Insurance: Mathematics and Economics 113 (2023) 96–121

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Hedging longevity risk under non-Gaussian state-space stochastic mortality models: A mean-variance-skewness-kurtosis approach

Johnny Siu-Hang Li [a,*], Yanxin Liu [b], Wai-Sum Chan [c]

[a] Department of Statistics and Actuarial Science, University of Waterloo, Canada
[b] Department of Finance, University of Nebraska-Lincoln, USA
[c] Department of Finance, The Chinese University of Hong Kong, Hong Kong, China

ARTICLE INFO

Article history:
Received November 2021
Received in revised form July 2023
Accepted 10 August 2023
Available online 19 August 2023

JEL classification:
C10
C53

Keywords:
Higher moments
State-space models
Polynomial goal programming
Longevity risk

ABSTRACT

Longevity risk has recently become a high profile risk among insurers and pension plan sponsors. One way to mitigate longevity risk is to build a hedge using derivatives that are linked to mortality indexes. Longevity hedging methods are often based on the normality assumption, considering only the variance but no other (higher) moments. However, strong empirical evidence suggests that mortality improvement rates are driven by asymmetric and fat-tailed distributions, so that existing longevity hedging methods should be expanded to incorporate higher moments. This paper fills the gap by adopting a mean-variance-skewness-kurtosis approach based on non-Gaussian extensions of commonly-used stochastic mortality models, formulated in a state-space setting. On the basis of a general representation of these models, the authors derive (approximate) analytical expressions for the moments of the present values of the hedging instruments and the liability being hedged. These expressions are then integrated with a polynomial goal programming model, from which the optimal hedge portfolio is identified. Finally, the paper demonstrates the theoretical results with a real mortality data set and a range of hedger preferences.

1. Introduction

Longevity risk refers to the risk that individuals live longer than expected. The risk creates challenges, not only for the individual who needs an income for a period longer than expected after retirement, but also for governments, defined benefit retirement funds, and life insurers who face retirement-related liabilities that increase as a result of improved life expectancy.

Roughly speaking, longevity risk is comprised of two components, namely, ‘micro’ and ‘macro’ longevity risks. Micro longevity risk refers to the uncertainty surrounding a known, fixed survival distribution. This piece of uncertainty is diversifiable, and is negligible when the number of lives in the annuity/pension portfolio is sufficiently large. Macro longevity risk, in contrast, refers to the uncertainty associated with the underlying survival distribution itself. Equivalently speaking, it represents the uncertainty associated with the sequence of death rates/probabilities that are not yet realized. It is important to note that macro longevity risk is systematic, as any change in the underlying survival distribution affects all lives in the portfolio. Although macro longevity risk is not diversifiable, it may be transferred to a counterparty who is willing to take it for a risk premium and/or diversification benefits.

The longevity risk market started in the United Kingdom in less than two decades ago. The market provides a platform for institutions to transfer their longevity risk exposures through various means, most notably capital market derivatives. In 2007, J.P. Morgan introduced the LifeMetrics Index, a mortality index that is composed of age-specific death probabilities in certain populations. The first capital markets transaction involving the LifeMetrics Index reportedly took place in January 2008 (Blake et al., 2013). In this transaction, Lucida (a monoline insurer in the United Kingdom) transferred part of its longevity risk exposures to J.P. Morgan through a derivative known as q-forward, written on the LifeMetrics index for the male population of England and Wales. Since then, a number of other investment banks, including Credit Suisse, Deutsche Börse, Deutsche Bank, Goldman Sachs, and Société Générale, have also launched tradable mortality indexes and written derivative securities on these indexes with insurers and pension plan sponsors in various countries. Longevity-linked

* Corresponding author.
E-mail address: shli@uwaterloo.ca (J.S.-H. Li).

https://doi.org/10.1016/j.insmatheco.2023.08.001
0167-6687/

securities have become a new class of financial assets, with a potential market size of US$60 to 80 trillion (Michaelson and Mulholland, 2014).

There are many parallels between macro longevity risk and typical financial risks (e.g., equity and interest rate) in terms of modeling. First, to quantify longevity risk and to price longevity-linked securities, an econometric model capturing the stochastic evolution of mortality rates over time and/or year-of-birth is necessary. Following the spirit of constructing interest rate models which preserve the stylized properties of interest rates such as term-structures and mean-reversion (Duan, 2016; Goliński and Zaffaroni, 2016; Lee and Tse, 1991; Tse, 1998), any econometric model for macro longevity risk should be carefully structured so that the resulting forecasts of future mortality are biologically reasonable. For instance, when applied to mortality at pensionable ages, the model should yield projected mortality rates that are monotonically increasing with age at any given time point. In addition, as with econometric modeling of typical financial risks, the modeler should consider goodness-of-fit to historical data when choosing a macro longevity risk model (Ling and Tong, 2011).

Second, similar to stock returns, mortality improvements are not necessarily (log-)normally distributed. Econometricians have found strong empirical evidence supporting the non-normality of stock returns, and have developed various non-normal models by which skewness and excess kurtosis of stock returns can be captured (Bakshi et al., 2003; Mills, 1995; Peiro, 1999). Portfolio optimization methods which incorporate higher moments have also been developed accordingly (Harvey et al., 2010; Jondeau and Rockinger, 2006; Prakash et al., 2003). On the other hand, there has been a growing concern among researchers working on the field of longevity risk that mortality improvements are not normally distributed. Non-normality is evident in the historical mortality improvements of various populations, and a number of attempts have been made to relax the normality assumption in typical stochastic mortality models (econometric models for macro longevity risk). For instance, Ahmadi and Gaillardetz (2014) extended the Cairns-Blake-Dowd model (a commonly used stochastic mortality model) to incorporate skewness and excess kurtosis of mortality improvements. Similar extensions have been considered by Giacometti et al. (2009) and Wang et al. (2011, 2013).

The potential non-normality of mortality improvements highlights the need for considering higher moments in developing capital markets solutions for longevity risk. In the literature, however, the effectiveness of a longevity risk is typically measured in terms of reduction in variance, without considering any higher moment that may also be indicative of the risk inherent to the hedged and unhedged positions (Cairns, 2011; Cairns et al., 2014; Li and Hardy, 2011; Zhou and Li, 2017). Furthermore, existing moment-based methods for longevity hedge optimization consider either variance only if the cost of hedging is ignored or mean and variance only if the cost of hedging is taken into account (Li et al., 2017; Liu and Li, 2017, 2018; Xue et al., 2018; Zhang et al., 2017). These methods are clearly inadequate when the distribution of mortality improvements is non-normal and the hedger concerns with, for example, the tail risk associated with his/her portfolio on top of volatility. Non-moment-based methods such as longevity Greeks and mortality durations (Li and Hardy, 2011; Li and Luo, 2012; Plat, 2011) somewhat ameliorate the problem of having a sole focus on mean and/or variance, but they are still unable to incorporate the hedger's preferences concerning various moments explicitly.

In this paper, we fill the literature gap by proposing a mean-variance-skewness-kurtosis approach to optimizing longevity hedges. In the proposed approach, the hedge portfolio (the vector of the notional amounts of the hedging instruments) is optimized when all of the first four moments of the distribution of portfolio values are considered simultaneously. The optimization challenge is overcome by employing the polynomial goal programming method, previously considered by Lai et al. (2006), Leung et al. (2001), Prakash et al. (2003) and Sun and Yan (2003) in the context of banking and investment. When applied to the context of our research, the polynomial goal programming method enables the hedger to connect multiple objectives concerning the four moments, and to specify his/her preference concerning each of the four moments. We further augment the polynomial goal programming method to incorporate a budget constraint, which guarantees that the (expected) cost of hedging is no greater than a certain fraction of the value of the liability being hedged. This additional feature makes our set-up a closer resemblance to real-life situations, in which institutions typically have finite budgets for risk management purposes.

The proposed hedging strategy is built on a general state-space representation of stochastic mortality models. The general state-space representation encompasses most discrete-time stochastic mortality models used in practice, including the Lee-Carter model (Lee and Carter, 1992), the Cairns-Blake-Dowd model (Cairns et al., 2006) and the Renshaw-Haberman model (Renshaw and Haberman, 2006). We believe that the flexibility provided by our proposed method is essential, because, based on the experience of Cairns et al. (2009) and Li et al. (2015), the optimal choice of a stochastic mortality model is heavily data dependent.

The implementation of the polynomial goal programming method depends on the first four moments of the values of the hedging instruments and the liability being hedged. In principle, these moments can be obtained using (nested) simulations, but the computational effort entailed may reduce the practicability of our proposed hedging strategies. To mitigate this problem, we derive approximate analytical formulas for calculating these moments. Expressed in terms of the parameters in the general state-space representation of stochastic mortality models, the approximation formulas are readily applicable in various situations. We apply our theoretical results to a real mortality data set, considering a range of hedger preferences. Through this application, the limitations of mean-variance hedging are demonstrated.

The remainder of this paper is organized as follows. Section 2 presents the general state-space representation of stochastic mortality models, and discusses the distributional assumptions we make. Section 3 presents the set-up of our work and specifies the hedge objectives. Section 4 details the polynomial goal programming method through which the optimal hedge portfolio is identified. Section 5 derives the approximation formulas for calculating the moments of the values of the hedging instruments and the liability being hedged. Section 6 provides a numerical illustration of our theoretical results. Finally, concluding remarks are made in Section 7.

2. Stochastic mortality models

2.1. Preliminaries

Macro longevity risk can be modeled by stochastic mortality models. This sub-section introduces stochastic mortality models by way of two examples. When defining the models, it is assumed that they are estimated to historical mortality data over a calibration window that spans $t = t_a$ to $t = t_b$ and an age range that covers $x = x_a$ to $x = x_b$.

2.1.1. The Lee-Carter model

The Lee-Carter model is defined as

$$\ln(m(x,t)) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \epsilon(x,t).$$

The quantity being modeled here is $m(x,t)$, which represents the central rate of death at age x in year t. It is defined as the ratio of the total number of deaths over the age interval of $[x, x+1)$ during the time interval of $[t-1, t)$ (i.e., the beginning of year t to the end of year t) to the corresponding number of persons at risk. As this quantity is bounded below by zero, a log transform is used.

In the model, parameters $\beta_x^{(1)}$, $x = x_a, \ldots, x_b$, represent the age effect, which refers to the variation of the base (historical average) level of mortality across ages. The dynamics of $\ln(m(x,t))$ at all ages are driven by a single time-varying index, $\kappa_t^{(2)}$, which, in many applications, is assumed to follow a random walk with a constant drift. However, parameters $\beta_x^{(2)}$, $x = x_a, \ldots, x_b$, enable $\ln(m(x,t))$ at different ages to respond differently to the time-varying index, and as such, the expected rates of mortality improvement at different ages are allowed to be different. Finally, $\epsilon(x,t)$ represents the observation error.

2.1.2. The Cairns-Blake-Dowd model

The Cairns-Blake-Dowd model is defined as

$$\ln\left(\frac{q(x,t)}{1-q(x,t)}\right) = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \epsilon(x,t).$$

The quantity being modeled here is $q(x,t)$, the one-year conditional probability of death for age x and year t. It is defined as the probability that an individual aged x exact at time $t-1$ (i.e., the beginning of year t) dies during the time interval of $[t-1, t)$. Being a probability, this quantity is bounded between zero and one, and for this reason a logit transform is used.[1]

The model contains two (instead of one) time-varying indexes, $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. It assumes that in any given year t, $q(x,t)$ after a logit transform is a linear function of age x, with $\kappa_t^{(1)}$ being the intercept and $\kappa_t^{(2)}$ being the gradient. A reduction in $\kappa_t^{(1)}$ represents a level shift in the curve of $\ln(q(x,t)/(1-q(x,t))$ against x, or equivalently speaking, an equal improvement in mortality (in logit scale) at all ages. On the other hand, a reduction in $\kappa_t^{(2)}$ represents a flattening of the curve of $\ln(q(x,t)/(1-q(x,t))$ against x, or equivalently speaking, a more substantial improvement in mortality (in logit scale) at higher ages (ages above the average age $\bar{x}$). In many applications, $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are assumed to follow a bivariate random walk with a constant drift vector.

2.1.3. Model selection

Models for $\ln(m(x,t))$ and $\ln(q(x,t)/(1-q(x,t))$ are not nested. However, we may still compare them by considering their deviance residuals, as suggested by Cairns et al. (2009) and Li et al. (2015).

As the conclusion of the model selection process is heavily data-dependent, to maximize the applicability of our research, we derive our proposed hedging strategies on the basis of a general representation that encompasses most of the discrete-time stochastic mortality models in the literature.

2.2. The general state-space representation

As pointed out by Fung et al. (2017, 2018) and Liu and Li (2016, 2021), most discrete-time stochastic mortality models in the literature, including those mentioned in the previous sub-section, can be expressed in a state-space form. The general state-space representation of stochastic mortality models is composed of an observation equation and a transition equation. The former links the observed values of either $\ln(m(x,t))$ or $\ln(q(x,t)/(1-q(x,t))$ to the time-related and/or year-of-birth-related indexes, whereas the latter captures the random evolution of the time-related and/or year-of-birth-related[2] indexes.

