, du \int_0^{1+\tau+t_{j+1}-s} \boldsymbol{\sigma}'(u, x - t_{j+1} + s)\, du \right) \right.
$$

$$
\left. - \left(\int_0^{\tau+t_{j+1}-s} \boldsymbol{\sigma}(u, x - t_{j+1} + s)\, du \int_0^{\tau+t_{j+1}-s} \boldsymbol{\sigma}'(u, x - t_{j+1} + s)\, du \right) ds \right)
$$

$$
= \frac{1}{2} \left(\int_0^1 \left(\int_0^{2+\tau-s} \boldsymbol{\sigma}(u, x - 1 + s)\, du \int_0^{2+\tau-s} \boldsymbol{\sigma}'(u, x - 1 + s)\, du \right) \right.
$$

$$
\left. - \left(\int_0^{1+\tau-s} \boldsymbol{\sigma}(u, x - 1 + s)\, du \int_0^{1+\tau-s} \boldsymbol{\sigma}'(u, x - 1 + s)\, du \right) ds \right)
$$

$$= \int_0^1 \frac{1}{2} \left(\int_{1+\tau-s}^{2+\tau-s} \boldsymbol{\sigma}(u, x-1+s) \, du \int_{1+\tau-s}^{2+\tau-s} \boldsymbol{\sigma}'(u, x-1+s) \, du \right)$$
$$+ \left(\int_{1+\tau-s}^{2+\tau-s} \boldsymbol{\sigma}(u, x-1+s) \, du \int_0^{1+\tau-s} \boldsymbol{\sigma}'(u, x-1+s) \, du \right) \, ds,$$

where the last step follows from writing the integrals in the first term as $\int_0^{2+\tau-s} = \int_0^{1+\tau-s} + \int_{1+\tau-s}^{2+\tau-s}$.

$\square$

Proof. Proof of Proposition 4.1

Inserting $\tau = 0$ in Equation (8) immediately yields (13). Further define $Y_t = Z_t \exp\{-\boldsymbol{H} t\} = \int_0^t \exp\{-\boldsymbol{H} s\} \boldsymbol{J} \, dW_s$. Apply Itō's formula on $Z_t = \exp\{\boldsymbol{H} t\} Y_t$, we obtain:

$$dZ_t = \boldsymbol{H} \exp\{\boldsymbol{H} t\} Y_t \, dt + \exp\{\boldsymbol{H} t\} \, dY_t$$
$$= \boldsymbol{H} \underbrace{\exp\{\boldsymbol{H} t\} Y_t}_{Z_t} \, dt + \exp\{\boldsymbol{H} t\} \exp\{-\boldsymbol{H} t\} \boldsymbol{J} \, dW_t$$
$$= \boldsymbol{H} Z_t \, dt + \boldsymbol{J} \, dW_t.$$

The dynamics for Z_t can therefore be expressed with our choices for $\boldsymbol{H}$ and $\boldsymbol{J}$.

$\square$

B   Non-parametric and Factor Mortality Forecasting Models

Algorithm B.1. *Non-parametric mortality forecasting:*
For $i = 1, 2, \ldots, I$:

- *Starting from the mortality forecast at time $t_N = T$, $\{_\tau p_x^{(i)}(T)\} = \{_\tau p_x(T)\}$.*

- *For $t = 1, 2, \ldots, y$, sample $\boldsymbol{F} = (F(\tau, x))$ (with replacement) from $\{\boldsymbol{F}^1, \boldsymbol{F}^2, \ldots, \boldsymbol{F}^N\}$, and for each x set:*

$$_1 p_x^{(i)}(T+t) = \exp\{-F(0, x)\} \times \frac{_2 p_{x-1}^{(i)}(T+t-1)}{_1 p_{x-1}^{(i)}(T+t-1)}, \tag{18}$$

$$_2 p_x^{(i)}(T+t) = {_1 p_x^{(i)}}(T+t) \times \exp\{-F(1, x)\} \times \frac{_3 p_{x-1}^{(i)}(T+t-1)}{_2 p_{x-1}^{(i)}(T+t-1)},$$

$$\cdots \qquad \cdots$$

$$_\tau p_x^{(i)}(T+t) = {_{\tau-1} p_x^{(i)}}(T+t) \times \exp\{-F(\tau-1, x)\} \times \frac{_{\tau+1} p_{x-1}^{(i)}(T+t-1)}{_\tau p_{x-1}^{(i)}(T+t-1)},$$

$$\cdots \qquad \cdots$$

Under the assumptions of Proposition 2.1, the vectors $\{\boldsymbol{F}^1, \boldsymbol{F}^2, \ldots, \boldsymbol{F}^N\}$ are i.i.d. so we can generate I y-year-ahead forecasts by simply sampling with replacement (non-parametric bootstrap), as outlined in Algorithm B.1. We note that in this algorithm, we do not directly rely on the assumption of a Gaussian evolution in Equation (2). In particular, the algorithm is a simple way of producing mortality forecasts as long as the i.i.d. assumption (roughly) holds in the data. We can also exploit Gaussianity to obtain parametric bootstrap forecasts. More precisely, we can calculate the sample mean $\bar{\boldsymbol{F}}^{\cdot}$ and covariance matrix $\hat{\boldsymbol{\Sigma}}$ of $\{\boldsymbol{F}^1, \boldsymbol{F}^2, \ldots, \boldsymbol{F}^N\}$, so that we can simulate the random vector $\boldsymbol{F}$ in Algorithm B.1 by sampling from the corresponding multivariate Normal distribution.

We rely on an analogous approach for the statistical factor forecasting model. As the key modification, in the second step in Algorithm B.1, we do not sample $\boldsymbol{F}$ via bootstrapping, but rather rely on Equation (4) in the main text. More precisely, we simulate univariate standard Normal variables ε_ν, $\nu = 1, \ldots, D$, to generate $\boldsymbol{F}$ in each step.

C   Analyses of Hazard Models

C.1   Gaussian Gompertz Model

According to Equation (16), $(\tilde{Z}_t)$ follows a two-dimensional Ornstein–Uhlenbeck process. Therefore, we have:

$$\tilde{Z}_{t+\tau} = \exp\{\tilde{\boldsymbol{H}}\tau\} \cdot \tilde{Z}_t + \int_t^{t+\tau} \exp\{\tilde{\boldsymbol{H}}(t + \tau - s)\} \cdot \boldsymbol{J} \, dW_s.$$

That is, with known $\tilde{Z}_t$, $\tilde{Z}_{t+\tau}$ follows a multivariate Normal distribution with $\mathbb{E}(\tilde{Z}_{t+\tau}) = \exp\{\tilde{\boldsymbol{H}}\tau\}\tilde{Z}_t$ and $\mathrm{Cov}(\tilde{Z}_{t+\tau}) = \int_t^{t+\tau} \exp\{\tilde{\boldsymbol{H}}(t + \tau - s)\}\boldsymbol{J}\boldsymbol{J}' \exp\{\tilde{\boldsymbol{H}}(t + \tau - s)\}' \, ds$.

Furthermore, $(\tilde{Z}_t)$ is an affine process (Duffie et al., 2000, 2003), so that we obtain for the survival probabilities (Biffis, 2005):

$$
\begin{aligned}
{}_{T-t}p_{x+t}(t) &= \mathbb{E}\left[\exp\left\{-\int_t^T \mu_s(0, x + s)ds\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{-\int_t^T (k + c\exp(r(x + s)))\,\tilde{Z}_s^{(2)}\,ds\right\}\right] \\
&= \exp\left\{A(t, T, x) + B(t, T, x)\,\tilde{Z}_t\right\}.
\end{aligned}
$$

where $A = A(\cdot, T, x)$ and $B = (B^{(1)}, B^{(2)})' = B(\cdot, T, x)$ satisfy the following Riccati ordinary

differential equations:

$$A_t = -\frac{1}{2}(1-ab)^2(B^{(1)})^2 - a(1-ab)B^{(1)}B^{(2)} - \frac{1}{2}a^2(B^{(2)})^2, \tag{19}$$

$$B_t^{(1)} = 2bB^{(1)} - B^{(2)}, \tag{20}$$

$$B_t^{(2)} = (k + c\exp(r(x+t))) + b^2 B^{(1)} - d\,B^{(2)}, \tag{21}$$

with terminal constraints:

$$A(T,T,x) = 0, \quad B(T,T,x) = 0.$$

C.2   Gamma Gompertz Model

Iterations of $\tilde{Z}_t$.

An iterative application of (17) immediately results in:

$$\begin{aligned}
\tilde{Z}_{t+\tau} &= (\tilde{\boldsymbol{H}} + \boldsymbol{I})\,\tilde{Z}_{t+\tau-1} - \theta\,\boldsymbol{J} + \boldsymbol{J}\,\gamma_{t+\tau} \\
&= \cdots \\
&= \sum_{j=1}^{\tau}(\tilde{\boldsymbol{H}} + \boldsymbol{I})^{\tau-j}\,\boldsymbol{J}\,\gamma_{t+j} - \theta\sum_{j=1}^{\tau}(\tilde{\boldsymbol{H}} + \boldsymbol{I})^{\tau-j}\,\boldsymbol{J} + (\tilde{\boldsymbol{H}} + \boldsymbol{I})^{\tau}\,\tilde{Z}_t.
\end{aligned}$$

Since the expressions of hazard rate only depend on $\tilde{Z}_t^{(2)}$, we specifically consider:

$$\tilde{Z}_{t+\tau}^{(2)} = \sum_{j=1}^{\tau}((\tilde{\boldsymbol{H}} + \boldsymbol{I})^{\tau-j}\boldsymbol{J})^{(2)}\,\gamma_{t+j} - \theta\sum_{j=1}^{\tau}((\tilde{\boldsymbol{H}} + \boldsymbol{I})^{\tau-j}\boldsymbol{J})^{(2)} + ((\tilde{\boldsymbol{H}} + \boldsymbol{I})^{\tau})^{(2)}\,\tilde{Z}_t. \tag{22}$$

Distribution of $\tilde{Z}_{t+\tau}^{(2)}$.

Since each γ_{t+j} is i.i.d. Gamma distributed $\sim \Gamma(\theta^2, 1/\theta)$, we immediately have $((\tilde{\boldsymbol{H}}+\boldsymbol{I})^{\tau-j}\boldsymbol{J})^{(2)}\,\gamma_{t+j} \sim \Gamma(\theta^2, ((\tilde{\boldsymbol{H}}+\boldsymbol{I})^{\tau-j}\boldsymbol{J})^{(2)}/\theta)$. With different j, the scale parameter $(((\tilde{\boldsymbol{H}}+\boldsymbol{I})^{\tau-j}\boldsymbol{J})^{(2)}/\theta)$ varies, so the summation as in the first term on the right-hand side of (22) cannot be simply obtained as another Gamma distributed random variable.

That said, this topic has been covered in the existing literature. For example, Theorem 1 in Moschopoulos (1985) provides the exact density function of the summation of independent Gamma random variables. Alternatively, the summation itself can also be approximated as a Gamma random variable based on the Welch-Satterthwaite equation, when we view the Gamma distribution as a scaled chi-square distribution (Box, 1954). In particular, with component shapes

k_i and scales θ_i, the shape and scale of the summation can be expressed as:

$$k_{\text{sum}} = \frac{(\sum_i \theta_i k_i)^2}{\sum_i \theta_i^2 k_i}, \tag{23}$$

and

$$\theta_{\text{sum}} = \frac{\sum_i \theta_i k_i}{k_{\text{sum}}}. \tag{24}$$

We rely on the approximation in our calculations, and we verified via simulation analyses that it is quite accurate in our applications (the approximation error is negligible).

Closed-form expression of $_\tau p_x(t)$.

For one-year survival probabilities, we have $p_x(t) = \exp(-(k + c\exp(rx)) \times \tilde{Z}_t^{(2)})$. Therefore, the τ-year realized survival probability is:

$$\prod_{i=0}^{\tau-1} p_{x+i}(t+i) = \exp\left(\sum_{i=0}^{\tau-1} -(k + c\exp(r(x+i))) \times \tilde{Z}_{t+i}^{(2)}\right).$$

Plugging in Equation (22) and rearranging terms, we obtain:

$$\prod_{i=0}^{\tau-1} p_{x+i}(t+i) = \exp\left\{-\sum_{i=0}^{\tau-1}(k + c\exp(r(x+i)))\left[((\tilde{H}+I)^i)^{(2)}\tilde{Z}_t - \theta\sum_{j=1}^{i}((\tilde{H}+I)^{i-j}J)^{(2)}\right]\right.$$

$$\left. - \underbrace{\sum_{i=0}^{\tau-1}(k + c\exp(r(x+i)))\left(\sum_{j=1}^{i}((\tilde{H}+I)^{i-j}J)^{(2)}\gamma_{t+j}\right)}_{\gamma_{\text{sum}}}\right\}. \tag{25}$$

The random variable γ_{sum} can be rewritten as:

$$\gamma_{\text{sum}} = \sum_{i=1}^{\tau-1}\gamma_{t+i} \times \left(\sum_{j=1}^{\tau-i}(k + c\exp(r(x+j))) \times ((\tilde{H}+I)^{j-1}J)^{(2)}\right).$$

Since each term within the summation follows a Gamma distribution, that is:

$$\gamma_{t+i} \times \left(\sum_{j=1}^{\tau-i}(k + c\exp(r(x+j))) \times ((\tilde{H}+I)^{j-1}J)^{(2)}\right) \sim \Gamma(\underbrace{\theta^2}_{k_i}, \underbrace{\sum_{j=1}^{\tau-i}(k + c\exp(r(x+j))) \times ((\tilde{H}+I)^{j-1}J)^{(2)}/\theta}_{\theta_i}),$$

their summation, γ_{sum}, can once again be approximately Gamma via (23) and (24). For simplicity, we refer to its shape and scale parameters as k_{sum} and θ_{sum}.

From (25), we can then calculate the expectation of the realized survival probability, $_\tau p_x(t)$:

$$_\tau p_x(t) \;=\; \exp\left\{-\sum_{i=0}^{\tau-1}(k+c\exp(r(x+i)))\left[((\tilde{\boldsymbol{H}}+\boldsymbol{I})^i)^{(2)}\tilde{Z}_t-\theta\sum_{j=1}^{i}((\tilde{\boldsymbol{H}}+\boldsymbol{I})^{i-j}\boldsymbol{J})^{(2)}\right]\right\}$$
$$\times\mathbb{E}\left[\exp(-\gamma_{\text{sum}})\right].$$

With γ_{sum} Gamma distributed, the random variable $\exp(-\gamma_{\text{sum}})$ follows a so-called negative log-Gamma distribution, with expected value:

$$\mathbb{E}[\exp(-\gamma_{\text{sum}})]=\frac{1}{(1+\theta_{\text{sum}})^{k_{\text{sum}}}}.$$

Therefore, we obtain (approximate) survival probabilities in closed-form:

$$_\tau p_x(t)=\frac{1}{(1+\theta_{\text{sum}})^{k_{\text{sum}}}}\times\exp\left\{-\sum_{i=0}^{\tau-1}(k+c\exp(r(x+i)))\left[((\tilde{\boldsymbol{H}}+\boldsymbol{I})^i)^{(2)}\tilde{Z}_t-\theta\sum_{j=1}^{i}((\tilde{\boldsymbol{H}}+\boldsymbol{I})^{i-j}\boldsymbol{J})^{(2)}\right]\right\}.$$
$$(26)$$

Relying on the survival probabilities expressed in (26), we can then determine life expectancies and annuity present values.

Impact of θ.

To illustrate the impact of the choice of θ, Figure A.1 provides box-whisker plots of future life expectancies for 70-year-old U.S. females $\overset{\circ}{e}_{70}\,(t_0+\tau)$ for $\tau=1$ (left-hand panel) and $\tau=10$ (right-hand panel) for different choices of θ in the Gamma Gompertz model. As indicated in the main text, if θ is larger, the resulting Gamma distribution is more similar to a Normal distribution, so that the Gamma Gompertz model performs very similarly to the Gaussian model, even for the short time horizon $\tau=1$. For longer time horizons, due to the aggregation of the annual innovations, the results are very similar even for the smaller choice of θ. As indicated in the main text, we choose $\theta=1$, which performs well and avoids negative mortality across all our calculations.

D   Model Extensions

D.1   Extension to Multi-Factor Forward Mortality Models

In the main text, we limit our consideration to models with a single source of randomness. This section briefly describes that we can follow a similar procedure for devising multi-factor mortality

(a) $\overset{\circ}{e}_{70}\,(t_0 + 1)$ (b) $\overset{\circ}{e}_{70}\,(t_0 + 10)$
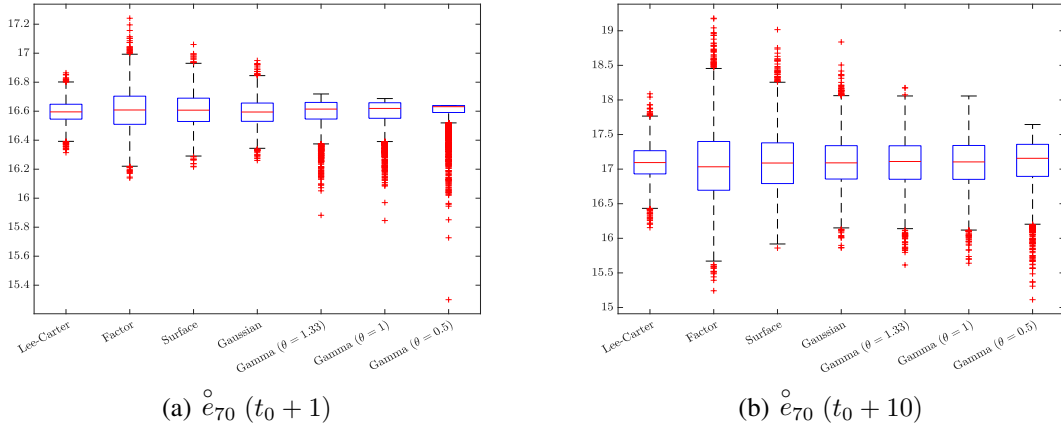
Figure A.1: Showing realizations of expected future life-time for age-70 in 1 year (left panel) and age-70 in 10 years (right panel), $t_0 = 2016$. Left-to-right: Lee-Carter predictions; one-factor model (9); mortality-surface model (w\ self-consistency); Gaussian Gompertz-hazard model; Gamma Gompertz-hazard model with $\theta = 1.33$, $\theta = 1.0$, and $\theta = 0.5$.

surface models. The foundation is the following proposition, which is similar to a result in Angelini and Herzel (2005) for interest rate modeling:

Proposition D.1. *Let* $\boldsymbol{\sigma}(\tau, x) = (\sigma_1(\tau, x), \ldots, \sigma_D(\tau, x))$, *where each function* $\sigma_i(\tau, x)$ *is of the form:*

$$\sigma_i(\tau, x) = \boldsymbol{C}_i(x + \tau) \times \exp\left\{\boldsymbol{H}_i\,\tau\right\} \times \boldsymbol{J}_i, \tag{27}$$

$\boldsymbol{C}_i(\cdot) \in \mathbb{R}^{1 \times m_i}$, $\boldsymbol{H}_i \in \mathbb{R}^{m_i \times m_i}$, $\boldsymbol{J}_i = \mathbb{R}^{m_i \times 1}$, $m_i \in \mathbb{N}$, $i = \{1, 2, \ldots, D\}$. *Then* $\boldsymbol{\sigma}(\tau, x)$ *is also of the form as in Equation (7), i.e. the model allows for a Gaussian finite-dimensional realization, where* $\boldsymbol{C}(x) = [\boldsymbol{C}_1(x), \ldots, \boldsymbol{C}_D(x)]$, $\boldsymbol{H} = \text{diag}\left\{\boldsymbol{H}_1, \ldots, \boldsymbol{H}_D\right\}$, *and* $\boldsymbol{J} = \text{diag}\left\{\boldsymbol{J}_1, \ldots, \boldsymbol{J}_D\right\}$.

Proof. Proof.

Stacking the σ_i's in a vector of dimension D, we obtain:

$$[\sigma_1(\tau, x), \ldots, \sigma_D(\tau, x)]$$
$$= [\boldsymbol{C}_1(x + \tau) \times \exp\left\{\boldsymbol{H}_1\,\tau\right\} \times \boldsymbol{J}_1, \ldots, \boldsymbol{C}_D(x + \tau) \times \exp\left\{\boldsymbol{H}_D\,\tau\right\} \times \boldsymbol{J}_D]$$
$$= \underbrace{[\boldsymbol{C}_1(x + \tau), \ldots, \boldsymbol{C}_D(x + \tau)]}_{\boldsymbol{C}(x+\tau)} \times \begin{pmatrix} \exp\{\boldsymbol{H}_1\,\tau\} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \exp\{\boldsymbol{H}_2\,\tau\} & & \boldsymbol{0} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \exp\{\boldsymbol{H}_D\,\tau\} \end{pmatrix} \times \underbrace{\begin{pmatrix} \boldsymbol{J}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J}_2 & & \boldsymbol{0} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{J}_D \end{pmatrix}}_{\boldsymbol{J}}$$
$$= \boldsymbol{C}(x + \tau) \times \underbrace{\exp\left\{\text{diag}\left\{\boldsymbol{H}_1, \ldots, \boldsymbol{H}_D\right\}\,\tau\right\}}_{\exp\{\boldsymbol{H}\tau\}} \times \boldsymbol{J}.$$

□

By relying on this representation, similarly to Equation (10), we can express $\boldsymbol{F}$ via:

$$
\begin{aligned}
F^j(\tau, x) &= \mathbb{E}[F(\tau, x)] + \sum_{i=1}^{D} \int_{\tau}^{\tau+1} \boldsymbol{C}_i(x+\nu) \, \exp\{\boldsymbol{H}_i \, \nu\} \, d\nu \int_{t_j}^{t_{j+1}} \exp\{\boldsymbol{H}_i \, (t_{j+1}-s)\} \, \boldsymbol{J}_i \, dW_s^{(i)} \\
&\approx \mathbb{E}[F(\tau, x)] + \sum_{i=1}^{D} \int_{\tau}^{\tau+1} \boldsymbol{C}_i(x+\nu) \, \exp\{\boldsymbol{H}_i \, (\nu+1)\} \, d\nu \, \boldsymbol{J}_i \, (W_{t_{j+1}}^{(i)} - W_{t_j}^{(i)}),
\end{aligned}
$$

where $(W_{t_{j+1}}^{(i)} - W_{t_j}^{(i)})$, $i = \{1, 2, \ldots, D\}$, are standard Normal and correspond to the ith component of the driving Brownian motion—in direct analogy to the factor model (4) in the main text. Hence, Proposition D.1 allows us to proceed analogously to a single-factor model, particularly separately for each independent factor.

D.2   Model with Catastrophe Component

Our focus in this paper is on longevity risk, although it is possible to extend the model to incorporate shocks originating from global pandemics and/or mortality catastrophes. Following Bauer and Kramer (2016), we can generalize the Gompertz model for the hazard ((15)/(16) in the main text) by adding a mean-reverting jump process describing such extremal scenarios:

$$
\mu_t(0, x) = (k + c \exp(rx)) \times \tilde{Z}_t^{(2)} + \Gamma_t, \tag{28}
$$

where

$$
d\Gamma_t = -\kappa \Gamma_t \, dt + dJ_t
$$

and (J_t) is a compound Poisson process with positive jump size distribution. Importantly, the extended model (28) remains in the class of exponentially-affine processes (Duffie et al., 2000), so tractability of the model will be preserved: For the calculation of survival probabilities, it is sufficient to solve simple ordinary differential equations, similarly to Part C.1 above.

Other extensions are also possible. For instance, while in the paper we treat sample data with different genders and from different countries separately in the principal component analysis, an alternative approach could be the use of common principal component, with which we could deal with multiple populations in one single analysis.

(a) Time series ACF and PACF   (b) $\overset{\circ}{e}_{70}\ (t_0 + 10)$
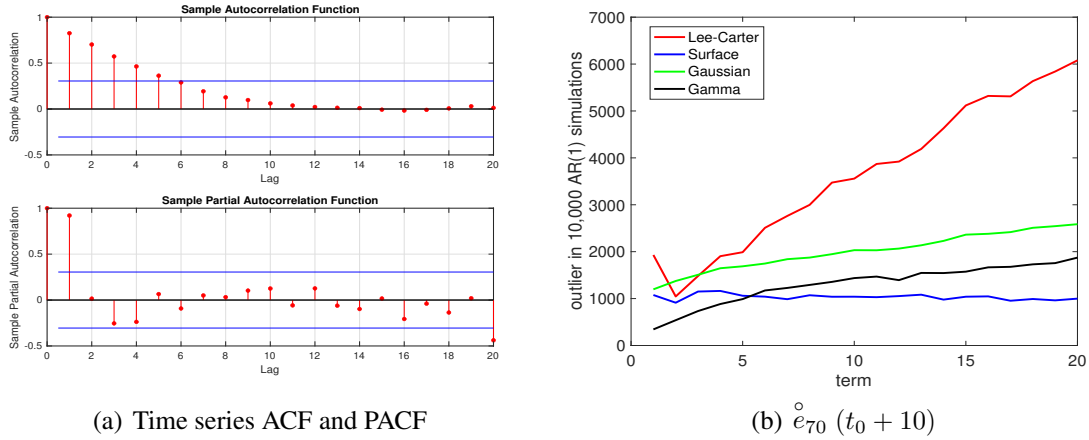
Figure A.2: Showing results for ARIMA time series models fit to cohort life expectancies for 70-year old U.S. females as forecast based on the Lee-Carter model (1974-2016). Left-hand side: Autocorrelation (ACF) and partial autocorrelation function (PACF) of the time series. Right-hand side: Counts how many of 10,000 simulated AR(1) paths are outside of the prediction intervals from Figure 1.

E   Analyses of Lee-Carter Prediction Intervals

To demonstrate that the Lee-Carter prediction intervals for cohort life expectancy are *too narrow* from a historical perspective, we carry out two separate analyses.

For the first, we proceed similarly to Lee and Carter (1992, p. 667) and fit univariate autoregressive moving average (ARIMA) models to our time series of cohort life expectancies for seventy-year old females illustrated in Figure 1 in the main text (1974-2016). We then use the resulting time series models to simulate 10,000 cohort life expectancy trajectories for the subsequent twenty years, and record—in each year—how many paths are outside the Lee-Carter prediction intervals (the red-dashed curves in Figure 1).

For selecting suitable ARIMA(p,q) models, we rely on the Box-Jenkins methodology. Figure A.2(a) shows plots of the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The ACF is exponentially decaying to zero, pointing to an AR process ($q = 0$). The PACF exhibits a spike at lag 1 but none beyond, suggesting $p = 1$. Thus, we fit an AR(1) process. The red curve in Panel (b) of Figure A.2 then records how many out of the 10,000 simulated paths are outside of the Lee-Carter prediction intervals each year over the next twenty years.