The observation equation can be expressed as follows:

$$\vec{y}(t) = \vec{d} + \mathbf{B}\vec{\alpha}(t) + \vec{\epsilon}(t), \tag{1}$$

where $\vec{y}(t)$ is the vector of observations (the vector of the observed values of $\ln(m(x,t))$ or $\ln(q(x,t)/(1-q(x,t))$ for $x = x_a, \ldots, x_b$), $\vec{d}$ is a constant vector (which is either a vector of zeros or a vector of age-effect parameters), $\vec{\alpha}(t)$ is the vector of hidden states (the time-related and/or year-of-birth-related indexes), $\mathbf{B}$ is the design matrix (which determines how the observations are related to the hidden states) and $\vec{\epsilon}(t)$ is the vector of observation errors at time t.

The transition equation characterizes the random evolution of the vector of hidden states through a first-order Markov process, given by

$$\vec{\alpha}(t) = \vec{c} + \mathbf{A}\vec{\alpha}(t-1) + \vec{\eta}(t), \tag{2}$$

where $\vec{c}$ is a constant vector (the vector of constant drifts), $\mathbf{A}$ is the transition matrix that captures the serial dependence of the hidden states, and $\vec{\eta}(t)$ is the vector of random innovations. It is assumed that $\vec{\eta}(t)$ has a zero mean vector, and that $\vec{\eta}(t)$ and $\vec{\eta}(s)$ are uncorrelated if $t \neq s$.

[1] The logit transform of a quantity y is $\ln(y/(1-y))$.
[2] A year-of-birth related index is one that depends on $t-x$.

2.3. Distributional assumptions

In the literature, it is often assumed that the innovation vector $\vec{\eta}(t)$ is normally distributed. However, as discussed in Section 1 (and demonstrated in Section 6), such a distributional assumption may be inadequate when the model is used for risk management purposes.

Due to the inadequacy of the normality assumption, in this paper, we assume instead the innovation vector $\vec{\eta}(t)$ follows the multivariate skew-*t* distribution introduced by Azzalini and Capitanio (2003), a four-parameter distribution that incorporates not only mean and variance but also skewness and excess kurtosis.

We use

$$\vec{Y} \sim \mathrm{St}_n(\vec{\xi}, \mathbf{\Omega}, \vec{\zeta}, \nu)$$

to denote an n-dimensional random vector $\vec{Y}$ which follows the multivariate skew-*t* distribution introduced by Azzalini and Capitanio (2003), with a location parameter of $\vec{\xi}$ (an n-by-1 constant vector), a dispersion parameter of $\mathbf{\Omega}$ (a full rank symmetric n-by-n constant matrix), a skewness parameter of $\vec{\zeta}$ (an n-by-1 constant vector) and ν (a constant scalar) degrees of freedom. The density function of $\vec{Y}$ can be expressed as follows:

$$f_{\vec{Y}}(\vec{y}) = 2t_n(\vec{y};\nu)T_1\left(\vec{\zeta}'\mathbf{\Lambda}^{-1}(\vec{y}-\vec{\xi})\left(\frac{\nu+n}{(\vec{y}-\vec{\xi})'\mathbf{\Omega}^{-1}(\vec{y}-\vec{\xi})+\nu}\right)^{1/2};\nu+n\right),$$

where

$$\mathbf{\Lambda} = \mathrm{diag}(\omega_{1,1},\ldots,\omega_{n,n})^{1/2}$$

with $\omega_{i,i}$ representing the i-th diagonal element in $\mathbf{\Omega}$, $T_1(\cdot;\nu+n)$ denotes the distribution function of the univariate (symmetric) Student's t distribution with $\nu+n$ degrees of freedom, and

$$t_n(\vec{y};\nu) = \frac{\Gamma\{(\nu+n)/2\}}{|\mathbf{\Omega}|^{1/2}(\pi\nu)^{n/2}\Gamma(\nu/2)}\left(1+\frac{(\vec{y}-\vec{\xi})'\mathbf{\Omega}^{-1}(\vec{y}-\vec{\xi})}{\nu}\right)^{-(\nu+n)/2},$$

with Γ representing the gamma function, is the density function of a multivariate t distribution with ν degrees of freedom.

Noting that $\vec{\eta}(t)$ should have a zero mean vector and that $\mathrm{E}(\vec{\eta}(t)) \neq \vec{\xi}$ if $\vec{\eta}(t) \sim \mathrm{St}_n(\vec{\xi}, \mathbf{\Omega}, \vec{\zeta}, \nu)$, the location parameter $\vec{\xi}$ is set to a *non-zero* vector such that the resulting expectation of $\vec{\eta}(t)$ is a zero mean vector. The moments of $\vec{\eta}(t)$ play a crucial role in our proposed hedging strategies, which aim to optimize the moments of the hedged position. Azzalini and Capitanio (2003) provided analytical expressions for various moments of random vectors following the multivariate skew-*t* distribution with $\vec{\xi}=0$, but their results are insufficient for the context of our research because, as previously mentioned, $\vec{\xi}$ should be a *non-zero* vector when the distribution is used to model innovation vectors. Therefore, we reformulate the results of Azzalini and Capitanio (2003) to obtain analytical expressions for the moments of random variables following $\mathrm{St}_n(\vec{\xi}, \mathbf{\Omega}, \vec{\zeta}, \nu)$ with $\vec{\xi} \neq \vec{0}$. These analytical expressions and their derivations are presented in Appendix A.

3. Set-up

In this section, we present the set-up for our proposed hedging strategies. Similar to most of the existing longevity hedging strategies, we focus on trend risk only, so the randomness arising from the observation error $\epsilon(x,t)$ for any x and t is ignored. To facilitate our presentation, we use $\tilde{q}(x,t)$ and $\tilde{m}(x,t)$ to represent $q(x,t)$ and $m(x,t)$, respectively, when the observation error $\epsilon(x,t)$ is ignored (being set to zero).

3.1. The liability being hedged

The liability being hedged is a portfolio of T_L-year temporary life annuities that are issued to individuals who are aged x_L at time t_0. At the end of each year during its term, the annuity makes a payment of \$1 to the annuitant if the annuitant is alive. We assume that the annuity portfolio is sufficiently large, so that micro longevity risk is negligible.

We let L be the sum of the per contract cash flows in time-t_0 dollars, with survivorship being taken into account. We have

$$L = \sum_{u=1}^{T_L}\left(e^{-ru}\prod_{s=1}^{u}\tilde{p}(x_L+s, t_0+s)\right), \tag{3}$$

where r is the interest rate at which cash flows are discounted, and

$$\tilde{p}(x,t) := 1-\tilde{q}(x,t) \approx \exp(-\tilde{m}(x,t))$$

represents the probability that an individual survives to age x, given that he/she is alive and aged $x-1$ exact at the beginning of year t.[3]

We use $V_L(t)$ to represent the time-t value of the entire (paid and unpaid) annuity liability on a per contract basis, measured in time-t_0 dollars. We can express $V_L(t)$ as

$$V_L(t) = \mathrm{E}(L|\mathcal{F}_t), \quad t \geq t_0,$$

where $\mathcal{F}_t$ denotes the information up to and including time t. It is clear from the definition that $V_L(t)$ is a known constant in either one or both of the following conditions:

[3] The approximation that $\tilde{q}(x,t) \approx 1-\exp(-\tilde{m}(x,t))$ is exact when the force of mortality is constant between two consecutive integer ages.

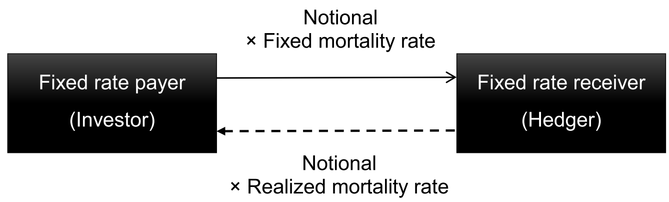

Fig. 1. An illustration of the cash flows exchanged between the two counterparties of a q-forward at maturity.

1. when it is viewed at any time point at or after time t (i.e., given $\mathcal{F}_s$ for $s \geq t$);
2. when it is viewed at a time point at or beyond time $t_0 + T_L$, as all possible liability cash flows should have been made by time $t_0 + T_L$.

Otherwise (i.e., given $\mathcal{F}_s$ for $s < \min(t, t_0 + T_L)$), $V_L(t)$ is a random variable.

3.2. The hedging instruments

The hedging instruments used in our proposed hedging strategies are q-forwards. A q-forward is a zero-coupon swap with the floating leg being the realized mortality rate at a certain age (the reference age) at the time when the swap matures, and the fixed leg being the forward mortality rate which is pre-determined when the swap is launched. The exchange of cash flows between counterparties at maturity is illustrated in Fig. 1.

Hedgers wishing to reduce their exposure to longevity risk (the risk that future mortality is lighter than expected) should participate in q-forwards as a fixed-rate receiver. With such a position, hedgers will receive net payments from the fixed-rate payer if future mortality turns out to be lighter than expected, and the net payments received can offset the correspondingly higher pension/annuity liabilities.

Let us suppose that the hedge portfolio consists of m q-forwards, all of which are launched at the same time point t_I, where $t_I \geq t_0$. For $j = 1, \ldots, m$, we denote the jth q-forward's reference age by x_j, time-to-maturity by T_j, and forward mortality rate by $q^f(x_j, t_I + T_j)$.

For consistency reasons, regardless of when the q-forward is launched, we measure its payoff in time-t_0 dollars, where t_0, as previously defined, is the time point when the life annuities are issued. Per \$1 notional, the payoff of the j-th q-forward at maturity, measured in time-t_0 dollars and from the fixed-rate receiver's perspective, is given by

$$H(j, t_I) = e^{-r\times(t_I - t_0 + T_j)}(q^f(x_j, t_I + T_j) - \tilde{q}(x_j, t_I + T_j)). \tag{4}$$

We further use

$$V_H(t; j, t_I) = \mathrm{E}(H(j, t_I)|\mathcal{F}_t), \quad t \geq t_I,$$

to represent the time-t value of the jth q-forward, measured in time-t_0 dollars. Of course, $V_H(t; j, t_I)$ is non-random if $t \geq t_I + T_j$, because in this case the mortality rate to which the floating leg is linked is already realized. Also, according to its definition, $V_H(t; j, t_I)$ is non-random when it is viewed at any time point at or beyond time t (i.e., given $\mathcal{F}_s$ for $s \geq t$). Otherwise (i.e., given $\mathcal{F}_s$ for $s < \min(t, t_I + T_J)$), $V_H(t; j, t_I)$ is a random variable.

Our proposed hedging strategies take costs of hedging into account. In this regard, the risk premium demanded by the counterparty (i.e., the fixed-rate payer) should factor into the forward mortality rate. Following Li and Hardy (2011), we use the following formula to determine the forward mortality rate for the jth q-forward in the hedge porfolio:

$$q^f(x_j, t_I + T_j) = (1 - T_j \times \ell \times v(x_j)) \times \hat{q}(x_j, t_I + T_j), \tag{5}$$

where

- $\hat{q}(x_j, t_I + T_j)$ represents the best estimate of $q(x_j, t_I + T_j)$ as of the issue date t_I, obtained by setting $\epsilon(x, s)$ and $\vec{\eta}(s)$ for all x and $s = t_I + 1, \ldots, t_I + T_j$ to zero.
- $v(x_j)$ is the estimated volatility of the yearly changes in the death probability at age x_j, and
- $\ell > 0$ is the market price of risk, which reflects the compensation to the fixed-rate payer for taking on longevity risk exposures from the fixed-rate receiver.

3.3. Formulating longevity hedges

In the insurance industry, the biggest concern of an insurer is the randomness associated with the value of its liability in one year, because in many jurisdictions, the capital requirement for an insurer is determined by a risk measure that is calculated from the distribution of the insurer's liability values in one year. For instance, Solvency II requires any insurer operating in the European Union to hold solvency capital that is no smaller than the Value-at-Risk (VaR) of the insurer's liability at the 99.5% confidence level. The Swiss Solvency Test, which is implemented in Switzerland, is broadly similar, but is based on the tail-VaR (also known as conditional tail expectation) at the 99% confidence level instead.

Accordingly, our goal is to mitigate the randomness associated with the value of the annuity liability in one year. Suppose that we formulate a longevity hedge at time t, and that for simplicity all of the m q-forwards used in the hedge are freshly launched at the same time. Our goal means that the hedger should acquire $N_{j,t}$ notional amount of the jth q-forward, for $j = 1, \ldots, m$, such that the randomness associated with

$$V_P(t+1;\vec{N}_t) = V_L(t+1) - \sum_{j=1}^{m} N_{j,t} V_H(t+1;j,t),$$

where $\vec{N}_t = (N_{1,t},\dots,N_{m,t})'$, is mitigated compared to that associated with $V_L(t+1)$.