As is evident, the paths cross the 95% bands far more frequently than predicted. In fact, for terms beyond 15 years, the majority of the paths is outside of the Lee-Carter prediction intervals. Figure A.2(b) also shows results for our models with long-term risk. Here, the frequencies are more stable and more modest (e.g., $\approx 10\%$ of paths for the Surface model across terms), although they still exceed the 5% confidence level. Of course this result depends on the underlying time

(a) Prediction interval at $t_0 - 20$ years (1996)   (b) Prediction interval at $t_0 - 10$ years (2006)
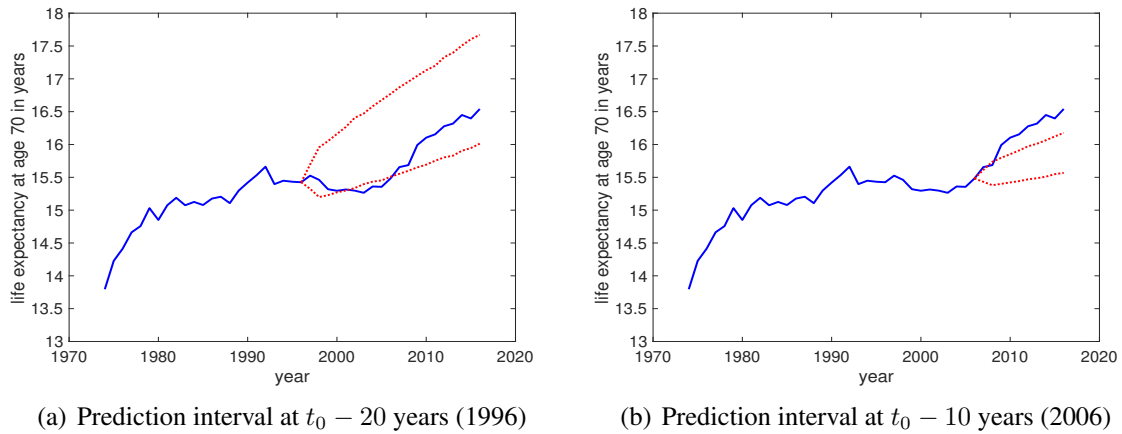
Figure A.3:   Showing cohort life expectancies for 70-year old U.S. females based on the Lee-Carter model until 2016 (blue solid line) and 95% prediction intervals (red dashed lines) based on past data until 1996 (left-hand side) and until 2006 (right-hand side).

series model, and different choices yield different result. For instance, for a simple exponential smoothing model with growth (ARIMA(0,1,1)), simulated trajectories are within the prediction bands based on our models with long-term risk for about 97% of the paths. However, here also (as within all of our experiments), far more than 5% of paths fall outside of the Lee-Carter prediction intervals ($\approx 10\%$ for ARIMA(0,1,1), which was the lowest figure in our tests).

To present an analysis that does not depend on the (time series) model, we also carry out backtests. More precisely, we carry out the exact same analyses as for the Lee-Carter intervals in Figure 1 but calculate prediction intervals based on past historical data until 2006 (10 years before the end of our series) and 1996 (20 years before the end of our series). Figure A.3 show the results, where the solid blue line is exactly the same as in Figure 1 before 2016. Strikingly, the historical paths fall outside the prediction intervals for both periods, and for the 2006 interval for most of the subsequent years. The paths are within the prediction intervals for the models with long-term risk proposed in this paper for both time periods. This provides further evidence that the Lee-Carter prediction intervals for cohort life expectancy are *too narrow* from a historical perspective.

F   Analysis on Alternative Data Sets

In this part, we repeat the analyses from the main text for alternative data sets/sources. In particular, we use (i) U.K. pensioner life tables (as published by the *Institute and Faculty of Actuaries* (IFoA)[1]) and (ii) mortality forecasts generated based on combinations of gender (male, female), region

---

[1]See the website of the *Continuous Mortality Investigation* (CMI), `www.actuaries.org.uk/learn-and-develop/continuous-mortality-investigation`.

(England & Wales, France, Japan, U.S., and West Germany) from the *Human Mortality Database*,[2] and mortality forecasting method (Lee-Carter, CBD, and P-splines). Our findings are generally analogous to the main text, which is the reason that we only show results for our primary data set there. Thus, we omit detailed discussions and only highlight notable differences here.

Figures A.4 through A.13 as well as Tables A.1 through A.5 show results from the factor analysis/model of F (cf. Section 3.2). More precisely, the figures show the first and second eigenvectors for each of the five regions (England & Wales, France, Japan, U.S., and West Germany) with all gender/forecasting method combinations. The tables show the four greatest eigenvalues (factor loadings) for each of the same five regions and the same gender/forecasting method combinations. Once again, we observe that the leading eigenvector explains the vast majority of the variation in the data, akin to Table 1 in the main text. As is clear, the leading eigenvectors (*slope factor*) in all cases exhibit similar shapes to Figure 2 in the main text (and hence the same arguments apply).

Figures A.4 through A.13 also illustrate the second eigenvector. As pointed out in the main text, the shape for the second factor (eigenvector) u_2 is drastically different than that of the leading factor—but again it is similar/systematic across all considered sets. While we again observe a mostly increasing pattern as a function of age x in the near future (small τ), the factor entries change signs when viewed as a function of τ (for fixed x). Hence, for a young to middle-aged individual (indexed by x), a random shock may affect mortality in the near term qualitatively differently than in the long term, implying that the *curvature* of the individual mortality curves will change. Again in analogy to yield curve forecasting, we refer to this factor as the *curvature factor*.

Tables A.6 through A.13 show maximum likelihood estimation results of our mortality surface model based on the remaining representative data sets/sources in Table 1, in analogy to Table 2 in Section 3.3 in the main text. No standard errors are reported for the U.K. pensioner life tables in Table A.13, where the limited sample size ($N = 6$) makes it impractical to sample F using the non-parametric bootstrap. The findings for the different genders/regions/forecasting approaches under source (ii) in Tables A.6 through A.12 are similar as in the main text: The parameters capture the patterns in the leading eigenvectors, and we observe similar estimates with and without the self-consistency constraint across all cases. We find more substantial differences in the point estimates for the U.K. pensioner data in Table A.13. Potential reasons include changes in the relevant (insured) population and/or the underlying forecasting method, although it is impossible to obtain conclusive insights based on the sample size ($N = 6$).

Figures A.14 through A.21 show prediction intervals of the expected future life-time for age-40 and 70 cohorts one year from the terminal year using different approaches and the same represen-

---

[2]Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at `www.mortality.org` or `www.humanmortality.de`.

| Factor | Female | | Male | |
	Value	Weight	Value	Weight
Lee-Carter				
λ_1	0.0415	94.84%	0.0504	87.25%
λ_2	0.0010	2.20%	0.0033	5.74%
λ_3	0.0005	1.18%	0.0019	3.30%
λ_4	0.0004	0.91%	0.0010	1.81%
CBD				
λ_1	0.1410	97.20%	0.5089	96.24%
λ_2	0.0038	2.65%	0.0170	3.22%
λ_3	0.0002	0.12%	0.0027	0.51%
λ_4	0.0000	0.03%	0.0001	0.03%
P-spline				
λ_1	0.1979	99.32%	0.2009	98.79%
λ_2	0.0013	0.63%	0.0021	1.01%
λ_3	0.0001	0.03%	0.0003	0.16%
λ_4	0.0000	0.02%	0.0001	0.04%

Table A.1: Showing the absolute values and relative weights of the four greatest eigenvalues based on standard principal component analysis on the England & Wales population (left: female; right: male) with different forecasting methods (top: Lee-Carter; middle: CBD; bottom: P-spline).

tative data sets, paralleling Figure 4 in the main text and Figure A.1 above. Here the presentation deviates in a few instances. First, it is difficult to generate random paths of *age-specific* mortality rates using the P-spline method. We therefore only show results based on the remaining four approaches in Figures A.15 and A.18. Similarly, for the U.K. pensioner data (Figure A.21), we only show three approaches due to no documented underlying model and limited sample size that renders the non-parametric approach impractical. Aside from these differences in presentation, the findings parallel the main text: The prediction intervals based on the forward models proposed in this paper closely resemble each other, and are far wider than the conventional forecasts where available. The sole exception is the "w\ s.c." approach for the U.K. pensioner data (Figure A.21), which is again due to potential changes in population or approach, and/or due to data limitations. We leave the consideration of a different, potentially richer set of life tables and/or price data for life-contingent securities (Source (iii) in Sec. 3.1 in the main text) for future research.

	Factor	Female		Male	
		Value	Weight	Value	Weight
Lee-Carter					
	λ_1	0.0383	96.03%	0.0763	91.28%
	λ_2	0.0008	2.09%	0.0033	3.99%
	λ_3	0.0003	0.77%	0.0016	1.88%
	λ_4	0.0002	0.55%	0.0011	1.26%
CBD					
	λ_1	0.0741	96.81%	0.1538	95.45%
	λ_2	0.0022	2.85%	0.0050	3.07%
	λ_3	0.0002	0.28%	0.0023	1.41%
	λ_4	0.0000	0.05%	0.0001	0.01%
P-spline					
	λ_1	0.1191	99.49%	0.1237	99.18%
	λ_2	0.0005	0.42%	0.0006	0.48%
	λ_3	0.0001	0.06%	0.0004	0.30%
	λ_4	0.0000	0.02%	0.0000	0.03%

Table A.2: Showing the absolute values and relative weights of the four greatest eigenvalues based on standard principal component analysis on the France population (left: female; right: male) with different forecasting methods (top: Lee-Carter; middle: CBD; bottom: P-spline).

Factor	Female		Male	
	Value	Weight	Value	Weight
Lee-Carter				
λ_1	0.0197	92.27%	0.0384	91.31%
λ_2	0.0006	2.66%	0.0017	4.08%
λ_3	0.0004	1.88%	0.0009	2.18%
λ_4	0.0004	1.73%	0.0005	1.23%
CBD				
λ_1	0.0838	96.30%	0.2988	97.21%
λ_2	0.0019	3.38%	0.0079	2.56%
λ_3	0.0002	0.25%	0.0005	0.17%
λ_4	0.0001	0.06%	0.0002	0.06%
P-spline				
λ_1	0.0501	98.64%	0.1018	99.16%
λ_2	0.0006	1.17%	0.0006	0.62%
λ_3	0.0001	0.15%	0.0002	0.16%
λ_4	0.0000	0.02%	0.0000	0.04%

Table A.3: Showing the absolute values and relative weights of the four greatest eigenvalues based on standard principal component analysis on the Japan population (left: female; right: male) with different forecasting methods (top: Lee-Carter; middle: CBD; bottom: P-spline).

| Factor | Female | | Male | |
	Value	Weight	Value	Weight
Lee-Carter				
λ_1	0.0241	94.47%	0.0401	90.36%
λ_2	0.0009	3.44%	0.0023	5.17%
λ_3	0.0002	0.94%	0.0011	2.37%
λ_4	0.0001	0.45%	0.0005	1.20%
CBD				
λ_1	0.0297	98.54%	0.0598	98.23%
λ_2	0.0004	1.19%	0.0007	1.15%
λ_3	0.0001	0.24%	0.0003	0.56%
λ_4	0.0000	0.02%	0.0000	0.05%
P-spline				
λ_1	0.1481	98.22%	0.0830	97.53%
λ_2	0.0023	1.50%	0.0017	1.97%
λ_3	0.0003	0.22%	0.0004	0.44%
λ_4	0.0001	0.04%	0.0000	0.03%

Table A.4: Showing the absolute values and relative weights of the four greatest eigenvalues based on standard principal component analysis on the U.S. population (left: female; right: male) with different forecasting methods (top: Lee-Carter; middle: CBD; bottom: P-spline).

| Factor | Female | | Male | |
	Value	Weight	Value	Weight
Lee-Carter				
λ_1	0.0237	90.84%	0.0273	80.38%
λ_2	0.0010	3.78%	0.0027	7.98%
λ_3	0.0005	2.07%	0.0016	4.81%
λ_4	0.0004	1.66%	0.0010	2.89%
CBD				
λ_1	0.0506	98.31%	0.1679	98.43%
λ_2	0.0007	1.43%	0.0023	1.35%
λ_3	0.0001	0.25%	0.0003	0.20%
λ_4	0.0000	0.01%	0.0000	0.02%
P-spline				
λ_1	0.1088	99.63%	0.0366	99.19%
λ_2	0.0002	0.22%	0.0002	0.46%
λ_3	0.0002	0.14%	0.0001	0.29%
λ_4	0.0000	0.00%	0.0000	0.03%

Table A.5:   Showing the absolute values and relative weights of the four greatest eigenvalues based on standard principal component analysis on the West Germany population (left: female; right: male) with different forecasting methods (top: Lee-Carter; middle: CBD; bottom: P-spline).

	Parameter point estimate (st. dev.)				
	k	c	r	a	b
w\ self-consistency					
	-2.84×10^{-6}	1.02×10^{-8}	0.106	18.06	0.0088
	(4.54×10^{-7})	(2.44×10^{-9})	(0.002)	(2.02)	(0.0006)
wo\ self-consistency					
	-2.67×10^{-6}	1.16×10^{-8}	0.105	18.23	0.0088
	(4.49×10^{-7})	(1.93×10^{-9})	(0.001)	(2.05)	(0.0005)

Table A.6: U.S. female population & CBD approach: Showing maximum likelihood point estimates as well as associated standard deviations for the parameters k, c, r, a, and b. The standard deviations are obtained based on one hundred bootstrap samples of the $\boldsymbol{F}$ realizations. We consider both cases of with the self-consistency constraint imposed or without in the estimation.

	Parameter point estimate (st. dev.)				
	k	c	r	a	b
w\ self-consistency					
	-5.23×10^{-6}	1.27×10^{-8}	0.113	5.16	0.0070
	(1.60×10^{-6})	(4.91×10^{-9})	(0.003)	(0.66)	(0.0027)
wo\ self-consistency					
	-1.91×10^{-6}	1.36×10^{-8}	0.114	5.41	0.0067
	(1.55×10^{-6})	(5.11×10^{-9})	(0.003)	(0.60)	(0.0029)

Table A.7: U.S. female population & P-Spline approach: Showing maximum likelihood point estimates as well as associated standard deviations for the parameters k, c, r, a, and b. The standard deviations are obtained based on one hundred bootstrap samples of the $\boldsymbol{F}$ realizations. We consider both cases of with the self-consistency constraint imposed or without in the estimation.