In studying risk mitigation, we consider the first four moments of $V_P(t+1;\vec{N}_t)$, given information up to and including time t. More specifically, the notional amounts $N_{1,t},\dots,N_{m,t}$ are determined by

- minimizing
$$M^{(1)}(\vec{N}_t) = \mathrm{E}[V_P(t+1;\vec{N}_t)|\mathcal{F}_t],$$
which represents the mean of the hedged position, so that the (average) cost of hedging is kept low,
- minimizing
$$M^{(2)}(\vec{N}_t) = \mathrm{E}[(V_P(t+1;\vec{N}_t) - M^{(1)}(\vec{N}_t))^2|\mathcal{F}_t],$$
which represents the variance of the hedged position, so that the volatility of the liability values (net of the values of the hedging instruments) is kept low,
- minimizing
$$M^{(3)}(\vec{N}_t) = \frac{\mathrm{E}[(V_P(t+1;\vec{N}_t) - M^{(1)}(\vec{N}_t))^3|\mathcal{F}_t]}{(M^{(2)}(\vec{N}_t))^{3/2}},$$
which reflects the skewness of the hedged position, so that the probability of having extremely large liability values (net of the values of the hedging instruments) is kept low, and
- minimizing
$$M^{(4)}(\vec{N}_t) = \frac{\mathrm{E}[(V_P(t+1;\vec{N}_t) - M^{(1)}(\vec{N}_t))^4|\mathcal{F}_t]}{(M^{(2)}(\vec{N}_t))^{2}},$$
which indicates the kurtosis of the hedged position, so that the tails of the hedged position's distribution are kept thin.

The formulation of the optimization problem above is detailed in the next section.

While the benefits of reducing mean and variance are obvious, the rationales behind minimizing skewness and kurtosis merit an illustration. Everything being equal, hedgers typically aim to reduce the right tail of V_P. To illustrate the effect of reducing skewness to the right tail of the distribution of V_P, in Fig. 2 we compare two skew-normal distributions with the same mean (0) and variance (1), but different skewness (−0.45 and −0.93, respectively). The distribution with a larger (less negative) skewness resembles the somewhat negatively skewed unhedged position that we are confronting, while the other distribution resembles a hedged position with a reduced (more negative) skewness but the same lower moments. The distribution with a smaller skewness clearly has a lower 99.5th percentile, highlighting the benefit of reducing skewness when considering the far tail. However, the diagram also suggests that reducing skewness is not necessarily beneficial if we are not looking far enough into the right tail. In particular, the 80th percentile of the distribution with a smaller skewness is higher compared to the other distribution.

To demonstrate the benefit of reducing kurtosis, in Fig. 3 we compare two skew-t distributions with the same mean (0), variance (1) and skewness (−0.9), but different kurtosis. The distribution with a smaller kurtosis (which resembles a hedged position with a reduced kurtosis) comes with a lower 99.5th percentile compared to the other distribution (which resembles the unhedged position), suggesting that a longevity hedge that minimizes kurtosis has a strong potential to reduce Value-at-Risk at a high confidence level.

Finally, we remark that in the expressions for $M^{(3)}(\vec{N}_t)$ and $M^{(4)}(\vec{N}_t)$, the denominators ($(M^{(2)}(\vec{N}_t))^{3/2}$ and $M^{(2)}(\vec{N}_t))^2$, respectively) are used for standardization purposes. In addition to standardized moments, centered moments are also considered in our investigation. It is found that the results produced on the basis of standardized moments and centered moments are similar.

4. Polynomial goal programming

In this section, we present a formal formulation of the optimization problem set out in Section 3.3. Let us begin by stating the constraints we use in the optimization problem:

- $N_{j,t} \geq 0$ for all $j = 1,\dots,m$
This constraint means that the hedger cannot participate in a q-forward as a fixed-rate payer. We impose this constraint because, as discussed in Section 3.2, one should participate in a q-forward as a fixed-rate receiver if he/she wishes to hedge his/her exposure to longevity risk.
- $\sum_{j=1}^{m} N_{j,t}\mathrm{E}[V_H(t+1;j,t)|\mathcal{F}_t] \leq C \times \mathrm{E}[V_L(t+1)|\mathcal{F}_t]$
Recall that we assume that the q-forwards used in the hedge come with a cost, with a market price of risk $\ell > 0$. We further assume that the hedger has a budget of C (as a fraction of the value of the annuity liability) for hedging. This constraint makes our set-up a closer resemblance to real-life situations, as in practice hedgers do have a budget for hedging (see, e.g., Callan Institute, 2017).

Accordingly, the optimization problem in question can be formulated as follows:

$$\begin{cases} \text{Minimize} & M^{(i)}(\vec{N}_t),\ i = 1,2,3,4 \\ \text{Subject to} & N_{j,t} \geq 0,\ j = 1\dots,m \\ & \sum_{j=1}^{m} N_{j,t}\mathrm{E}[V_H(t+1;j,t)|\mathcal{F}_t] \leq C \times \mathrm{E}[V_L(t+1)|\mathcal{F}_t] \end{cases}$$

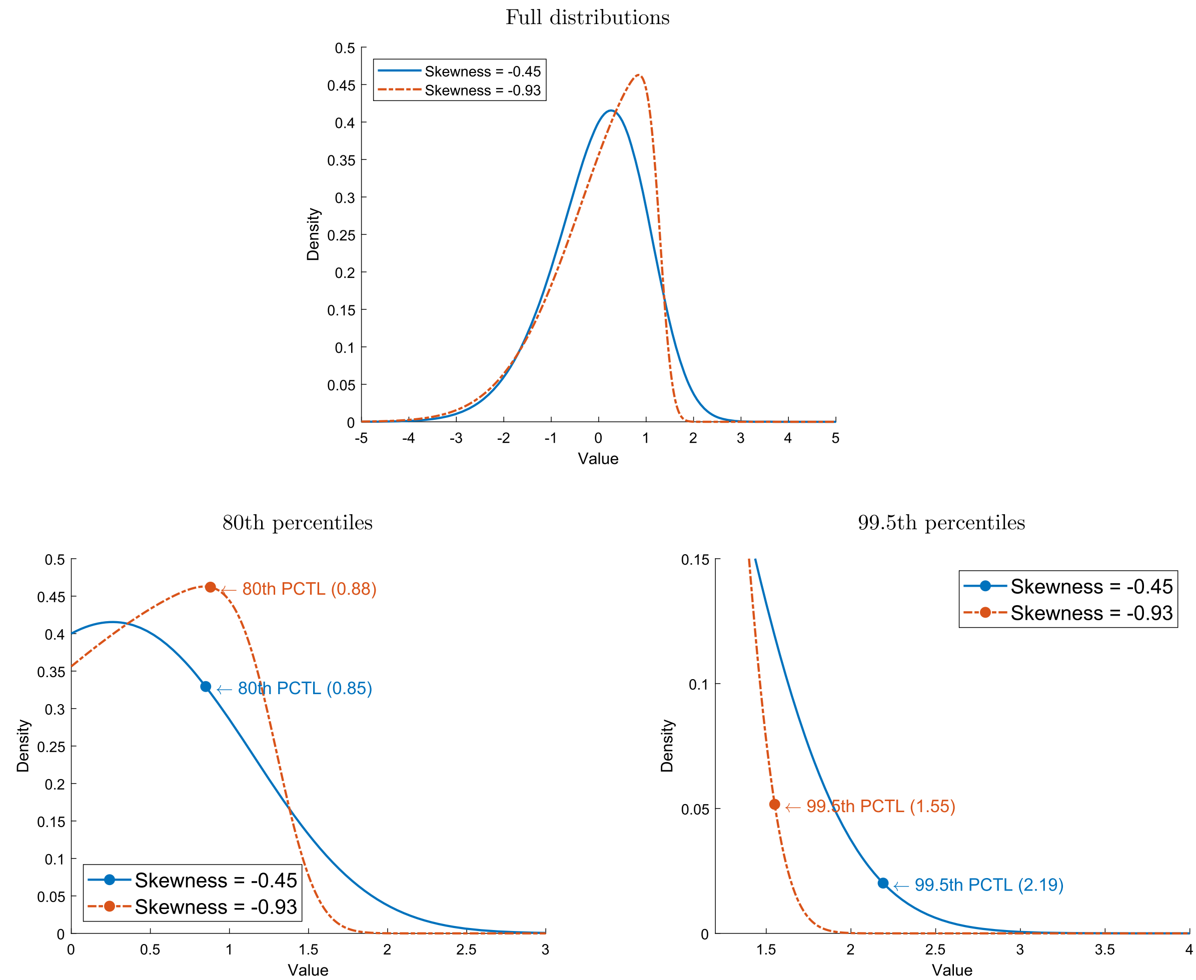

Fig. 2. Illustrating the effect of reducing skewness on the tail of the distribution of V_P: skew-normal distributions with the same mean (0) and variance (1), but different skewness (−0.45 and −0.93, respectively).

To solve the optimization problem, we adapt the polynomial goal programming approach considered by Lai et al. (2006). This approach makes the optimization problem solvable by consolidating the four objectives into one single objective function.

In the polynomial goal programming approach, we first determine the aspired levels of the four moments. Concerning the optimization problem we are confronting, the aspired level of a moment is the lowest achievable level of the moment when the objectives concerning the other moments are not taken into consideration. The aspired levels of the four moments can be determined by solving the following sub-problem independently for each $i = 1, 2, 3, 4$:

$$\begin{cases} \text{Minimize} & M^{(i)}(\vec{N}_t) \\ \text{Subject to} & N_{j,t} \geq 0,\ j = 1 \ldots, m \\ & \sum_{j=1}^{m} N_{j,t}\mathrm{E}[V_H(t+1; j, t)|\mathcal{F}_t] \leq C \times \mathrm{E}[V_L(t+1)|\mathcal{F}_t] \end{cases}$$

We use $\mathcal{M}^{(i)}$ to represent the aspired level of $M^{(i)}(\vec{N}_t)$ (i.e., the solution to the sub-problem above).

Next, we consider the deviation between $M^{(i)}(\vec{N}_t)$ and $\mathcal{M}^{(i)}$ for each $i = 1, 2, 3, 4$, measured with the following metric:

$$\left| \frac{M^{(i)}(\vec{N}_t) - \mathcal{M}^{(i)}}{\mathcal{M}^{(i)}} \right|.$$

We further allow the hedger to have asymmetric preferences concerning the four moments. A constant λ_i is used to represent the hedger's emphasis on the ith moment, where $i = 1, 2, 3, 4$. The higher the value of λ_i is, the more the hedger concerns about the ith moment. The values of λ_i for $i = 1, 2, 3, 4$ should be specified before executing the optimization.

On the basis of the deviation metric, the aspired levels, and the pre-specified values of λ_i for $i = 1, 2, 3, 4$, the original optimization problem can be reformulated as follows:

$$\begin{cases} \text{Minimize} & \lambda_1 \left| \frac{g_1}{\mathcal{M}^{(1)}} \right| + \lambda_2 \left| \frac{g_2}{\mathcal{M}^{(2)}} \right| + \lambda_3 \left| \frac{g_3}{\mathcal{M}^{(3)}} \right| + \lambda_4 \left| \frac{g_4}{\mathcal{M}^{(4)}} \right| \\ \text{Subject to} & M^{(i)}(\vec{N}_t) - g_i = \mathcal{M}^{(1)}, i = 1, 2, 3, 4 \\ & N_{j,t} \geq 0,\ j = 1 \ldots, m \\ & \sum_{j=1}^{m} N_{j,t}\mathrm{E}[V_H(t+1; j, t)|\mathcal{F}_t] \leq C \times \mathrm{E}[V_L(t+1)|\mathcal{F}_t] \\ & g_i \geq 0,\ i = 1, 2, 3, 4 \end{cases}$$

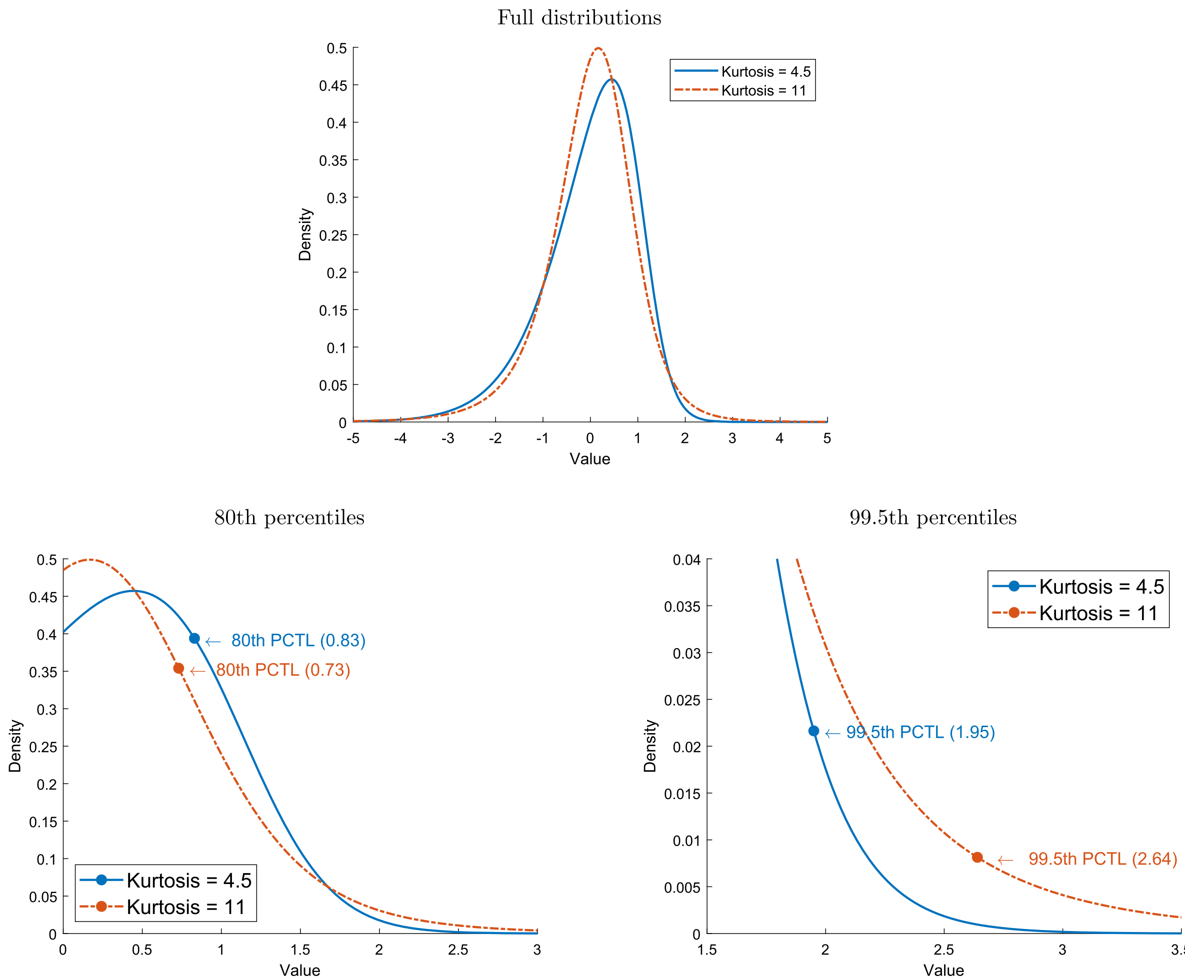

Fig. 3. Illustrating the effect of reducing skewness on the tail of the distribution of V_P: skew-t distributions with the same mean (0), variance (1) and skewness (-0.9), but different kurtosis (4.5 and 11, respectively).

In the above, g_i is the goal variable measuring how less favorable $M^{(i)}(\vec{N}_t)$ is compared to its aspired level $\mathcal{M}^{(i)}$, and as such, the last constraint $g_i \geq 0$ is imposed. On solving this optimization problem, we obtain the optimized notional amounts, $\hat{N}_{j,t}$ for $j = 1, \ldots, m$, of the q-forwards in the hedge portfolio.

5. Implementation

5.1. The required inputs

In this section, we discuss how the polynomial goal programming approach can be implemented in the context of this research. We begin by identifying the inputs required for the implementation.

Let us consider

$$(V_L(t+1), V_H(t+1; 1, t), \ldots, V_H(t+1; m, t))',$$

the vector of the time-$(t+1)$ values of the annuity liability and the m q-forwards (which are freshly issued at time t). We use

$$\mathbf{M}_t = \{\mu_{i,t}; i = 1, \ldots, m+1\},$$
$$\mathbf{V}_t = \{\sigma_{ij,t}; i, j = 1, \ldots, m+1\},$$
$$\mathbf{S}_t = \{s_{iju,t}; i, j, u = 1, \ldots, m+1\}$$

and

$$\mathbf{K}_t = \{k_{ijuv,t}; i, j, u, v = 1, \ldots, m+1\}$$

to represent the first four moments of this vector, given information up to and including time t, respectively.[4] It can be shown that the first four moments of the hedged position $V_P(t+1; \vec{N}_t)$ can be rewritten in terms of $\mu_{i,t}$, $\sigma_{ij,t}$, $s_{iju,t}$ and $k_{ijuv,t}$ as follows:

[4] The third and fourth moments represented by these notations are pre-standardized.

$$M^{(1)}(\vec{N}_t) = \mathrm{E}[V_P(t+1;\vec{N}_t)|\mathcal{F}_t] = \sum_{i=1}^{m+1} w_{i,t}\mu_{i,t},$$

$$M^{(2)}(\vec{N}_t) = \mathrm{E}[(V_P(t+1;\vec{N}_t) - M^{(1)}(\vec{N}_t))^2|\mathcal{F}_t] = \sum_{i=1}^{m+1}\sum_{j=1}^{m+1} w_{i,t}w_{j,t}\sigma_{ij,t},$$

$$M^{(3)}(\vec{N}_t) = \frac{\mathrm{E}[(V_P(t+1;\vec{N}_t) - M^{(1)}(\vec{N}_t))^3|\mathcal{F}_t]}{(M^{(2)}(\vec{N}_t))^{3/2}} = \frac{\sum_{i=1}^{m+1}\sum_{j=1}^{m+1}\sum_{u=1}^{m+1} w_{i,t}w_{j,t}w_{u,t}s_{iju,t}}{\left(\sum_{i=1}^{m+1}\sum_{j=1}^{m+1} w_{i,t}w_{j,t}\sigma_{ij,t}\right)^{3/2}}$$

and

$$M^{(4)}(\vec{N}_t) = \frac{\mathrm{E}[(V_P(t+1;\vec{N}_t) - M^{(1)}(\vec{N}_t))^4|\mathcal{F}_t]}{(M^{(2)}(\vec{N}_t))^{2}} = \frac{\sum_{i=1}^{m+1}\sum_{j=1}^{m+1}\sum_{u=1}^{m+1}\sum_{v=1}^{m+1} w_{i,t}w_{j,t}w_{u,t}w_{v,t}k_{ijuv,t}}{\left(\sum_{i=1}^{m+1}\sum_{j=1}^{m+1} w_{i,t}w_{j,t}\sigma_{ij,t}\right)^{2}},$$

where $w_{1,t} = 1$, and $w_{i,t} = -N_{i-1,t}$ for $i = 2, \ldots, m+1$.

It is now clear that the optimization problem depends critically on $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$. In principle, we can obtain these four arrays of values using simulations, but the computation effort entailed may reduce the practicality of the proposed hedging strategies. To get a better idea about the computational effort needed, let us suppose that we wish to evaluate the performance of a hedge that is established at time t. Without analytical solutions, we first need to simulate a large number of sample paths of future mortality rates (or, equivalently speaking, sample paths of innovation vectors) from time t. Then, for each of the simulated sample paths, we need another set of simulated sample paths to calculate $\mathbf{M}_{t+1}$, $\mathbf{V}_{t+1}$, $\mathbf{S}_{t+1}$ and $\mathbf{K}_{t+1}$, thereby creating the situation of 'simulations on simulations'.

To reduce the computational effort needed, we propose to calculate $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$ using approximate analytical formulas. In the next two sub-sections, we discuss how such approximate analytical formulas may be obtained.

5.2. Approximating L and $H(j,t)$

The approximate analytical formulas for calculating $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$ are based on first-order Taylor's expansions of L and $H(j,t)$.

Let us first focus on L, the sum of the annuity liability's cash flows measured in time-t_0 dollars, taken into account of survivorship. When calculating $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$ (i.e., calibrating the hedge at time t), we consider a first-order Taylor's expansion of L around $\vec{\eta}(t+1), \ldots, \vec{\eta}(t_0+T_L)$ (the innovation vectors between time $t+1$ and the time when the annuity liability runs off completely). More specifically, when calculating $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$ for $t_0 \le t < t_0 + T_L$, the approximation of L takes the following form:

$$L \approx l := \hat{L} + \sum_{s=t+1}^{t_0+T_L}\left(\frac{\partial L}{\partial \vec{\eta}(s)}\right)'\left(\vec{\eta}(s) - \hat{\vec{\eta}}(s)\right), \tag{6}$$

where $\hat{\vec{\eta}}(s) = \mathrm{E}(\vec{\eta}(s)|\mathcal{F}_t) = \vec{0}$ for all $s = t+1, \ldots, t_0+T_L$, $\hat{L}$ is the value of L calculated by using the realized values of $\vec{\eta}(t_0), \ldots, \vec{\eta}(t)$ and setting $\vec{\eta}(s) = \hat{\vec{\eta}}(s) = \vec{0}$ for $s = t+1, \ldots t_0 + T_L$, and $\partial L/\partial\vec{\eta}(s)$ is the partial derivative of L with respect to $\vec{\eta}(s)$, evaluated at the realized values of $\vec{\eta}(t_0), \ldots, \vec{\eta}(t)$ and $\vec{\eta}(u) = \hat{\vec{\eta}}(u) = \vec{0}$ for $u = t+1, \ldots, t_0+T_L$.

Using equation (6), we obtain the following approximation of the time-$(t+1)$ value of the annuity liability (when viewed at time t when $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$ are calculated):

$$V_L(t+1) = \mathrm{E}(L|\mathcal{F}_{t+1}) \approx V_l(t+1) := \mathrm{E}(l|\mathcal{F}_{t+1}) = \hat{L} + \left(\frac{\partial L}{\partial \vec{\eta}(t+1)}\right)'\vec{\eta}(t+1). \tag{7}$$

This equation involves only $\vec{\eta}(t+1)$ for the following two reasons. First, given information up to and including time $t+1$, the expectation of $\vec{\eta}(s)$ for any $s > t+1$ is a zero vector. Second, given information up to and including time $t+1$, $\vec{\eta}(t+1)$ is a known vector.

An analytical expression (written in terms of the parameters in the general state-space representation of stochastic mortality models) for computing $\partial L/\partial\vec{\eta}(t+1)$ is derived in Appendix B.

We then consider $H(j,t)$, the payoff (measured in time-t_0 dollars) of the jth q-forward in the hedge portfolio that is formed at $t \ge t_0$.[5] When calculating $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$ (i.e., calibrating the hedge at time t), we consider a first-order Taylor's expansion of $H(j,t)$ around $\vec{\eta}(t+1), \ldots, \vec{\eta}(t+T_j)$ (the innovation vectors between time $t+1$ and the time when the jth q-forward matures). In more detail, when calculating $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$ for $t_0 \le t < t_0 + T_L$, the approximation of $H(j,t)$ takes the following form:

$$H(j,t) \approx h(j,t) := \hat{H}(j,t) + \sum_{s=t+1}^{t+T_j}\left(\frac{\partial H(j,t)}{\partial \vec{\eta}(s)}\right)'\left(\vec{\eta}(s) - \hat{\vec{\eta}}(s)\right), \tag{8}$$

[5] In our set-up, it is assumed that all q-forwards in the hedge portfolio formed at time $t \ge t_0$ are freshly launched at time t (i.e., $t_l = t$).

for $j = 1, \ldots, m$, where $\hat{\vec{\eta}}(s) = \mathrm{E}(\vec{\eta}(s)|\mathcal{F}_t) = \vec{0}$ for all $s = t+1, \ldots, t+T_j$, $\hat{H}(j,t)$ is the value of $H(j,t)$ calculated by setting $\vec{\eta}(s) = \hat{\vec{\eta}}(s) = \vec{0}$ for all $s = t+1, \ldots, t+T_j$, and $\partial H(j,t)/\partial \vec{\eta}(s)$ is the partial derivative of $H(j,t)$ with respect to $\vec{\eta}(s)$, evaluated at $\vec{\eta}(u) = \hat{\vec{\eta}}(u) = \vec{0}$ for all $u = t+1, \ldots, t+T_j$.

Using equation (8), we get the following approximation of the time-$(t+1)$ value of the jth q-forward in the hedge portfolio formed at time t:

$$V_H(t+1; j,t) = \mathrm{E}[H(j,t)|\mathcal{F}_{t+1}] \approx V_h(t+1; j,t) := \mathrm{E}[h(j,t)|\mathcal{F}_{t+1}] = \hat{H}(j,t) + \left(\frac{\partial H(j,t)}{\partial \vec{\eta}(t+1)}\right)' \vec{\eta}(t+1).$$

This equation involves only $\vec{\eta}(t+1)$, because, given information up to and including time $t+1$, the expectation of $\vec{\eta}(s)$ for any $s > t+1$ is a zero vector and $\vec{\eta}(t+1)$ is a known vector. An analytical expression (written in terms of the parameters in the general state space representation of stochastic mortality models) for computing $\partial H(j,t)/\partial \vec{\eta}(t+1)$ is derived in Appendix B.

Cairns (2011) and Zhou and Li (2017) considered similar first-order approximations to approximate values of annuity liabilities and q-forwards, and confirmed the accuracy of the approximations for thousands of different realizations of $\vec{\eta}(t+1)$. A further examination of the accuracy of the approximation is provided in Section 6 where numerical illustrations are presented.

5.3. Approximate analytical expressions for $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$

Given $V_l(t+1)$ as an approximate for $V_L(t+1)$ and $V_h(t+1; j,t)$ as an approximate for $V_H(t+1; j,t)$, for $j = 1, \ldots, m$, we approximate $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$ using the first four moments of the random vector

$$(V_l(t+1), V_h(t+1; 1,t), \ldots, V_h(t+1; m,t))', \tag{9}$$

given information up to and including time t, respectively. As shown below, the approximation of $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$ can be calculated analytically.