	Parameter point estimate (st. dev.)				
	k	c	r	a	b
w\ self-consistency					
	7.95×10^{-6}	1.21×10^{-12}	0.201	16.88	0.0010
	(3.80×10^{-6})	(1.32×10^{-10})	(0.027)	(1.25)	(0.0008)
wo\ self-consistency					
	6.96×10^{-6}	1.27×10^{-12}	0.200	17.11	0.0010
	(4.29×10^{-6})	(1.03×10^{-10})	(0.028)	(1.69)	(0.0009)

Table A.8: U.S. male population & Lee-Carter approach: Showing maximum likelihood point estimates as well as associated standard deviations for the parameters k, c, r, a, and b. The standard deviations are obtained based on one hundred bootstrap samples of the $\boldsymbol{F}$ realizations. We consider both cases of with the self-consistency constraint imposed or without in the estimation.

	Parameter point estimate (st. dev.)				
	k	c	r	a	b
w\ self-consistency					
	-9.16×10^{-6}	1.04×10^{-8}	0.115	10.15	0.0067
	(2.25×10^{-6})	(2.74×10^{-9})	(0.003)	(2.27)	(0.0010)
wo\ self-consistency					
	-7.11×10^{-6}	1.28×10^{-8}	0.113	10.38	0.0066
	(1.42×10^{-6})	(2.69×10^{-9})	(0.002)	(2.55)	(0.0009)

Table A.9: England & Wales female population & CBD approach: Showing maximum likelihood point estimates as well as associated standard deviations for the parameters k, c, r, a, and b. The standard deviations are obtained based on one hundred bootstrap samples of the $\boldsymbol{F}$ realizations. We consider both cases of with the self-consistency constraint imposed or without in the estimation.

	Parameter point estimate (st. dev.)				
	k	c	r	a	b
w\ self-consistency					
	-6.24×10^{-7}	1.13×10^{-8}	0.110	5.52	0.0133
	(9.07×10^{-7})	(3.21×10^{-9})	(0.003)	(0.83)	(0.0008)
wo\ self-consistency					
	2.44×10^{-7}	1.00×10^{-8}	0.112	6.19	0.0133
	(5.33×10^{-7})	(2.67×10^{-9})	(0.003)	(0.46)	(0.0008)

Table A.10: West German male population & P-Spline approach: Showing maximum likelihood point estimates as well as associated standard deviations for the parameters k, c, r, a, and b. The standard deviations are obtained based on one hundred bootstrap samples of the $\boldsymbol{F}$ realizations. We consider both cases of with the self-consistency constraint imposed or without in the estimation.

	Parameter point estimate (st. dev.)				
	k	c	r	a	b
w\ self-consistency					
	-1.04×10^{-5}	6.31×10^{-8}	0.091	10.36	0.0116
	(2.61×10^{-6})	(2.36×10^{-8})	(0.003)	(3.03)	(0.0004)
wo\ self-consistency					
	-9.70×10^{-6}	6.19×10^{-8}	0.091	10.35	0.0115
	(2.57×10^{-6})	(2.28×10^{-8})	(0.003)	(3.05)	(0.0003)

Table A.11: French female population & Lee-Carter approach: Showing maximum likelihood point estimates as well as associated standard deviations for the parameters k, c, r, a, and b. The standard deviations are obtained based on one hundred bootstrap samples of the $\boldsymbol{F}$ realizations. We consider both cases of with the self-consistency constraint imposed or without in the estimation.

	Parameter point estimate (st. dev.)				
	k	c	r	a	b
w\ self-consistency					
	-4.61×10^{-6}	2.83×10^{-8}	0.098	14.09	0.0112
	(3.34×10^{-6})	(3.03×10^{-8})	(0.009)	(1.57)	(0.0008)
wo\ self-consistency					
	-2.55×10^{-6}	2.48×10^{-8}	0.099	14.19	0.0112
	(3.75×10^{-6})	(3.09×10^{-8})	(0.010)	(1.59)	(0.0008)

Table A.12: Japanese male population & Lee-Carter approach: Showing maximum likelihood point estimates as well as associated standard deviations for the parameters k, c, r, a, and b. The standard deviations are obtained based on one hundred bootstrap samples of the $\boldsymbol{F}$ realizations. We consider both cases of with the self-consistency constraint imposed or without in the estimation.

	Parameter point estimate				
	k	c	r	a	b
w\ self-consistency					
	-4.14×10^{-4}	1.86×10^{-6}	0.063	5.39	0.0319
wo\ self-consistency					
	-1.2×10^{-3}	1.85×10^{-5}	0.046	0.55	0.0270

Table A.13: U.K. pensioner: Showing maximum likelihood point estimates for the parameters k, c, r, a, and b. We consider both cases of with the self-consistency constraint imposed or without in the estimation.

Additional References

Angelini F, Herzel S (2005) Consistent Calibration of HJM Models to CAP Implied Volatilities. *J. Futures Markets* 25: 1093-1120.

Bauer D, Kramer F (2016) The Risk of a Mortality Catastrophe. *J. Bus. Econom. Statist.* 34: 391-405.

Biffis E (2005) Affine processes for dynamic mortality and actuarial valuations. *Insurance, Math., Econom.* 37: 443-468.

Box GEP (1954) Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification. *Ann. Math. Statist.* 25: 290-302.

Da Prato G, Zabczyk J (2014) *Stochastic equations in infinite dimensions*, 2nd ed. (Cambridge university press, Cambridge, MA).

Duffie D, Pan J, Singleton KJ (2000) Transform Analysis and Asset Pricing for Affine Jump-diffusions. *Econometrica* 68: 1343-1376.

Duffie D, Filipović D, Schachermayer W (2003) Affine processes and applications in finance. *Ann. Appl. Probab.* 13: 984-1053.

Moschopoulos PG (1985) The distribution of the sum of independent gamma random variables. *Ann. Inst. Stat. Math.* 37: 541-544.
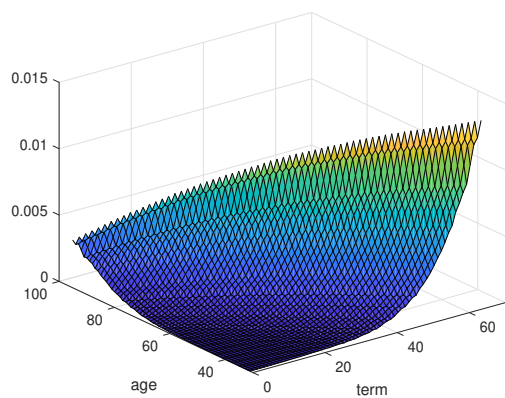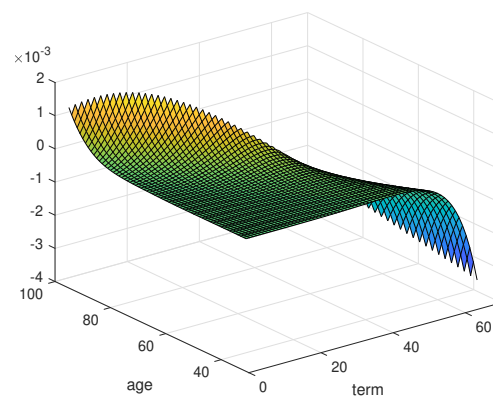
(a) female, Lee-Carter model
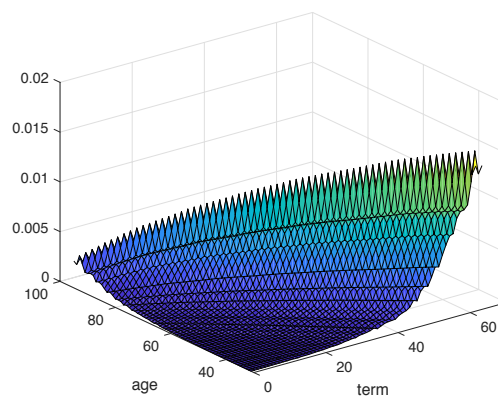
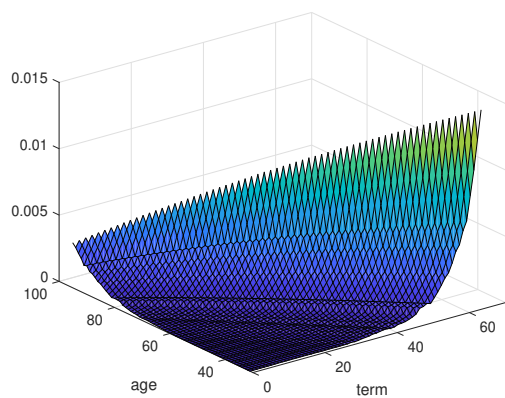(b) male, Lee-Carter model
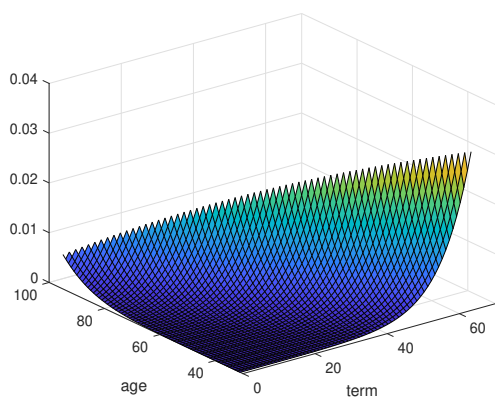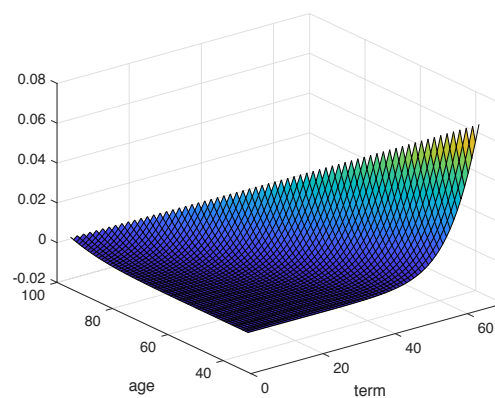
(c) female, CBD model
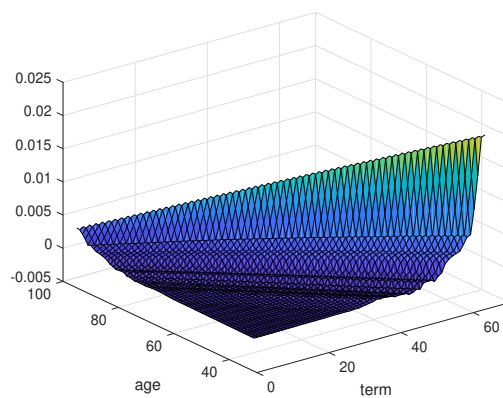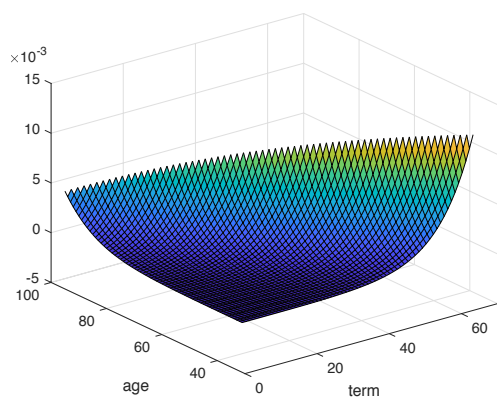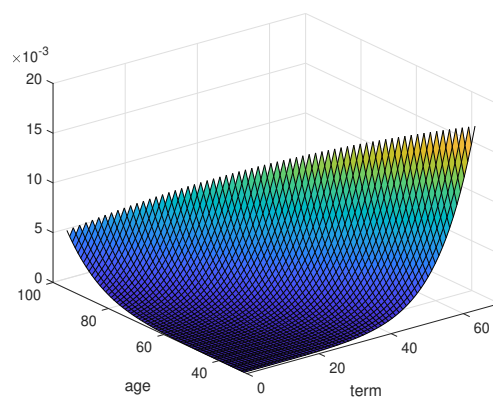
(d) male, CBD model
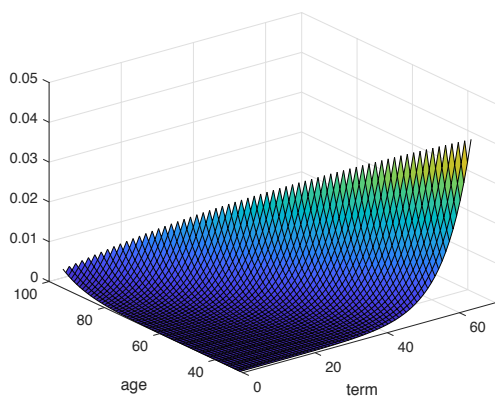
(e) female, P-spline model
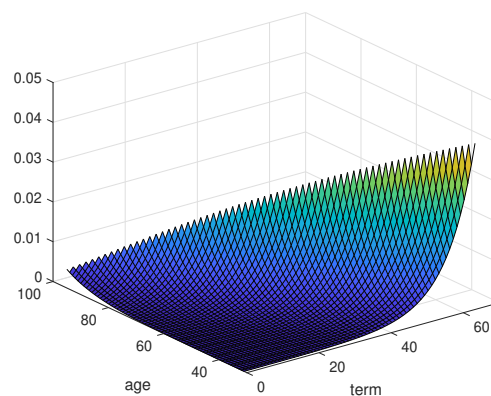
(f) male, P-spline model