First, assuming that the stochastic mortality model used contains n innovations at a given time point, we can rewrite $V_l(t+1)$ and $V_h(t+1; j,t)$ for $j = 1, \ldots, m$ in scalar forms as follows:

$$V_l(t+1) = \hat{L} + \sum_{k=1}^{n} (D_k(0, t+1) \times \eta_k(t+1)) \tag{10}$$

and

$$V_h(t+1; j,t) = \hat{H}(j,t) + \sum_{k=1}^{n} (D_k(j, t+1) \times \eta_k(t+1)), \quad j = 1, \ldots, m, \tag{11}$$

where

$$D_k(0, t+1) = \frac{\partial L}{\partial \eta_k(t+1)}$$

and

$$D_k(j, t+1) = \frac{\partial H(j,t)}{\partial \eta_k(t+1)}, \quad j = 1, \ldots, m.$$

Following the discussion in Section 2.3, given $\mathcal{F}_t$, the innovation vector at time $t+1$, that is, $\vec{\eta}(t+1) = (\eta_1(t+1), \ldots, \eta_n(t+1))'$, is assumed to follow $\mathrm{St}_n(\vec{\xi}, \boldsymbol{\Omega}, \vec{\zeta}, \nu)$. Using the distributional assumption on $\vec{\eta}(t+1)$ and equations (10) and (11), the first four moments of the random vector in expression (9) and hence the approximates of $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$ can be obtained as follows:

- Approximation of $\mathbf{M}_t$
  By taking expectation on equations (10) and (11) given $\mathcal{F}_t$, it can be shown that the expectation of the vector in expression (9) given $\mathcal{F}_t$ is $(\hat{L}, \hat{H}(1,t), \ldots, \hat{H}(m,t))'$. Therefore, we approximate $\mathbf{M}_t$ as $(\hat{L}, \hat{H}(1,t), \ldots, \hat{H}(m,t))'$, in which all elements, as explained in the previous sub-section, can be computed analytically.
- Approximation of $\mathbf{V}_t$
  We can approximate $\sigma_{ij,t}$ (i.e., the (i,j)th element in $\mathbf{V}_t$) as follows:

$$\begin{aligned}\sigma_{ij,t} &\approx \mathrm{Cov}\left(\sum_{k_1=1}^{n} \left(D_{k_1}(i-1, t+1) \times \eta_{k_1}(t+1)\right), \sum_{k_2=1}^{n} \left(D_{k_2}(j-1, t+1) \times \eta_{k_2}(t+1)\right) \middle| \mathcal{F}_t\right) \\ &= \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \left(D_{k_1}(i-1, t+1) \times D_{k_2}(i-1, t+1) \times \mathrm{Cov}(\eta_{k_1}(t+1), \eta_{k_2}(t+1)|\mathcal{F}_t)\right),\end{aligned}$$

  for $i, j = 1, 2, \ldots, m+1$, where $\mathrm{Cov}(\eta_{k_1}(t+1), \eta_{k_2}(t+1)|\mathcal{F}_t)$ can be obtained analytically from equation (22) in Appendix A.
- Approximation of $\mathbf{S}_t$

We can approximate $s_{iju,t}$ (i.e., the (i,j,u)th element in $\mathbf{S}_t$) as follows:

$$\begin{aligned} s_{iju,t} &\approx \mathrm{E}\left[\left(\sum_{k_1=1}^{n}\left(D_{k_1}(i-1,t+1)\times\eta_{k_1}(t+1)\right)\right)\times\left(\sum_{k_2=1}^{n}\left(D_{k_2}(j-1,t+1)\times\eta_{k_2}(t+1)\right)\right)\right.\\ &\quad\left.\times\left(\sum_{k_3=1}^{n}\left(D_{k_3}(u-1,t+1)\times\eta_{k_3}(t+1)\right)\right)\middle|\mathcal{F}_t\right]\\ &=\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\sum_{k_3=1}^{n}\Big(D_{k_1}(i-1,t+1)\times D_{k_2}(j-1,t+1)\times D_{k_3}(u-1,t+1)\\ &\quad\times\mathrm{E}(\eta_{k_1}(t+1)\eta_{k_2}(t+1)\eta_{k_3}(t+1)|\mathcal{F}_t)\Big), \end{aligned}$$

for $i,j,u=1,2,\ldots,m+1$, where $\mathrm{E}(\eta_{k_1}(t+1)\eta_{k_2}(t+1)\eta_{k_3}(t+1)|\mathcal{F}_t)$ can be calculated analytically by using equation (23) in Appendix A.

- Approximation of $\mathbf{K}_t$

We can approximate $k_{ijuv,t}$ (i.e., the (i,j,u,v)th element in $\mathbf{K}_t$) as follows:

$$\begin{aligned} k_{ijuv,t} &\approx \mathrm{E}\left[\left(\sum_{k_1=1}^{n}\left(D_{k_1}(i-1,t+1)\times\eta_{k_1}(t+1)\right)\right)\times\left(\sum_{k_2=1}^{n}\left(D_{k_2}(j-1,t+1)\times\eta_{k_2}(t+1)\right)\right)\right.\\ &\quad\left.\times\left(\sum_{k_3=1}^{n}\left(D_{k_3}(u-1,t+1)\times\eta_{k_3}(t+1)\right)\right)\times\left(\sum_{k_4=1}^{n}\left(D_{k_4}(v-1,t+1)\times\eta_{k_4}(t+1)\right)\right)\middle|\mathcal{F}_t\right]\\ &=\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\sum_{k_3=1}^{n}\sum_{k_4=1}^{n}\Big(D_{k_1}(i-1,t+1)\times D_{k_2}(j-1,t+1)\times D_{k_3}(u-1,t+1)\\ &\quad\times D_{k_4}(v-1,t+1)\times\mathrm{E}(\eta_{k_1}(t+1)\eta_{k_2}(t+1)\eta_{k_3}(t+1)\eta_{k_4}(t+1)|\mathcal{F}_t)\Big), \end{aligned}$$

for $i,j,u,v=1,2,\ldots,m+1$, where $\mathrm{E}(\eta_{k_1}(t+1)\eta_{k_2}(t+1)\eta_{k_3}(t+1)\eta_{k_4}(t+1)|\mathcal{F}_t)$ can be computed analytically with equation (24) in Appendix A.

With the approximates of $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$, the constrained optimization problem formulated in Section 4 can be implemented without simulations.

6. Numerical illustrations

In this section, we demonstrate our theoretical results with a stochastic mortality model that is estimated to real historical data.

6.1. Model and data

We consider the data from the male population of England and Wales over the age range of $x_a=60$ to $x_b=89$ and the calibration window of $t_a=1950$ to $t_b=2016$. The data set (which encompasses death and exposure counts) are obtained from the Human Mortality Database (www.mortality.org).

In what follows, we describe the processes in which we (a) identify a stochastic mortality model for the data set, (b) choose a time-series process for the dynamics of the time-varying indexes, and (c) specify the distribution for the innovation vectors in the chosen time-series process.

6.1.1. Stochastic mortality models

The candidate models we consider are the Lee-Carter model and the Cairns-Blake-Dowd model, which are probably the most frequently considered stochastic mortality models in the literature. These two models are not nested, so we compare them by considering the deviance residuals produced by them when they are fitted to the data set under consideration using the method of Poisson maximum likelihood (Wilmoth, 1993).

The resulting deviance residuals[6] are displayed in Fig. 4. For both models, the deviance residuals are reasonably random and do not exhibit any substantial systematic patterns. While the deviance residuals suggest that both models are viable, we choose to use the Cairns-Blake-Dowd model on grounds that it is more interpretable and that it possesses properties that facilitate the construction of index-based longevity risk transfers (see Chan et al., 2014).

6.1.2. Time-series processes for the dynamics of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$

Fig. 5 displays the estimates of the time-varying indexes $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ in the Cairns-Blake-Dowd model within the calibration window. When modeling the dynamics of the indexes, we restrict ourselves to the class of vector autoregressive (VAR) processes, as the data series seems too short for non-linear multivariate time-series processes such as multivariate GARCH.

On the basis of sample autocorrelation matrices (SCCM; see Table 1[7]), it is found that $\{(\kappa_t^{(1)},\kappa_t^{(2)})'\}$ is not stationarity, but the first difference of it, i.e., $\{(\Delta\kappa_t^{(1)},\Delta\kappa_t^{(2)})'\}$ appears to be stationary. We then model $\{(\Delta\kappa_t^{(1)},\Delta\kappa_t^{(2)})'\}$ with the simplest VAR process, i.e., VAR(0).

[6] We refer the reader to Li (2013) for the definition of deviance residuals in the context of stochastic mortality modeling.

[7] Simplified SCCMs are presented (Tsay, 2010). In each simplified SCCM, '+' indicates a value greater than twice the estimated standard error, '−' denotes a value less than minus twice the estimated standard error and '·' represents an insignificant value.

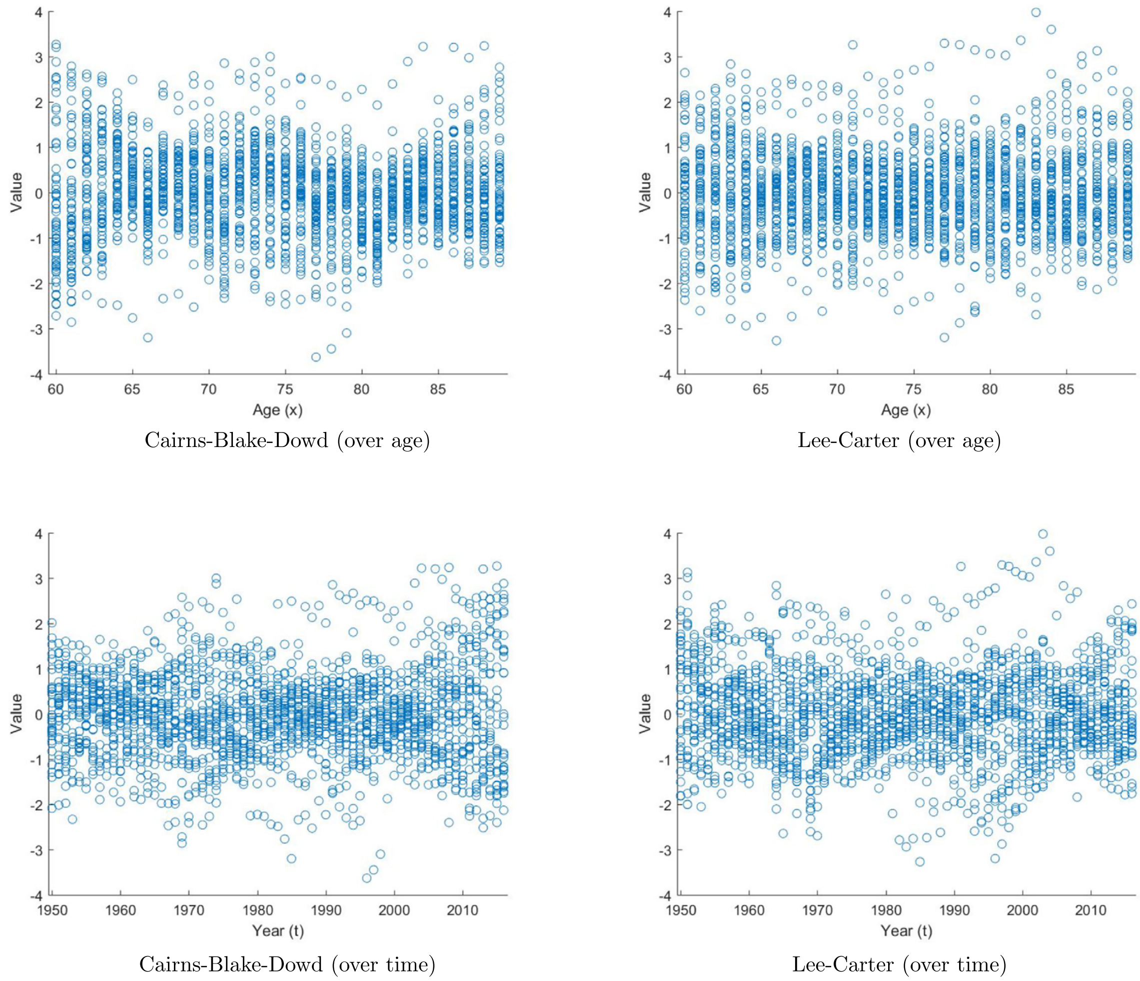

Fig. 4. Deviance residuals produced by the Lee-Carter model and the Cairns-Blake-Dowd model when fitted to the data set under consideration.

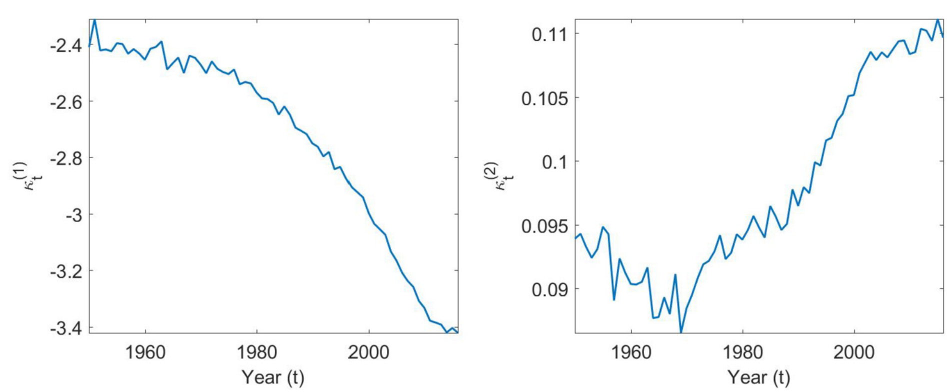

Fig. 5. Estimates of the time-varying indexes $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, for $t = 1950, \ldots, 2016$, in the Cairns-Blake Dowd model.

This in turn means that a bivariate random walk with drift is used to model $\{(\kappa_t^{(1)}, \kappa_t^{(2)})'\}$. The estimate of the drift vector $(c^{(1)}, c^{(2)})'$ is $(-0.0153, 2.3866 \times 10^{-4})'$.