Figure A.4: Showing the first eigenvectors (*slope factor*) that correspond to the mortality forecasts based on England & Wales population (left: female; right: male). Top panels: Lee-Carter model; middle panels: CBD model; bottom panels: P-spline model. Each eigenvector is plotted as a function of age (x) and term (τ), with $x \in \{31, \ldots, 95\}$, $\tau \in \{0, \ldots, 64\}$, and $x + \tau \in \{31, \ldots, 95\}$.
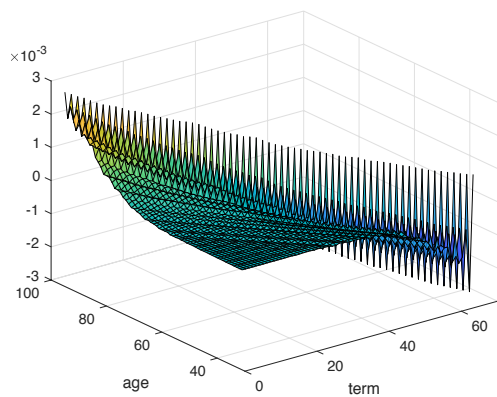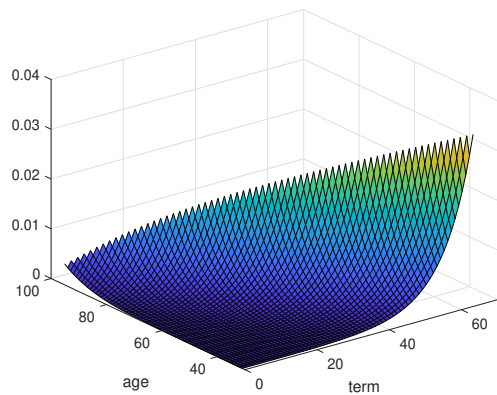
(a) female, Lee-Carter model
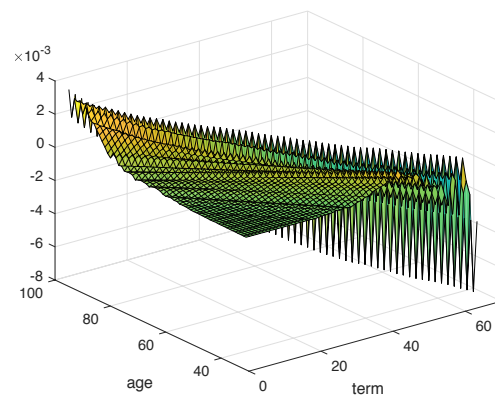
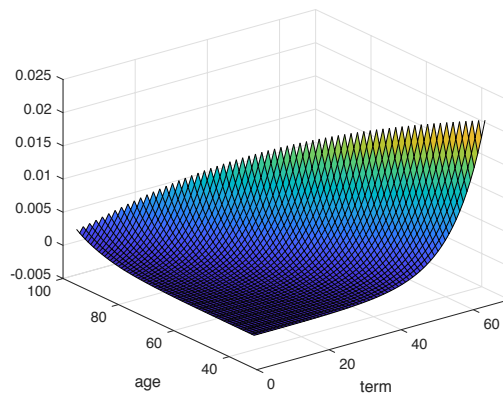(b) male, Lee-Carter model
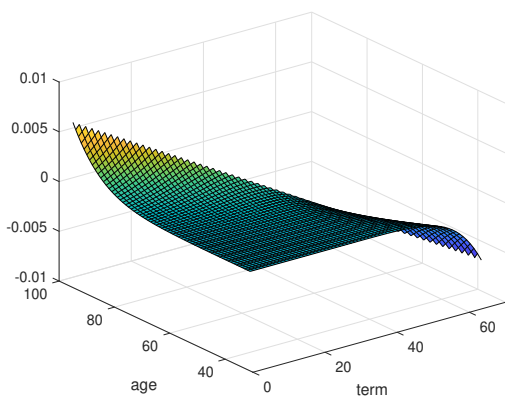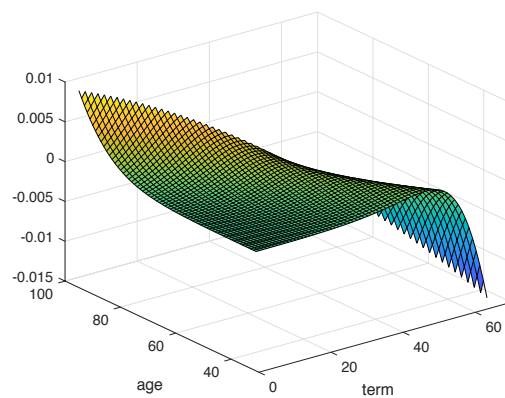
(c) female, CBD model

(d) male, CBD model
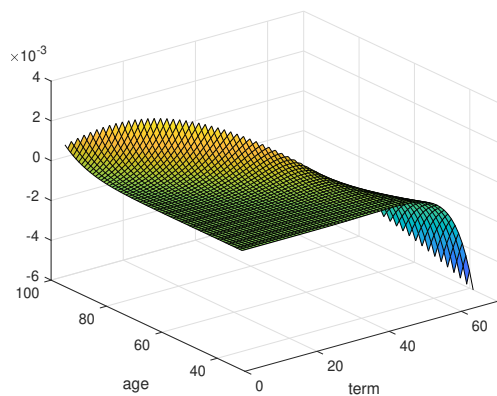
(e) female, P-spline model
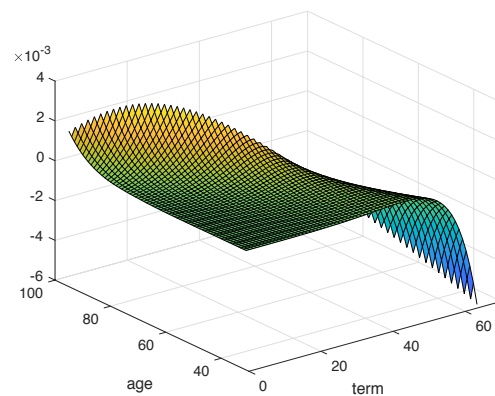
(f) male, P-spline model

Figure A.5: Showing the second eigenvectors (*curvature factor*) that correspond to the mortality forecasts based on England & Wales population (left: female; right: male). Top panels: Lee-Carter model; middle panels: CBD model; bottom panels: P-spline model. Each eigenvector is plotted as a function of age (x) and term (τ), with $x \in \{31, \ldots, 95\}$, $\tau \in \{0, \ldots, 64\}$, and $x + \tau \in \{31, \ldots, 95\}$.
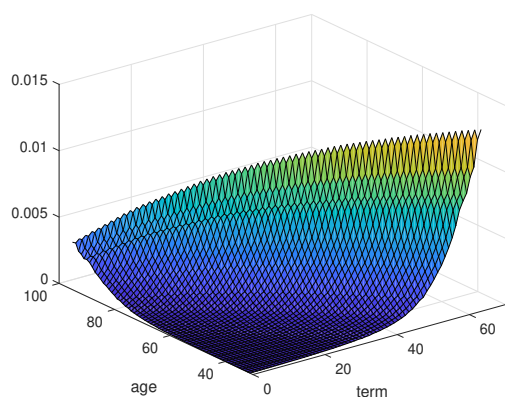
(a) female, Lee-Carter model
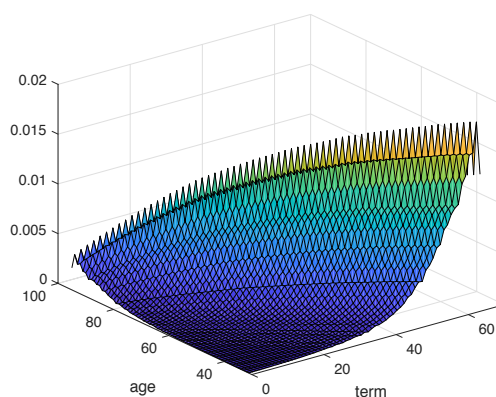
(b) male, Lee-Carter model
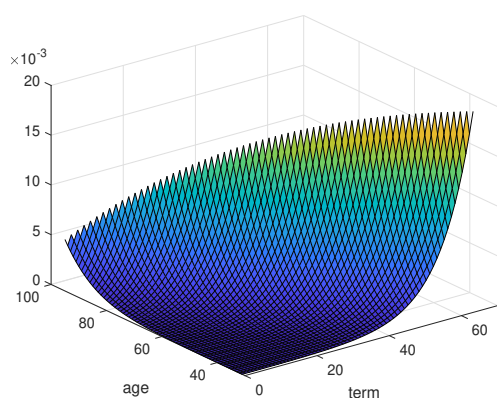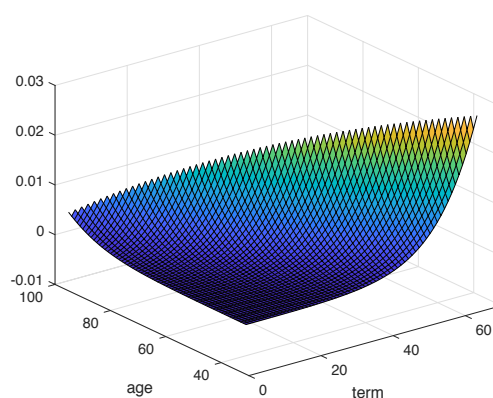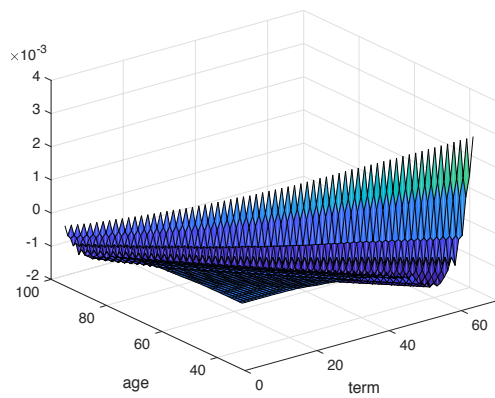
(c) female, CBD model
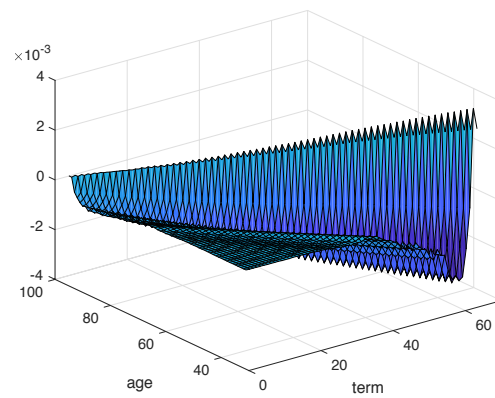
(d) male, CBD model
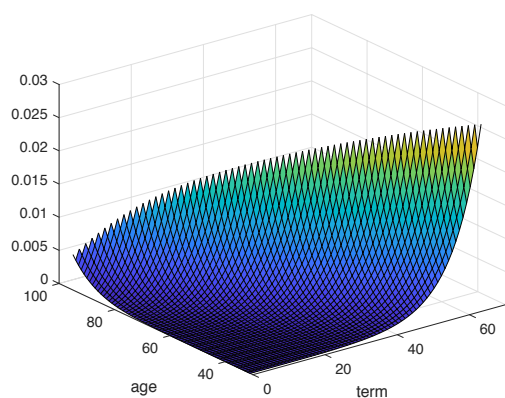
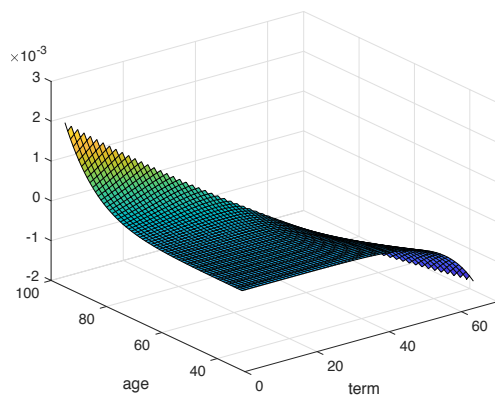(e) female, P-spline model
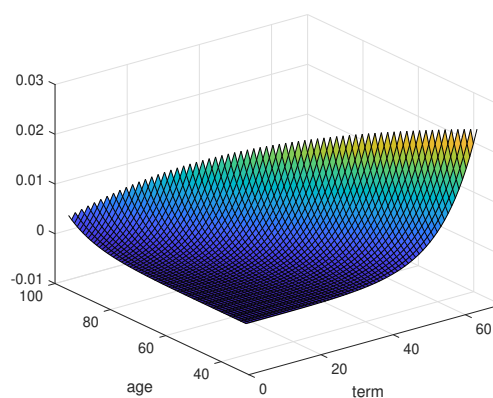
(f) male, P-spline model