We are aware that the series of the first difference of the time-varying indexes exhibits some lag-1 autocorrelation. We choose not to include an autoregressive term, in part because the lag-1 autocorrelation is only marginally significant, and in part because in the literature the time-varying indexes in stochastic mortality models are typically modeled by a random walk with drift. We remark that the random walk with drift assumption is used in the original versions of the Lee-Carter and Cairns-Blake-Dowd models (Lee and Carter, 1992; Cairns et al., 2006), and is being regarded as 'the universal pattern of mortality decline' by a group of renowned demographers (Tuljapurkar et al., 2000).

6.1.3. Distributions for the innovation vectors

We first examine if the normality assumption for the innovation vectors is adequate. When the Jarque-Bera test is applied to each of the two series of estimated innovations over $t = 1950, \ldots, 2016$, the resulting p-values are 0.0127 and 0.0020, respectively. These p-values

Table 1
Simplified SCCMs for $\{(\kappa_t^{(1)}, \kappa_t^{(2)})'\}$ and $\{(\Delta\kappa_t^{(1)}, \Delta\kappa_t^{(2)})'\}$.

Lag	1	2	3	4	5
SCCM for $\{(\kappa_t^{(1)}, \kappa_t^{(2)})'\}$	$\begin{pmatrix} + & - \\ - & + \end{pmatrix}$	$\begin{pmatrix} + & - \\ - & + \end{pmatrix}$	$\begin{pmatrix} + & - \\ - & + \end{pmatrix}$	$\begin{pmatrix} + & - \\ - & + \end{pmatrix}$	$\begin{pmatrix} + & - \\ - & + \end{pmatrix}$
Lag	1	2	3	4	5
SCCM for $\{(\Delta\kappa_t^{(1)}, \Delta\kappa_t^{(2)})'\}$	$\begin{pmatrix} - & - \\ - & - \end{pmatrix}$	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$

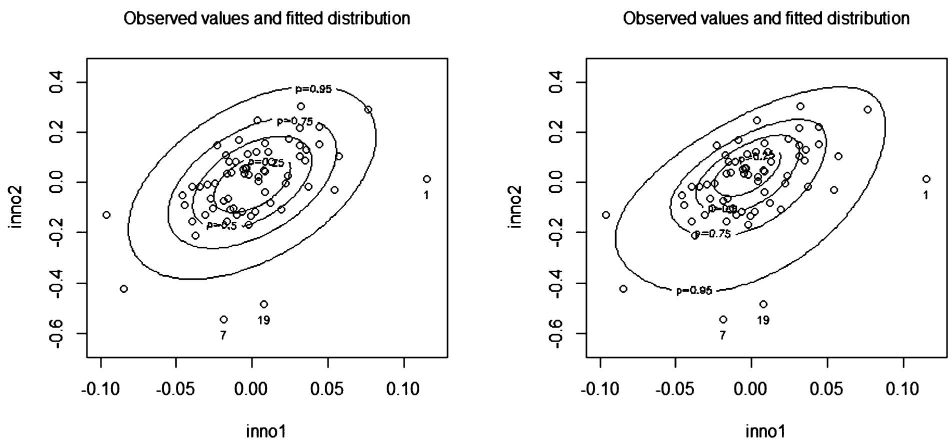

Fig. 6. Contour plots of the estimated bivariate normal and skew-t distributions.

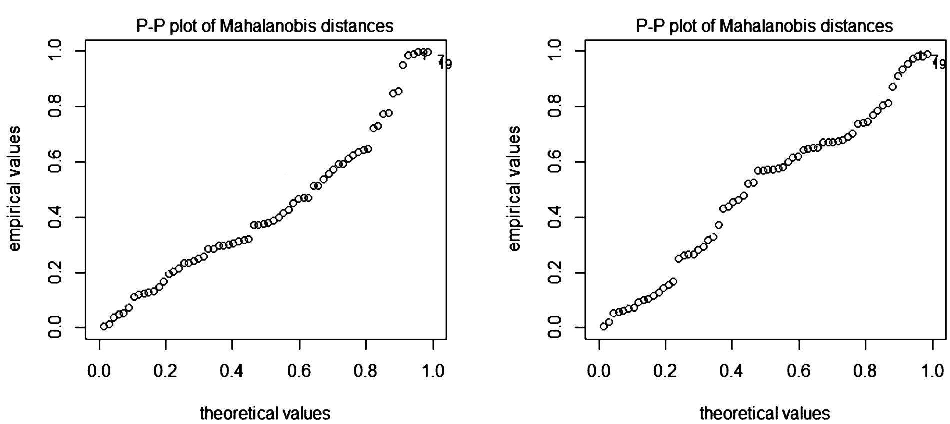

Fig. 7. p-p plots of the Mahalanobis distances generated from the estimated bivariate normal and skew-t distributions.

indicate that the marginal distributions of both $\eta_1(t)$ and $\eta_2(t)$ are not normal at the 5% level of significance. We also apply Henze-Zirkler's multivariate normality test to the series of the innovation vector $(\eta_1(t), \eta_2(t))'$ over $t = 1950, \ldots, 2016$. The resulting p-value is 0.0208, suggesting a rejection of the null hypothesis that the innovation vector is bivariate-normally distributed at the 5% level of significance. In light of the results of the normality tests, we opt to model the innovation vectors with the bivariate skew-t distribution.

The fitness of the estimated bivariate skew-t distribution is then analyzed by graphical means. The following observations are made:

- As shown in Fig. 6, the contour plot of the density function of the bivariate skew-t distribution is substantially more in line with the observed values, compared to that of the bivariate normal distribution.
- As shown in Fig. 7, the p-p plot of the Mahalanobis distances (Healy, 1968; Azzalini and Capitanio, 2003) generated from the estimated bivariate skew-t distribution is close to the 45 degree line, and suggests a substantial improvement in goodness-of-fit compared to the bivariate normal distribution.

Finally, a bivariate Kolmogorov-Smirnov goodness-of-fit test (Justel et al., 1997) is used to test the null hypothesis that the bivariate skewed-t distribution provides an adequate fit. From the data, it is found that the value of the test statistic is 0.08464. Using Monte-Carlo simulations, we generate the distribution of the test statistic under the null hypothesis (for a sample size of 66), and obtain the critical values reported in Table 2. It is clear that the null hypothesis cannot be rejected at any reasonable level of significance. For the reader's information, the estimates of the parameters in the bivariate skew-t distribution are provided in Table 3.

Table 2
Critical values of the bivariate Kolmogorov-Smirnov goodness-of-fit test when the sample size is 66.

Level of significance	0.1	0.05	0.01	0.001
Critical value	0.1563	0.1721	0.2040	0.2453

Table 3
Estimates of parameters in the skew-t distribution assumed for the innovation vector.

Parameter	$\vec{\xi}$	Ω	$\vec{\zeta}$	ν
Estimate	$\begin{pmatrix} -0.0061 \\ 0.0011 \end{pmatrix}$	$\begin{pmatrix} 7.2847\times 10^{-4} & 0.1306\times 10^{-4} \\ 0.1306\times 10^{-4} & 0.0234\times 10^{-4} \end{pmatrix}$	$\begin{pmatrix} 1.1803 \\ -2.1223 \end{pmatrix}$	5.3491

Table 4
Specifications of the four q-forwards used in the hedge portfolios.

j	x_j	T_j	v_{x_j}
1	60	10	2.25%
2	65	10	2.92%
3	70	10	3.66%
4	75	10	4.28%

6.2. General assumptions

We assume that the liability being hedged is a portfolio of life annuities with a term of $T_L = 25$ years. The annuities are issued to individuals from the male population of England and Wales, all aged $x_L = 60$ when the annuities are sold at the end of year $t_0 = 2016$. We consider a static hedge, which is established at the end of year $t_0 = 2016$.

Each annuity makes a payment of \$1 to the annuitant at the end of each year until the annuitant dies or the end of the term is reached, whichever is the earliest. As previously mentioned, we assume that the annuity portfolio is large enough so that micro longevity risk is negligible.

The hedging instruments used are $m = 4$ q-forwards, all of which are freshly launched at time $t_0 = 2016$ and linked to the mortality of English and Welsh males. Table 4 shows the reference age x_j and time-to-maturity T_j of each of the four q-forwards; it also gives, for each $j = 1, 2, 3, 4$, the value of $v(x_j)$, the estimated volatility of the yearly changes in the smoothed death probability at age x_j.[8] Given $v(x_j)$ and the assumption that the market price of risk ℓ is 0.25, we can calculate the forward mortality rate (i.e., the fixed leg) of the jth q-forward using equation (5).[9]

We set C to 0.5% in the implementation of the polynomial goal programming method. This assumption is equivalent to saying that the hedger has a budget of 0.5% of the value of the liability (i.e., a budget of $0.005 \times \mathrm{E}[V_L(t+1)|\mathcal{F}_t]$) for the hedge formulated at time t. Finally, an interest rate of $r = 0.01$ is used to discount all cash flows.

6.3. Implementing the approximation

To implement the constrained optimization of the hedge portfolio, we approximate $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$ using the analytical method presented in Section 5.3. We now demonstrate the calculations involved.

First, given the estimates of the parameters in the bivariate skew-t distribution for the innovation vectors (Table 3), we calculate the moments of the innovation vector $\vec{\eta}(t+1) = (\eta_1(t+1), \eta_2(t+1))'$, given information up to and including time t, using the expressions derived in Appendix A. The calculated second, third, and fourth moments of $\vec{\eta}(t+1)$ given $\mathcal{F}_t$ are reported in Table 5. Note that the parameter estimates in Table 3 imply that the first moment of $\vec{\eta}(t+1)$ given $\mathcal{F}_t$ is $\mathrm{E}(\vec{\eta}(t+1)|\mathcal{F}_t) = \vec{0}$.

Next, we express the Cairns-Blake-Dowd model in a state-space form. For the observation equation, we have

$$\vec{y}(t) = \begin{pmatrix} \ln \frac{q(x_a,t)}{1-q(x_a,t)} \\ \ln \frac{q(x_a+1,t)}{1-q(x_a+1,t)} \\ \vdots \\ \ln \frac{q(x_b,t)}{1-q(x_b,t)} \end{pmatrix}, \quad \vec{d} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & (x_a - \bar{x}) \\ 1 & (x_a + 1 - \bar{x}) \\ \vdots & \vdots \\ 1 & (x_b - \bar{x}) \end{pmatrix},$$

[8] The values of $v(x_j)$ are calculated from smoothed death probabilities because the pricing of q-forwards (which takes as $v(x_j)$ an input) should not include small sample risk that can be diversified away.

[9] In reality, the market price of risk ℓ might be time-varying as market participants' views on longevity risk might change over time. We assume in our numerical illustrations that the market price of risk does not vary with time, in part for simplicity and in part for the fact that the lack of transactions in the currently infantile longevity risk market prohibits us from setting up a time-varying model for ℓ.

Table 5
The calculated second, third, and fourth moments of $\vec{\eta}(t+1) = (\eta_1(t+1), \eta_2(t+1))'$, given information up to and including time t.

	$\eta_1(t+1)$	$\eta_2(t+1)$
$\eta_1(t+1)$	1.1323×10^{-3}	2.6931×10^{-5}
$\eta_2(t+1)$	2.6931×10^{-5}	2.5557×10^{-6}

Second moment

	$\eta_1(t+1)$		$\eta_2(t+1)$	
	$\eta_1(t+1)$	$\eta_2(t+1)$	$\eta_1(t+1)$	$\eta_2(t+1)$
$\eta_1(t+1)$	8.1938×10^{-6}	-4.1933×10^{-7}	-4.1933×10^{-7}	-1.4355×10^{-8}
$\eta_2(t+1)$	-4.1933×10^{-7}	-1.4355×10^{-8}	-1.4355×10^{-8}	-4.4271×10^{-9}

Third moment

	$\eta_1(t+1)$			
	$\eta_1(t+1)$		$\eta_2(t+1)$	
	$\eta_1(t+1)$	$\eta_2(t+1)$	$\eta_1(t+1)$	$\eta_2(t+1)$
$\eta_1(t+1)$	9.6862×10^{-6}	2.1497×10^{-7}	2.1497×10^{-7}	1.1405×10^{-8}
$\eta_2(t+1)$	2.1497×10^{-7}	1.1405×10^{-8}	1.1405×10^{-8}	5.1680×10^{-10}
	$\eta_2(t+1)$			
	$\eta_1(t+1)$		$\eta_2(t+1)$	
	$\eta_1(t+1)$	$\eta_2(t+1)$	$\eta_1(t+1)$	$\eta_2(t+1)$
$\eta_1(t+1)$	2.1497×10^{-7}	1.1405×10^{-8}	1.1405×10^{-8}	5.1680×10^{-10}
$\eta_2(t+1)$	1.1405×10^{-8}	5.1680×10^{-10}	5.1680×10^{-10}	6.5290×10^{-11}

Fourth moment

Table 6
Values of partial derivatives $D_k(j, t_0+1)$, for $k = 1, 2$ and $j = 0, 1, \ldots, 4$.