Figure A.6: Showing the first eigenvectors (*slope factor*) that correspond to the mortality forecasts based on France population (left: female; right: male). Top panels: Lee-Carter model; middle panels: CBD model; bottom panels: P-spline model. Each eigenvector is plotted as a function of age (x) and term (τ), with $x \in \{31, \ldots, 95\}$, $\tau \in \{0, \ldots, 64\}$, and $x + \tau \in \{31, \ldots, 95\}$.
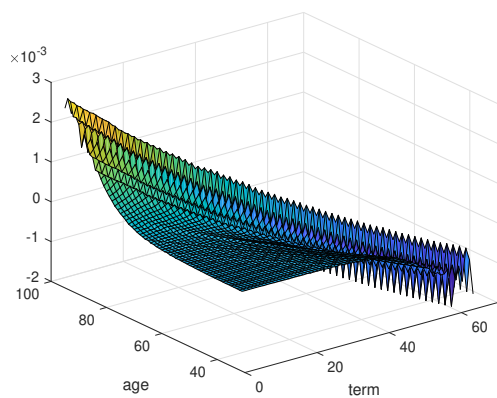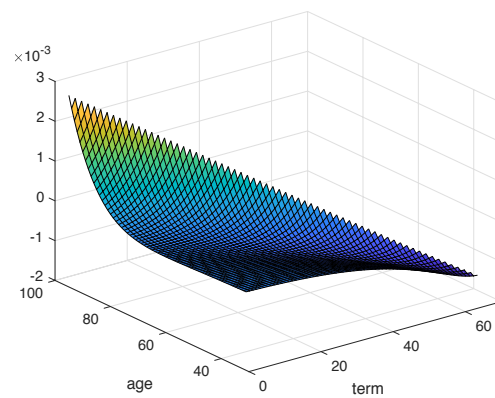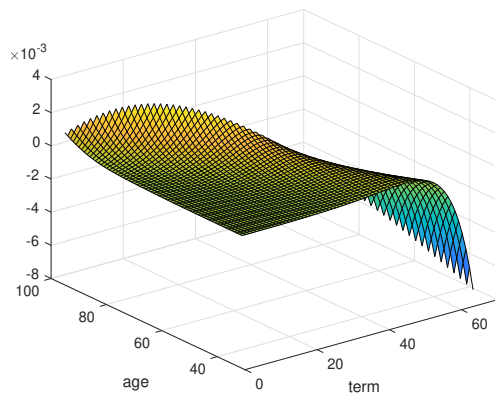
(a) female, Lee-Carter model
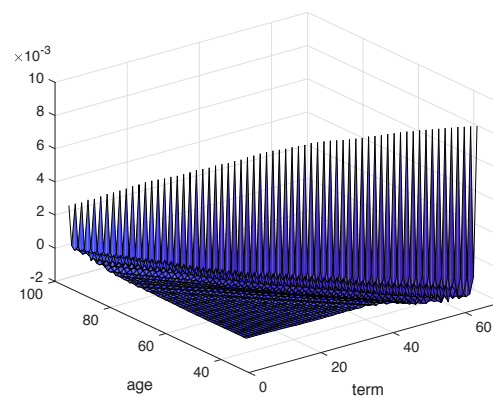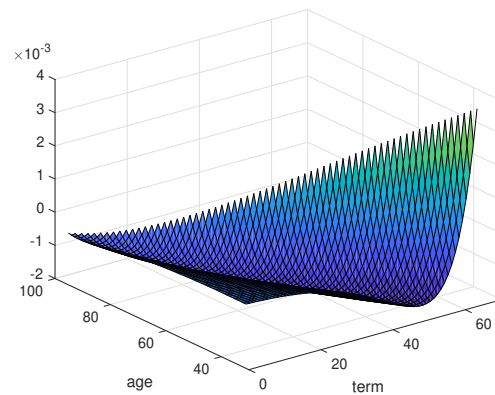
(b) male, Lee-Carter model
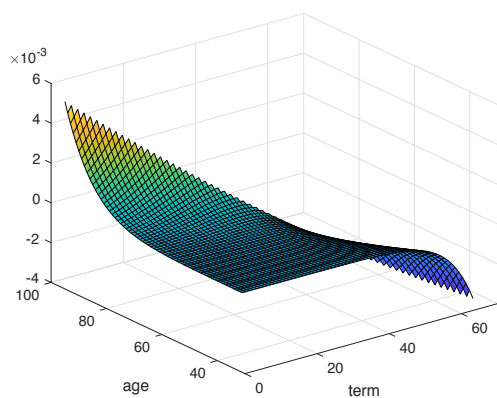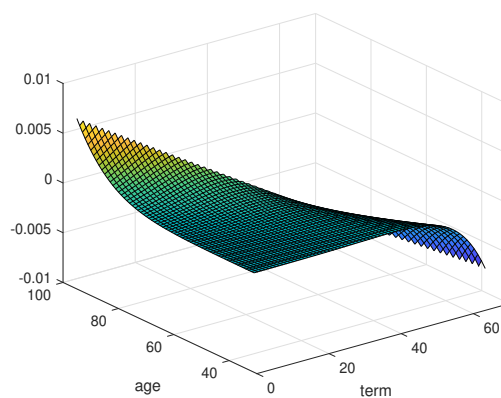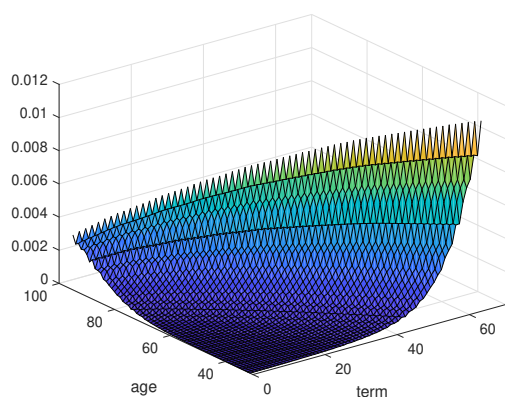
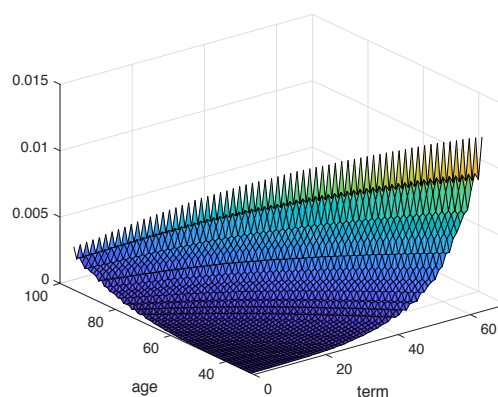(c) female, CBD model

(d) male, CBD model
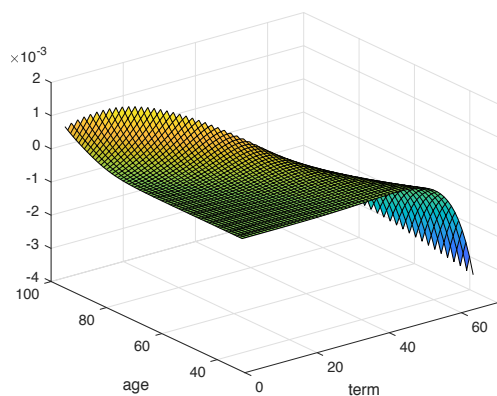
(e) female, P-spline model
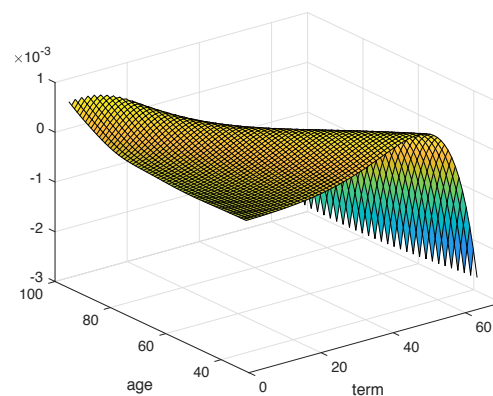
(f) male, P-spline model

Figure A.7: Showing the second eigenvectors (*curvature factor*) that correspond to the mortality forecasts based on France population (left: female; right: male). Top panels: Lee-Carter model; middle panels: CBD model; bottom panels: P-spline model. Each eigenvector is plotted as a function of age (x) and term (τ), with $x \in \{31, \ldots, 95\}$, $\tau \in \{0, \ldots, 64\}$, and $x + \tau \in \{31, \ldots, 95\}$.
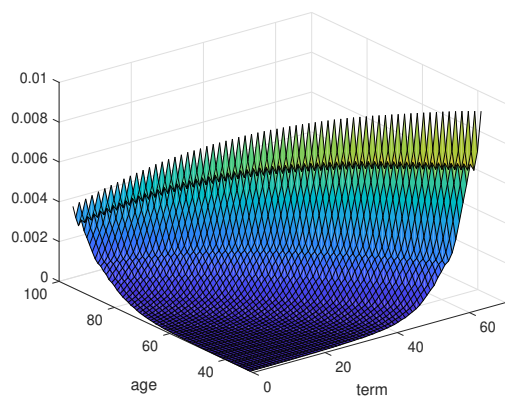
(a) female, Lee-Carter model
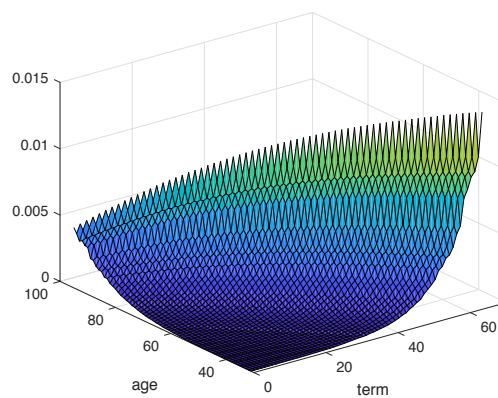
(b) male, Lee-Carter model
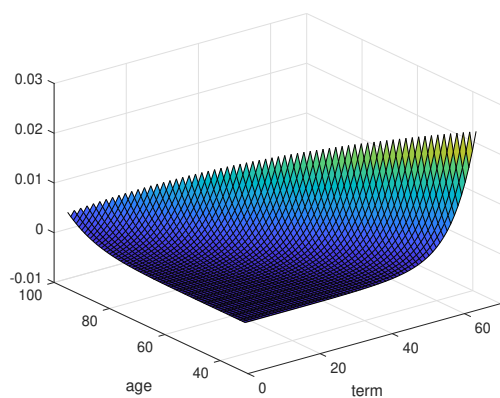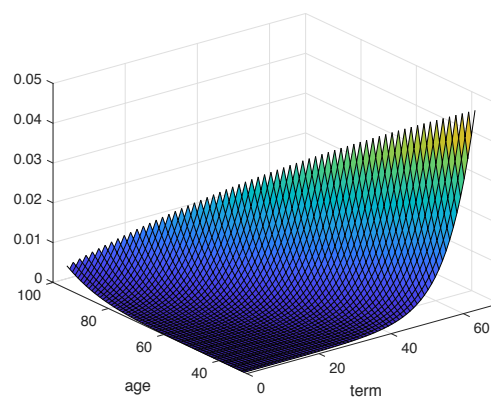
(c) female, CBD model

(d) male, CBD model
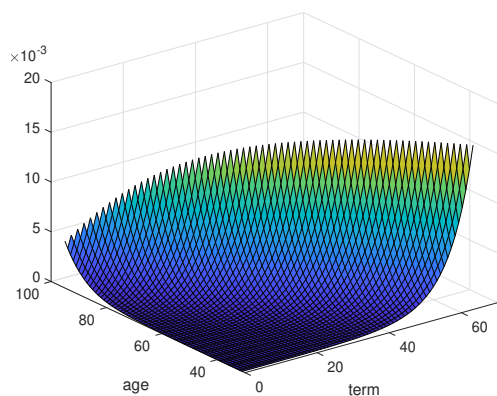
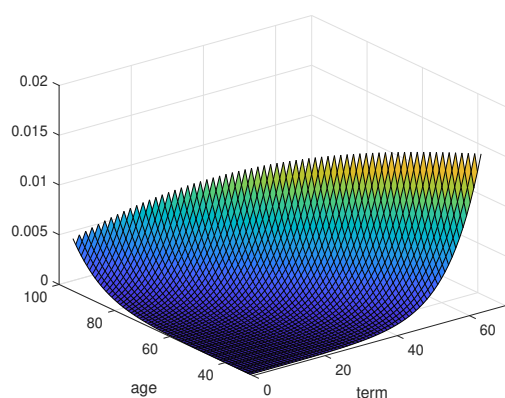(e) female, P-spline model
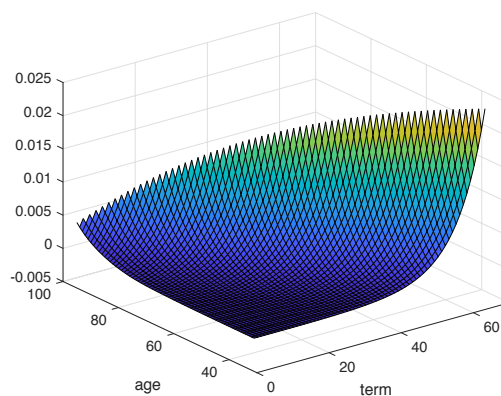
(f) male, P-spline model

Figure A.8: Showing the first eigenvectors (*slope factor*) that correspond to the mortality forecasts based on Japan population (left: female; right: male). Top panels: Lee-Carter model; middle panels: CBD model; bottom panels: P-spline model. Each eigenvector is plotted as a function of age (x) and term (τ), with $x \in \{31, \ldots, 95\}$, $\tau \in \{0, \ldots, 64\}$, and $x + \tau \in \{31, \ldots, 95\}$.
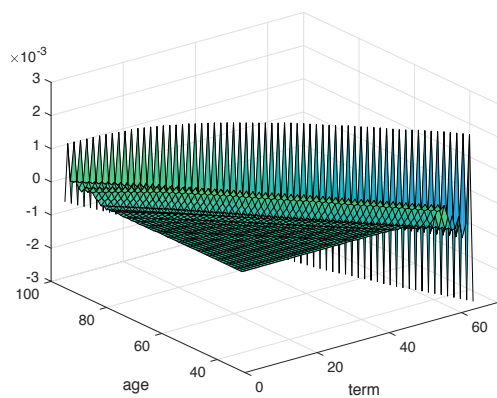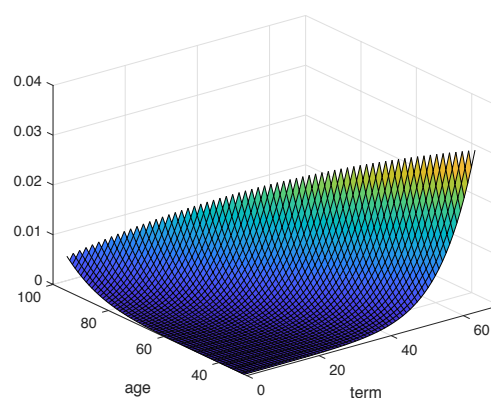
(a) female, Lee-Carter model
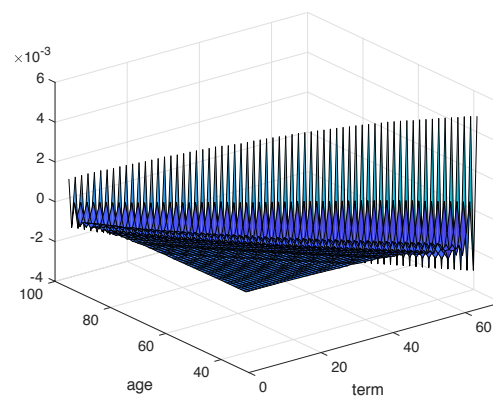
(b) male, Lee-Carter model
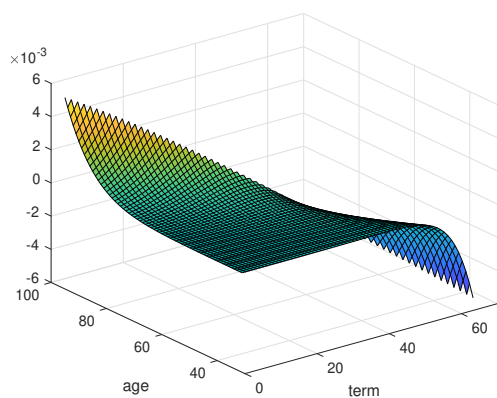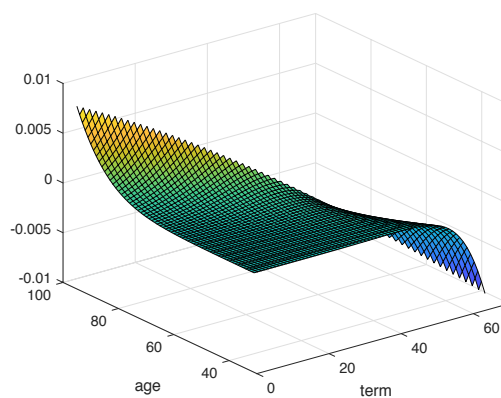
(c) female, CBD model