	$D_k(0, t_0+1)$	$D_k(1, t_0+1)$	$D_k(2, t_0+1)$	$D_k(3, t_0+1)$	$D_k(4, t_0+1)$
$k=1$	-3.0993	-0.0049	-0.0086	-0.0148	-0.0253
$k=2$	11.3312	0.0716	0.0814	0.0666	-0.0126

$$\vec{\alpha}(t) = \begin{pmatrix} \kappa_t^{(1)} \\ \kappa_t^{(2)} \end{pmatrix} \quad \text{and} \quad \vec{\epsilon}(t) = \begin{pmatrix} \epsilon(x_a, t) \\ \epsilon(x_a+1, t) \\ \vdots \\ \epsilon(x_b, t) \end{pmatrix}.$$

For the transition equation, we have

$$\vec{\alpha}(t) = \begin{pmatrix} \kappa_t^{(1)} \\ \kappa_t^{(2)} \end{pmatrix}, \quad \vec{c} = \begin{pmatrix} c^{(1)} \\ c^{(2)} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad \vec{\eta}(t) = \begin{pmatrix} \eta_1(t) \\ \eta_2(t) \end{pmatrix}.$$

Using the state-space parameterization and the formulas provided in Appendix B, we calculate all of the required partial derivatives, namely, $D_k(j, t+1)$ for $j = 0, 1, \ldots, 4$ and $k = 1, 2$. Table 6 shows the calculated values of $D_k(j, t_0+1)$, for $j = 0, 1, \ldots, 4$ and $k = 1, 2$, for the reader's reference.

Finally, with the calculated moments of $\vec{\eta}(t+1)$ given $\mathcal{F}_t$ and the partial derivatives, we use the analytical expressions provided in Section 5.3 to compute approximated values of $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$, which are subsequently fed into the constrained optimization problem formulated in Section 4 to obtain the optimized hedge portfolio.

6.4. Evaluating the approximation

As discussed in Section 5.3, the approximation formulas for $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$ are derived using $V_l(t+1)$ as an approximate for $V_L(t+1)$ and $V_h(t+1; j, t)$ as an approximate for $V_H(t+1; j, t)$, for $j = 1, \ldots, 4$. We now evaluate the accuracy of these approximations.

To evaluate the accuracy of these approximations, we generate 20,000 realizations of $\mathcal{F}_{t_0+1}$ given $\mathcal{F}_{t_0}$. For each of these realizations, we calculate a realization of $V_l(t_0+1) = \mathrm{E}(l|\mathcal{F}_{t_0+1})$, and a realization of $V_h(t_0+1; j, t_0) = \mathrm{E}(h(j, t_0)|\mathcal{F}_{t_0+1})$ for each $j = 1, \ldots, 4$. Then, for each of the 20,000 realizations of $\mathcal{F}_{t_0+1}$, we compute the arctangent absolute percentage errors in approximating $V_L(t_0+1)$ and $V_H(t_0+1; j, t_0)$, $j = 1, \ldots, 4$, as

$$\arctan\left|\frac{V_l(t_0+1) - V_L(t_0+1)}{V_L(t_0+1)}\right| \quad \text{and} \quad \arctan\left|\frac{V_h(t_0+1, j, t_0) - V_H(t_0+1, j, t_0)}{V_H(t_0+1, j, t_0)}\right|,$$

respectively. In calculating the arctangent absolute percentage errors, the values of $V_H(t_0+1, j, t_0)$ and $V_L(t_0+1)$ are obtained using full (nested) simulations. Finally, the arctangent absolute percentage errors are then averaged over the 20,000 realizations to obtain the mean

Table 7
The mean arctangent absolute percentage errors (MAAPE) in approximating $V_L(t_0+1)$ and $V_H(t_0+1; j, t_0)$, for $j = 1, \ldots, 4$.

	$V_L(t_0+1)$	$V_H(t_0+1; 1, t_0)$	$V_H(t_0+1; 2, t_0)$	$V_H(t_0+1; 3, t_0)$	$V_H(t_0+1; 4, t_0)$
MAAPE	0.0060%	2.3817%	1.0808%	0.7651%	1.1427%

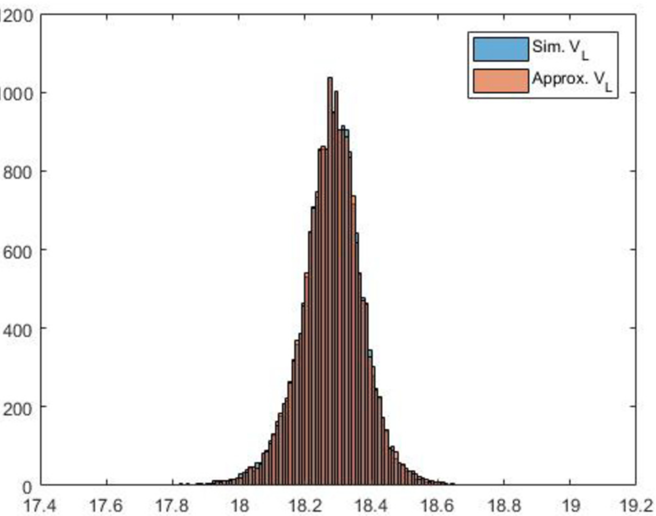

Fig. 8. Empirical distributions of $V_L(t_0+1)$ (obtained using full nested simulations) and $V_l(t_0+1)$ (obtained using the proposed approximation). (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

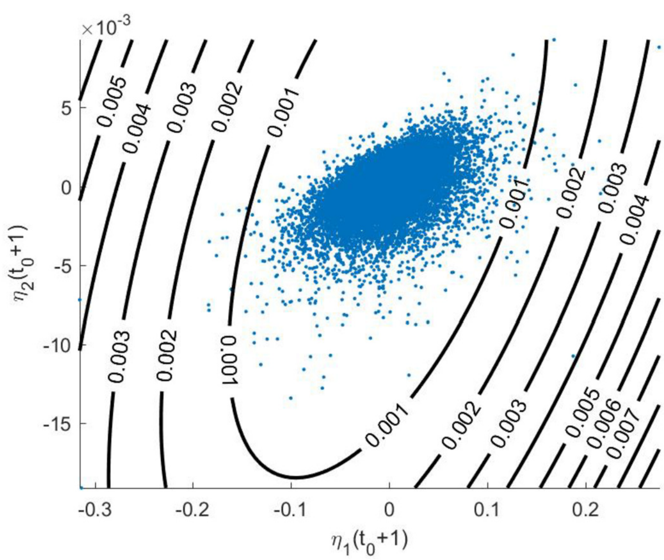

Fig. 9. Percentage errors in the approximation of $V_L(t_0+1)$ for different values of $\eta_1(t_0+1)$ and $\eta_2(t_0+1)$.

arctangent absolute percentage errors (MAAPE) in approximating $V_L(t_0+1)$ and $V_H(t_0+1; j, t_0)$ for $j = 1, \ldots, 4$.[10] As shown in Table 7, the MAAPEs are reasonably low.

In addition to reporting the MAAPEs, we consider two aspects that may help us more holistically assess the approximation accuracy. First, we compare the empirical distributions of $V_L(t_0+1)$ and $V_H(t_0+1; j, t_0)$ for $j = 1, \ldots, 4$ (obtained using full nested simulations) and the empirical distributions of $V_l(t_0+1)$ and $V_h(t_0+1; j, t_0)$ for $j = 1, \ldots, 4$ (obtained using the proposed approximation). Fig. 8 shows the empirical distributions of $V_L(t_0+1)$ and $V_l(t_0+1)$ side by side. The close proximity between the two empirical distributions substantiates the accuracy of the approximation. The comparisons between $V_h(t_0+1; j, t_0)$ and $V_H(t_0+1; j, t_0)$ for $j = 1, \ldots, 4$ are similar.

Second, we simulate 20,000 pairs of $\eta_1(t_0+1)$ and $\eta_2(t_0+1)$ (the innovations at time t_0+1), and evaluate the accuracy of the approximations of $V_L(t_0+1)$ and $V_H(t_0+1; j, t_0)$, $j = 1, \ldots, 4$, for each of the 20,000 pairs of simulated $\eta_1(t_0+1)$ and $\eta_2(t_0+1)$. As an example, let us consider the result for $V_L(t_0+1)$ (Fig. 9). The dots in the diagrams represent the 20,000 simulated pairs of $\eta_1(t_0+1)$ and

[10] Although mean absolute percentage error (MAPE) is a more commonly used metric, we measure approximation errors with MAAPE, because the latter is not reliable when the denominator (true value) is small (see Kim and Kim, 2016).

Table 8
Hedging results for static hedges implemented with nine different sets of risk preferences.

	Unhedged	Risk Preference (i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	Aspired Level
λ_1	–	1	1	1	1	1	1	1	1	1	–
λ_2	–	1	1	1	1	1	1	1	1	1	–
λ_3	–	0	0.1	0.5	1	5	0	0	0	0	–
λ_4	–	0	0	0	0	0	0.1	0.5	1	5	–
N_1	–	3.3700	2.6720	0.5145	0.0500	0.0208	3.2191	3.1593	2.9307	2.8435	–
N_2	–	0.0006	0.0000	0.0000	0.0008	0.0229	0.0031	0.0042	0.0026	0.0000	–
N_3	–	0.0013	0.0000	0.0000	0.0000	0.0018	0.0073	0.0097	0.0060	0.0033	–
N_4	–	0.0328	0.1003	0.3060	0.3500	0.3462	0.0431	0.0473	0.0715	0.0821	–
Mean	18.2798	18.3723	18.3723	18.3723	18.3723	18.3723	18.3723	18.3723	18.3723	18.3723	18.2798
Variance ($\times 10^{-3}$)	9.4427	3.4130	3.4290	4.1793	4.4774	4.4870	3.4139	3.4143	3.4162	3.4171	3.4106
Skewness	−0.4464	0.0745	−0.0444	−0.4195	−0.4996	−0.4998	0.0499	0.0403	0.0007	−0.0149	−0.5082
Kurtosis	6.2071	5.9309	5.8741	6.1476	6.2946	6.2949	5.9143	5.9084	5.8886	5.8826	5.8580

$\eta_2(t_0+1)$ given $\mathcal{F}_{t_0}$. The cloud of dots may therefore be seen as the possible range of $(\eta_1(t_0+1), \eta_2(t_0+1))'$. The contour lines represent the percentage errors in approximating $V_L(t_0+1)$. At the centroid of the cloud of dots, the percentage error is zero as the approximation is exact at $(\hat{\eta}_1(t_0+1), \hat{\eta}_2(t_0+1))'$ by definition. As the distance from the centroid increases, the percentage errors also increases. However, within the boundary of the cloud of dots, the percentage errors are no greater than 0.4%, suggesting that the accuracy of the quadratic approximation is very high over the possible range of $(\eta_1(t_0+1), \eta_2(t_0+1))'$.

6.5. The effect of risk preferences

Table 8 presents the results of the static hedges implemented using nine different sets of risk preferences. For the sake of completeness, we also show in Table 8 the aspired levels of $M^{(i)}(\vec{N}_{t_0})$ for $i=1,2,3,4$, and the values of $M^{(i)}(\vec{N}_{t_0})$ for $i=1,2,3,4$ when the annuity liability is left unhedged. All values of $M^{(i)}(\vec{N}_{t_0})$ shown in Table 8 are calculated with the analytical approximation method discussed in Section 5.3.

The following general observations are made. First, the unhedged position leads to the most desirable (smallest) value of $M^{(1)}(\vec{N}_{t_0})$ compared to all of the nine hedged positions. This outcome is due to the fact that acquiring any hedging instrument which comes with a cost would lead to a larger value of $M^{(1)}(\vec{N}_{t_0})$. Second, the aspired levels of $M^{(1)}(\vec{N}_{t_0}), \ldots, M^{(4)}(\vec{N}_{t_0})$ are the most desirable compared to the values of $M^{(1)}(\vec{N}_{t_0}), \ldots, M^{(4)}(\vec{N}_{t_0})$ resulting from the unhedged position and all of the nine hedged positions. This outcome is expected, as the aspired level of a moment, by definition, is the best achievable level of the moment when the objectives concerning the other moments are not taken into consideration.

Next, we discuss the observations that are made in four specific comparisons.

- Risk Preferences (i) vs. (ii)-(v)
  Risk Preference (i) is a simple mean-variance optimization. Compared to Risk Preference (i), Risk Preferences (ii)-(v), all of which incorporate skewness, yield more desirable (more negative) values of skewness, at the expense of less desirable values of mean and variance.
- Risk Preferences (i) vs. (vi)-(ix)
  Compared to Risk Preference (i), Risk Preferences (vi)-(ix), all of which incorporate kurtosis, yield more desirable (lower) values of kurtosis, at the expense of less desirable values of mean and variance.
- Risk Preferences (ii)-(v)
  The difference among Risk Preferences (ii), (iii), (iv) and (v) lies in the attitude towards skewness. Risk Preference (v) places the strongest emphasis on skewness, and not surprisingly, it leads to the most desirable (most negative) value of skewness. Risk Preference (ii), in contrast, places the weakest emphasis on skewness, and therefore results in the least desirable value of skewness.
- Risk Preferences (vi) to (ix)
  The difference among Risk Preferences (vi), (vii), (viii) and (ix) lies in the attitude towards kurtosis. Risk Preference (ix) places the strongest emphasis on kurtosis, and hence it leads to the most desirable (lowest) value of kurtosis. Risk Preference (vi), in contrast, places the weakest emphasis on kurtosis, and thus results in the least desirable value of kurtosis.