(d) male, CBD model
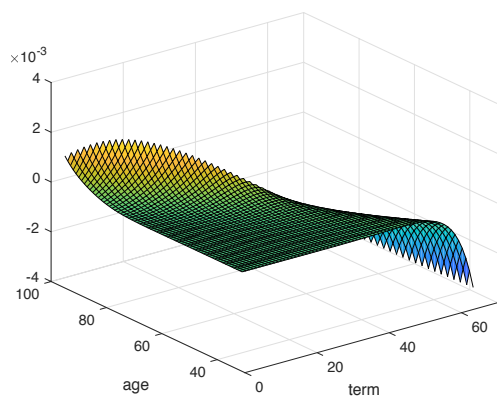
(e) female, P-spline model

(f) male, P-spline model

Figure A.9: Showing the second eigenvectors (*curvature factor*) that correspond to the mortality forecasts based on Japan population (left: female; right: male). Top panels: Lee-Carter model; middle panels: CBD model; bottom panels: P-spline model. Each eigenvector is plotted as a function of age (x) and term (τ), with $x \in \{31, \ldots, 95\}$, $\tau \in \{0, \ldots, 64\}$, and $x + \tau \in \{31, \ldots, 95\}$.
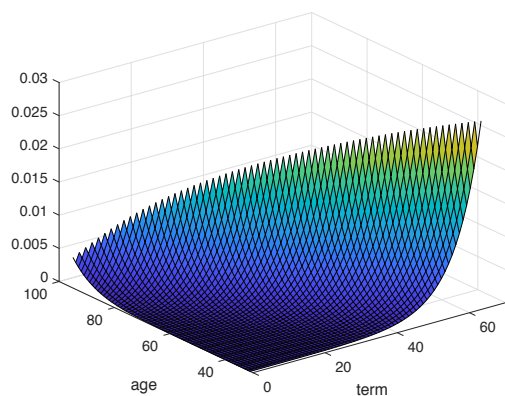
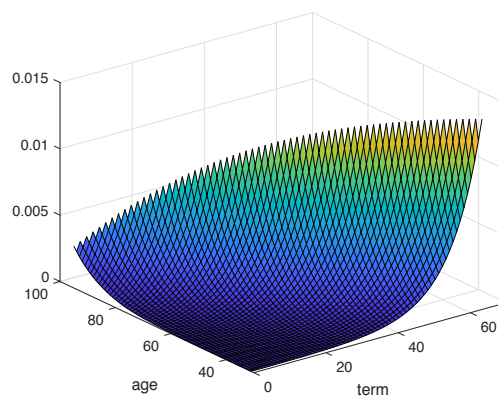(a) female, Lee-Carter model

(b) male, Lee-Carter model
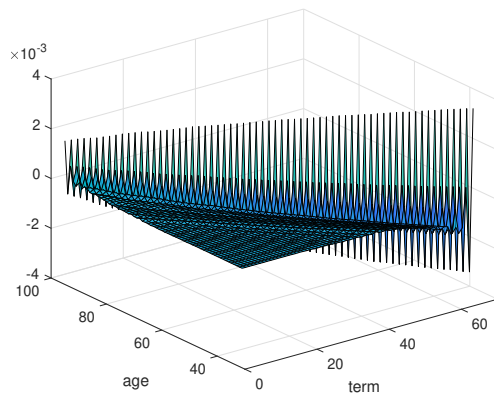
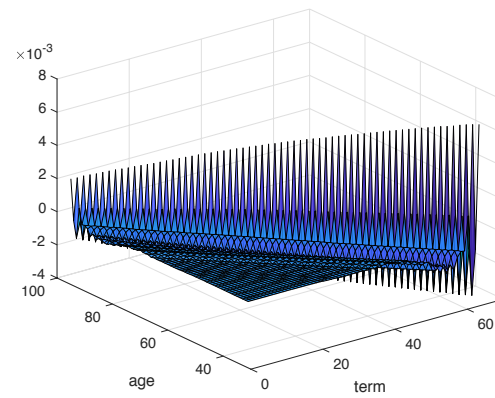(c) female, CBD model

(d) male, CBD model

(e) female, P-spline model

(f) male, P-spline model

Figure A.10: Showing the first eigenvectors (*slope factor*) that correspond to the mortality forecasts based on U.S. population (left: female; right: male). Top panels: Lee-Carter model; middle panels: CBD model; bottom panels: P-spline model. Each eigenvector is plotted as a function of age (x) and term (τ), with $x \in \{31, \ldots, 95\}$, $\tau \in \{0, \ldots, 64\}$, and $x + \tau \in \{31, \ldots, 95\}$.

(a) female, Lee-Carter model

(b) male, Lee-Carter model

(c) female, CBD model

(d) male, CBD model

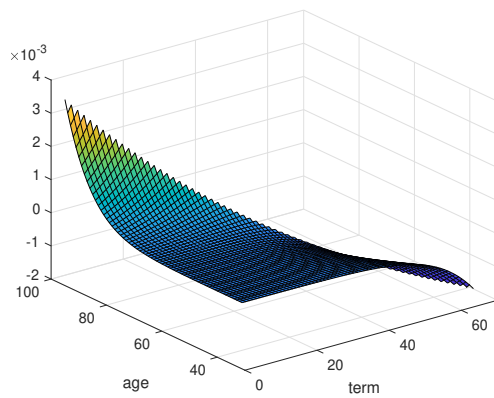(e) female, P-spline model

(f) male, P-spline model

Figure A.11: Showing the second eigenvectors (*curvature factor*) that correspond to the mortality forecasts based on U.S. population (left: female; right: male). Top panels: Lee-Carter model; middle panels: CBD model; bottom panels: P-spline model. Each eigenvector is plotted as a function of age (x) and term (τ), with $x \in \{31, \ldots, 95\}$, $\tau \in \{0, \ldots, 64\}$, and $x + \tau \in \{31, \ldots, 95\}$.
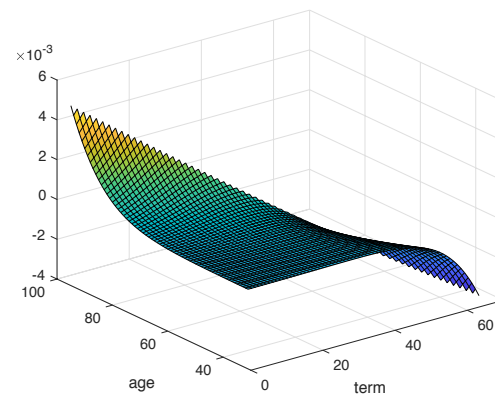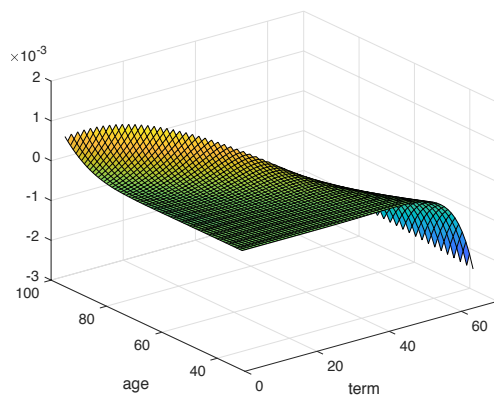
(a) female, Lee-Carter model

(b) male, Lee-Carter model

(c) female, CBD model

(d) male, CBD model

(e) female, P-spline model

(f) male, P-spline model

Figure A.12: Showing the first eigenvectors (*slope factor*) that correspond to the mortality forecasts based on West Germany population (left: female; right: male). Top panels: Lee-Carter model; middle panels: CBD model; bottom panels: P-spline model. Each eigenvector is plotted as a function of age (x) and term (τ), with $x \in \{31, \dots, 95\}$, $\tau \in \{0, \dots, 64\}$, and $x + \tau \in \{31, \dots, 95\}$.

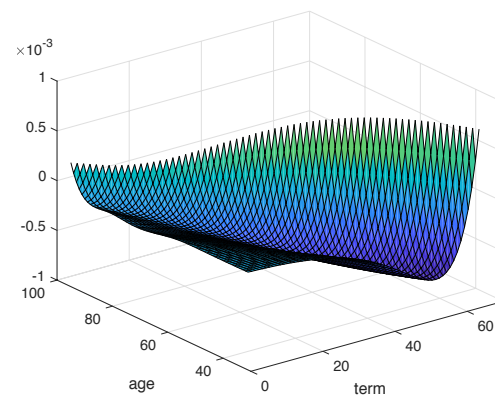(a) female, Lee-Carter model

(b) male, Lee-Carter model

(c) female, CBD model

(d) male, CBD model

(e) female, P-spline model

(f) male, P-spline model

Figure A.13: Showing the second eigenvectors (*curvature factor*) that correspond to the mortality forecasts based on West Germany population (left: female; right: male). Top panels: Lee-Carter model; middle panels: CBD model; bottom panels: P-spline model. Each eigenvector is plotted as a function of age (x) and term (τ), with $x \in \{31, \ldots, 95\}$, $\tau \in \{0, \ldots, 64\}$, and $x + \tau \in \{31, \ldots, 95\}$.

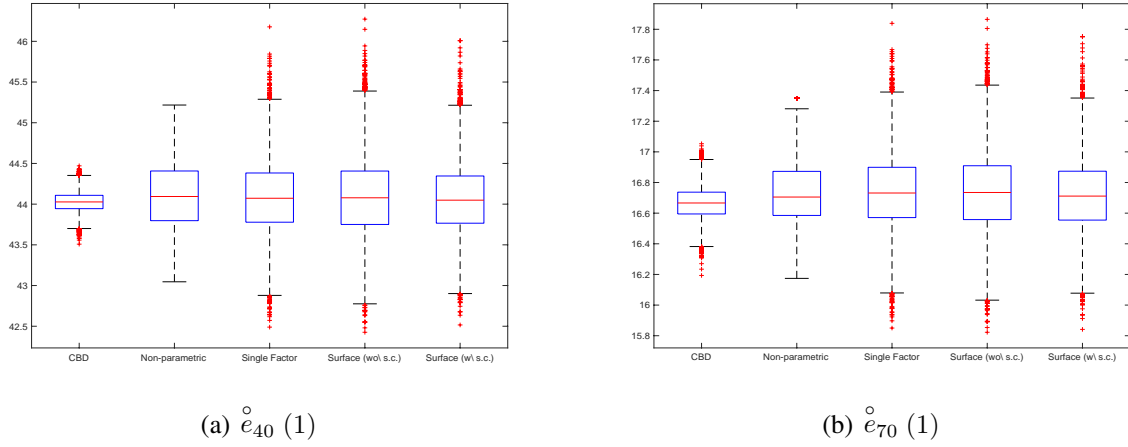(a) $\overset{\circ}{e}_{40}(1)$ (b) $\overset{\circ}{e}_{70}(1)$

Figure A.14: U.S. female population & CBD approach: Showing the prediction intervals of the expected future life-time for age-40 (left panel) and age-70 (right panel) cohorts at year 2017 under $10,000$ simulations. We consider five different approaches as shown by order: The CBD model with error terms; the non-parametric forecasting approach; the single-factor model; the mortality surface model without the self-consistency constraint; and the mortality surface model with the self-consistency constraint.

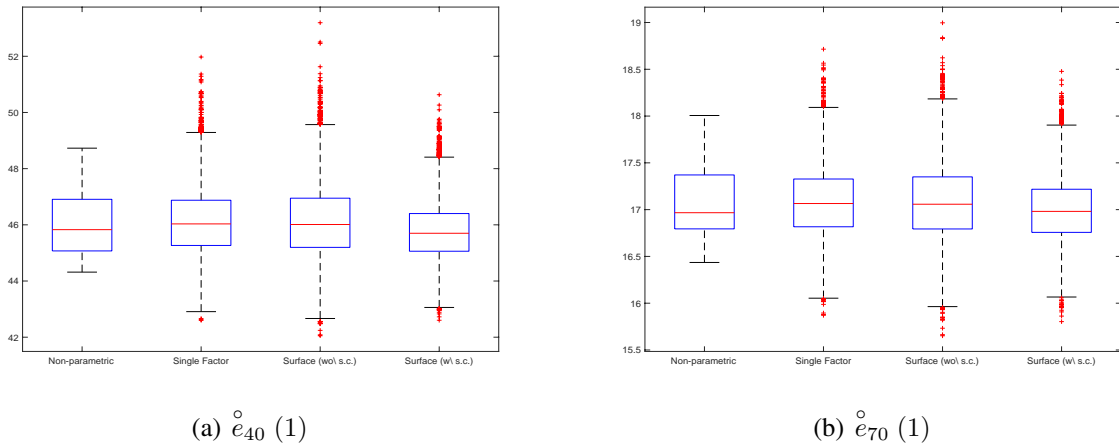(a) $\overset{\circ}{e}_{40}(1)$ (b) $\overset{\circ}{e}_{70}(1)$

Figure A.15: U.S. female population & P-spline approach: Showing the prediction intervals of the expected future life-time for age-40 (left panel) and age-70 (right panel) cohorts at year 2008 under $10,000$ simulations. We consider four different approaches as shown by order: The non-parametric forecasting approach; the single-factor model; the mortality surface model without the self-consistency constraint; and the mortality surface model with the self-consistency constraint.

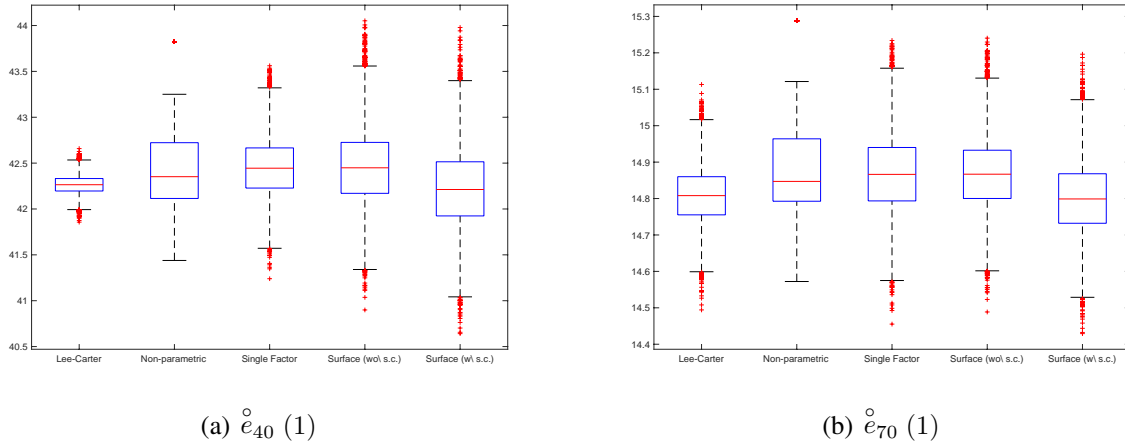(a) $\overset{\circ}{e}_{40}(1)$ (b) $\overset{\circ}{e}_{70}(1)$

Figure A.16: U.S. male population & Lee-Carter approach: Showing the prediction intervals of the expected future life-time for age-40 (left panel) and age-70 (right panel) cohorts at year 2017 under $10,000$ simulations. We consider five different approaches as shown by order: The Lee-Carter model with error terms; the non-parametric forecasting approach; the single-factor model; the mortality-surface model without the self-consistency constraint; and the mortality surface model with the self-consistency constraint.

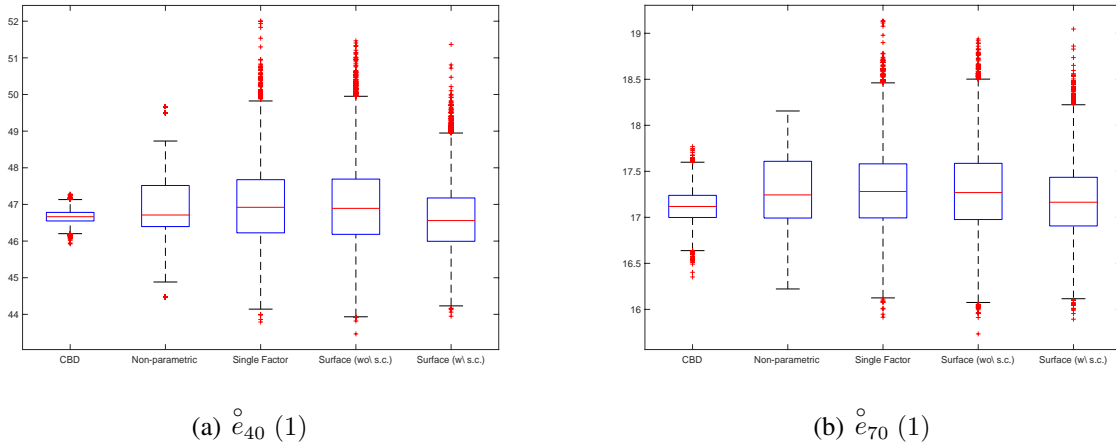(a) $\overset{\circ}{e}_{40}(1)$ (b) $\overset{\circ}{e}_{70}(1)$

Figure A.17: England & Wales female population & CBD approach: Showing the prediction intervals of the expected future life-time for age-40 (left panel) and age-70 (right panel) cohorts at year 2017 under $10,000$ simulations. We consider five different approaches as shown by order: The CBD model with error terms; the non-parametric forecasting approach; the single-factor model; the mortality surface model without the self-consistency constraint; and the mortality surface model with the self-consistency constraint.

(a) $\mathring{e}_{40}\,(1)$
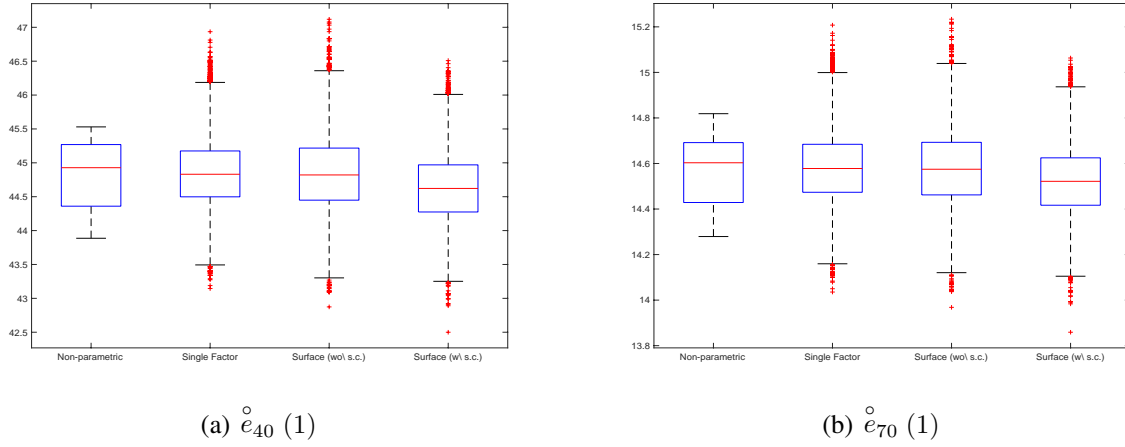
(b) $\mathring{e}_{70}\,(1)$

Figure A.18: West German male population & P-spline approach: Showing the prediction intervals of the expected future life-time for age-40 (left panel) and age-70 (right panel) cohorts at year 2008 under $10,000$ simulations. We consider four different approaches as shown by order: The non-parametric forecasting approach; the single-factor model; the mortality surface model without the self-consistency constraint; and the mortality surface model with the self-consistency constraint.

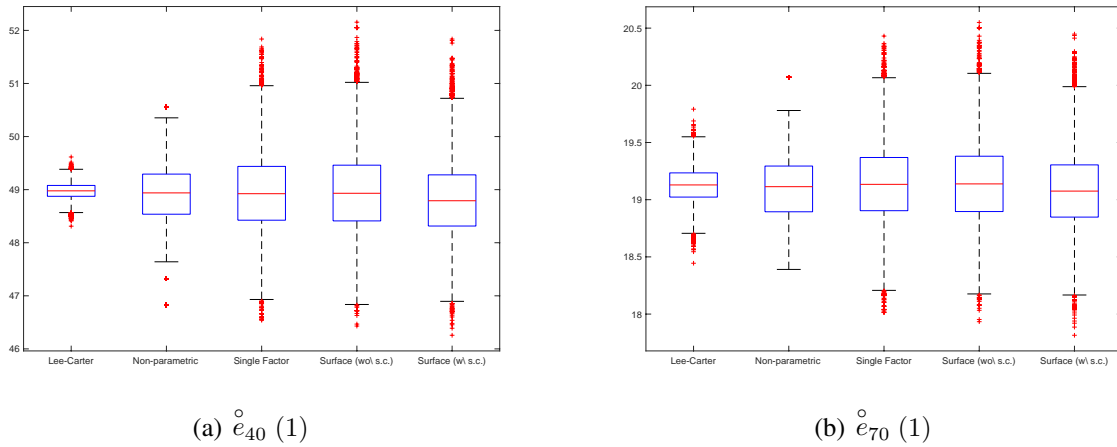(a) $\mathring{e}_{40}\,(1)$

(b) $\mathring{e}_{70}\,(1)$

Figure A.19: French female population & Lee-Carter approach: Showing the prediction intervals of the expected future life-time for age-40 (left panel) and age-70 (right panel) cohorts at year 2017 under $10,000$ simulations. We consider five different approaches as shown by order: The Lee-Carter model with error terms; the non-parametric forecasting approach; the single-factor model; the mortality-surface model without the self-consistency constraint; and the mortality surface model with the self-consistency constraint.

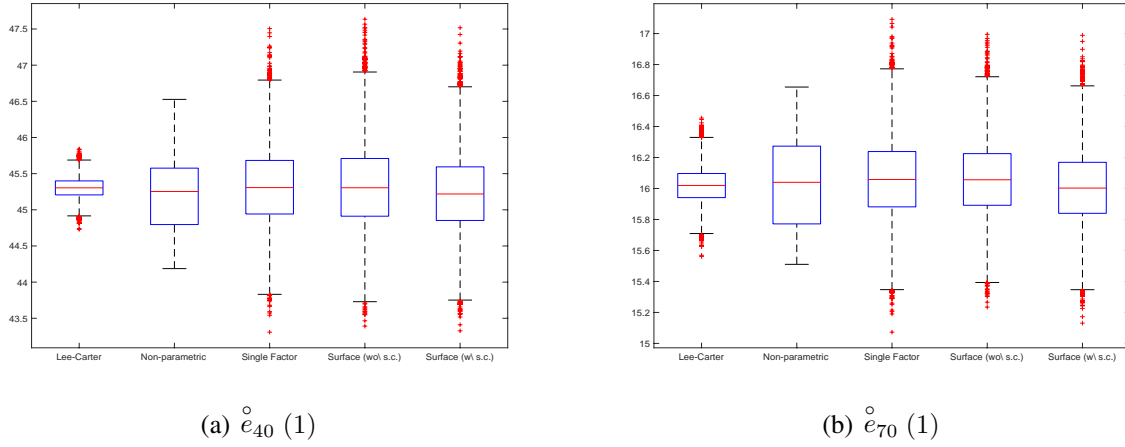(a) $\overset{\circ}{e}_{40}(1)$ (b) $\overset{\circ}{e}_{70}(1)$

Figure A.20: Japanese male population & Lee-Carter approach: Showing the prediction intervals of the expected future life-time for age-40 (left panel) and age-70 (right panel) cohorts at year 2017 under 10,000 simulations. We consider five different approaches as shown by order: The Lee-Carter model with error terms; the non-parametric forecasting approach; the single-factor model; the mortality-surface model without the self-consistency constraint; and the mortality surface model with the self-consistency constraint.

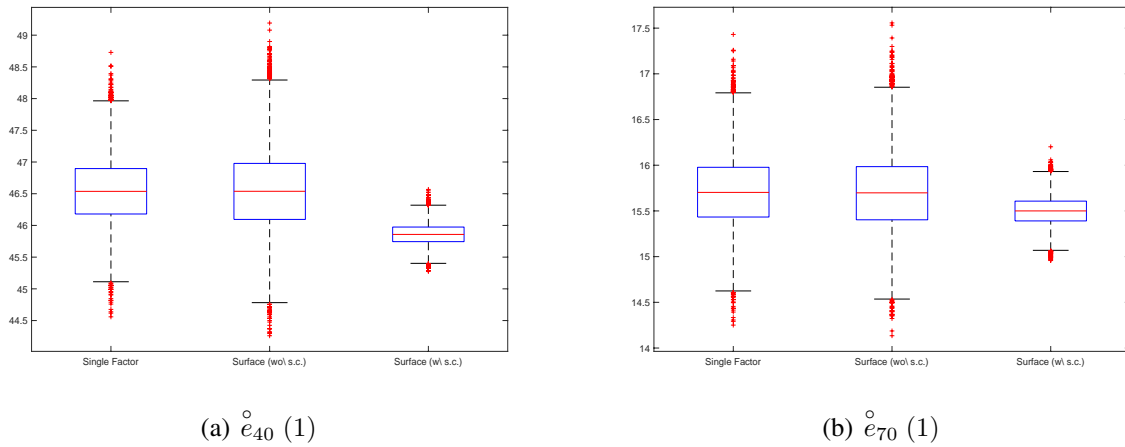(a) $\overset{\circ}{e}_{40}(1)$ (b) $\overset{\circ}{e}_{70}(1)$

Figure A.21: U.K. pensioner: Showing the prediction intervals of the expected future life-time for age-40 (left panel) and age-70 (right panel) cohorts at year 2006 under 10,000 simulations. We consider three different approaches as shown by order: The single-factor model; the mortality surface model without the self-consistency constraint; and the mortality surface model with the self-consistency constraint.