6.6. Analysis of gains and losses

In this subsection, we perform a gain-and-loss analysis for the static longevity hedge. First, let us consider the hedger's position at time t_0 when the hedge is established. At time t_0, the value of the hedger's position is given by

$$V_L(t_0) + P_H - \sum_{i=1}^{4} N_j(t_0) V_H(t_0; j, t_0) - P_a,$$

where $V_L(t_0)$ is the value of the liability at time t_0, $N_j(t_0)$ is the notional amount of the jth q-forward at time t_0, $V_H(t_0; j, t_0)$ is the time-t_0 value of jth q-forward launched at time t_0, P_H is the price of the hedge and P_a is the annuity premium received. We set

$$P_H = \sum_{i=1}^{4} N_j(t_0) V_H(t_0; j, t_0),$$

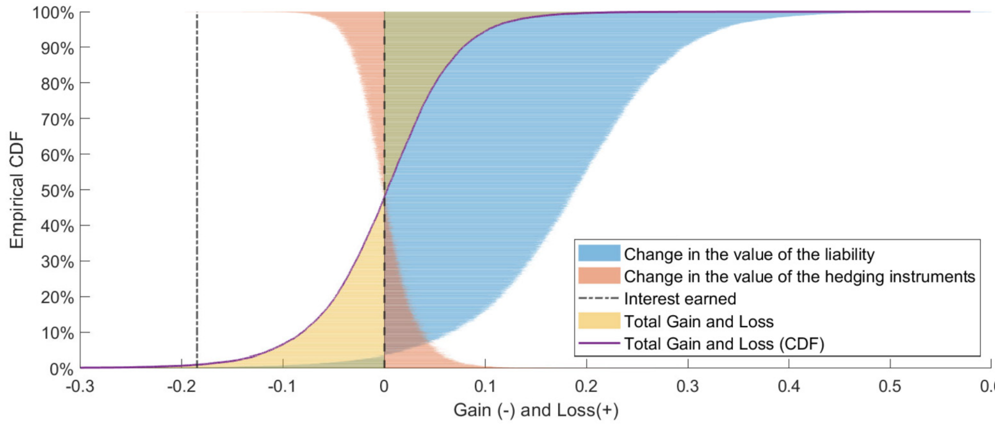

Fig. 10. Empirical cumulative distributions of the aggregate gain-and-loss and its components when $\lambda_1 = \lambda_2 = 1$ and $\lambda_3 = \lambda_4 = 0$.

which represents the total time-t_0 cost of the q-forwards, and set $P_a = V_L(t_0)$ so that P_a represents the actuarially fair annuity premium that makes the value of the hedger's position at time t_0 zero.

Then, at time $t_0 + 1$, the hedger realizes a gain or loss that arises from the following three sources:

1. Change in the value of the liability: $e^r V_L(t_0+1) - V_L(t_0)$
2. Changes in the values of q-forwards: $\sum_{j=1}^{4} N_j(t_0)(e^r V_H(t_0+1; j, t_0) - V_H(t_0; j, t_0))$
3. Interest earned on the annuity premium collected at time t_0 net of the time-t_0 cost of the hedge: $(e^r - 1)(P_a - P_H)$

Note that according to their definitions, both V_L and V_H represent values measured in time-t_0 dollars. As such, in the above, a one-period accumulation factor is multiplied to $V_L(t_0+1)$ and $V_H(t_0+1; j, t_0)$ so that the gains and losses from all sources are measured in time-(t_0+1) dollars consistently.

Aggregating the gains and losses from the three sources, we obtain the aggregate gain-and-loss from time t_0 to $t_0 + 1$:

$$e^r V_L(t_0+1) - V_L(t_0) - \sum_{j=1}^{4} N_j(t_0)(e^r V_H(t_0+1; j, t_0) - V_H(t_0; j, t_0)) - (e^r - 1)(P_a - P_H),$$

which can be simplified to

$$e^r \left(V_L(t_0+1) - V_L(t_0) - \sum_{j=1}^{4} N_j(t_0)\,(V_H(t_0+1; j, t_0) - V_H(t_0; j, t_0)) \right),$$

by considering the expression for P_H and P_a.

Based on the stochastic mortality model described in Section 6.1, the distribution of the aggregate gain-and-loss from time t_0 to $t_0 + 1$ is studied empirically. As an example, Fig. 10 shows the empirical cumulative distributions of the aggregate gain-and-loss and its components when we set λ_1 and λ_2 to one and λ_3 and λ_4 to zero.

To examine the effect of incorporating skewness and kurtosis into the multi-objective function on the distribution of aggregate gain-and-loss, we compare the standard deviation, Value-at-Risk and tail-Value-at-Risk, between a mean-variance longevity hedge (for which $\lambda_3 = \lambda_4 = 0$) and a mean-variance-skewness-kurtosis longevity hedge (for which $\lambda_3 > 0$ and/or $\lambda_4 > 0$). The results are tabulated in Table 9,[11] from which the following findings are drawn.

First, compared to a mean-variance longevity hedge (for which $\lambda_1 = \lambda_2 = 1$ and $\lambda_3 = \lambda_4 = 0$), a longevity hedge that incorporates skewness through a moderate risk preference parameter (say $\lambda_3 = 0.5$) is more effective in terms of the Value-at-Risk and tail-Value-at-Risk at a 99.5% confidence level. However, if the confidence level drops to, say 80%, then incorporating skewness into the multi-objective function brings no benefit. This outcome is consistent with the arguments provided in Section 3.3, which explain why incorporating skewness may increase (rather than decrease) the Value-at-Risk if the confidence level is low. In addition, if the risk preference parameter for skewness becomes excessively large (say $\lambda_3 = 5$), then the performance of the mean-variance-skewness longevity hedge may deteriorate and may not even be as good as that of the mean-variance longevity hedge in terms of 99.5% Value-at-Risk, because an excessively large value of λ_3 may lead the hedge to over-emphasize on skewness.

For fixed values of λ_1, λ_2 and λ_3 ($\lambda_1 = \lambda_2 = 1$; $\lambda_3 = 0.1$), incorporating kurtosis into the multi-objective function through a small to moderate risk preference parameter for kurtosis (say, $0.1 \le \lambda_4 \le 1$) results in a marginally better performance in terms of 99.5% Value-at-Risk. However, when the value of λ_4 is excessively large (say $\lambda_4 = 5$), the mean-variance-skewness-kurtosis hedge is no longer outperforming.

6.7. Comparison with alternative hedging strategies

In this section, we compare our proposed hedge with the following single-objective hedges:

[11] The gain-and-loss distribution for the reference hedge has a mean of 0.0005 and a standard deviation of 0.0593. At a 80% confidence level, the Value-at-Risk and tail-Value-at-Risk are 0.0437 and 0.0816, respectively; at a 99.5% confidence level, the Value-at-Risk and tail-Value-at-Risk are 0.1821 and 0.2302, respectively.

Table 9
Differences ($\times 10^4$) in mean, standard deviation, Value-at-Risk (80% and 99.5% confidence levels), and tail-Value-at-Risk (80% and 99.5% confidence levels) between various hedges and the reference mean-variance hedge (with $\lambda_1 = \lambda_2 = 1$ and $\lambda_3 = \lambda_4 = 0$). A negative difference indicates that a better performance compared to the reference mean-variance hedge.

Risk Preference	Mean	Standard Deviation	Value-at-Risk 80%	Value-at-Risk 99.5%	Tail-Value-at-Risk 80%	Tail-Value-at-Risk 99.5%
Mean-variance-skewness hedges						
1-1-0.1-0	−0.07	3.75	4.96	−33.01	−6.38	−79.17
1-1-0.5-0	−0.32	72.09	63.58	−2.45	52.48	−36.84
1-1-1-0	−0.35	85.63	75.78	18.39	65.61	−18.18
1-1-5-0	−0.39	87.99	77.47	24.56	68.55	−10.88
Mean-variance-skewness-kurtosis hedges						
1-1-0.1-0.1	−0.07	3.62	4.80	−33.81	−6.41	−78.65
1-1-0.1-0.5	−0.07	3.22	3.72	−35.03	−6.49	−76.98
1-1-0.1-1	−0.07	2.89	3.09	−35.46	−6.53	−75.46
1-1-0.1-5	−0.08	2.03	3.27	−31.57	−6.61	−71.31
Minimization of 99.5% Value-at-Risk						
N/A	−239.20	79.73	−179.37	−49.91	−141.99	−31.03
Minimization of semi-variance						
N/A	−0.63	2.65	2.64	−35.37	−6.89	−73.33
Minimization of 99.5% Tail-Value-at-Risk						
N/A	−0.63	5.95	6.06	−30.52	−5.94	−85.81

- A hedge that minimizes the portfolio's 99.5% Value-at-Risk:

$$\min_{N_{1,t_0},\dots,N_{4,t_0}} \mathrm{VaR}_\gamma \left(V_L(t_0+1) - \sum_{j=1}^{4} (N_{j,t_0} \times V_H(t_0+1, j, t_0)) \middle| \mathcal{F}_{t_0} \right), \tag{12}$$

  where $\gamma = 0.995$.
- A hedge that minimizes the portfolio's semi-variance:

$$\min_{N_{1,t_0},\dots,N_{4,t_0}} \mathrm{E}\left(\max\left(0, V_P(t_0+1; \vec{N}_{t_0}) - \mathrm{E}\left(V_P(t_0+1; \vec{N}_{t_0})\right)\right)^2 \middle| \mathcal{F}_{t_0} \right) \tag{13}$$

- A hedge that minimizes the portfolio's 99.5% tail-Value-at-Risk:

$$\min_{N_{1,t_0},\dots,N_{4,t_0}} \mathrm{E}\left(V_P(t_0+1; \vec{N}_{t_0}) \middle| V_P(t_0+1; \vec{N}_{t_0}) > \pi_\gamma, \mathcal{F}_{t_0} \right) \tag{14}$$

  where $\gamma = 0.995$,

$$V_P(t_0+1; \vec{N}_{t_0}) = V_L(t_0+1) - \sum_{j=1}^{4} (N_{j,t_0} \times V_H(t_0+1, j, t_0))$$

  and

$$\pi_\gamma = \mathrm{VaR}_\gamma \left(V_P(t_0+1; \vec{N}_{t_0}) \middle| \mathcal{F}_{t_0} \right).$$

As with the optimization for our proposed hedging strategy, all of the optimizations above are subject to the following two constraints: (1) $N_{j,t_0} \geq 0$ (no short-selling); (2) $\sum_{j=1}^{m} N_{j,t_0} \hat{H}_j \leq 0.5\% \hat{L}$ (the cost of hedging is less than 0.5% of the expected liability value).

By approximating $V_L(t_0+1)$ and $V_H(t_0+1, j, t_0)$ in equation (12) with $V_l(t_0+1)$ and $V_h(t_0+1, j, t_0)$, respectively, Liu and Li (2021) provide an analytical solution to the Value-at-Risk minimization problem. A brief summary of the analytical solution is provided in Appendix C.

We adopt the analytical solution provided by Liu and Li (2021) to obtain results for the hedge that minimizes 99.5% Value-at-Risk. For the other two single-objective hedges, the optimizations are performed using full nested simulations with 20,000 outer scenarios.[12] The results for the three single-objective hedges are compared against the mean-variance-skewness-kurtosis hedges in Table 9. The following observations can be made:

- Minimization of the portfolio's 99.5% Value-at-Risk leads to an impressively lower 99.5% Value-at-Risk; however, the Value-at-Risk-minimizing hedge underperforms the mean-variance-skewness-kurtosis hedges (with the range of risk preferences under consideration) in terms of both standard deviation (variance) and 99.5% tail-Value-at-Risk. Notably, the Value-at-Risk-minimizing hedge results in a standard deviation that is significantly higher than that of the unhedged position.

[12] As a sanity check, we also performed the Value-at-Risk minimization numerically using full nested simulations. It is found that the analytical and numerical results are the same up to four significant figures.

- Similarly, as expected, minimization of the portfolio's 99.5% tail-Value-at-Risk leads to a lower 99.5% tail-Value-at-Risk, but a higher 99.5% VaR and standard deviation compared to the mean-variance-skewness-kurtosis hedges (with the range of risk preferences under consideration).
- The semi-variance hedge beats the mean-variance-skewness-kurtosis hedge with risk preferences 1, 1, 0.1 and 1 in terms of standard deviation, but underperforms in terms of Value-at-Risk and tail-Value-at-Risk at both 80% and 99.5% confidence levels.

To conclude, if the hedger has a very specific hedging objective (e.g., to reduce the portfolio's Value-at-Risk as much as possible), then a hedge that optimizes a single risk measure might be more suited, but the drawback is that other risk measures for the portfolio may be compromised. In contrast, if the hedger does not have a specific hedging objective, then the proposed mean-variance-skewness-kurtosis hedge is an attractive alternative, as it provides a better all-round performance compared to single-objective hedges.

7. Concluding remarks

In this paper, we develop a mean-variance-skewness-kurtosis approach to optimizing longevity hedges. Compared to mean-variance methods in the literature, our proposed approach is more appropriate when the evolution of mortality is driven by non-normal distributions and the hedger concerns with not only the volatility but also the skewness and/or excess kurtosis of his/her portfolio.

To maximize applicability of our proposed approach, we base our derivations on a general state-space representation that encompasses most of the discrete-time stochastic mortality models used in practice. We derive approximate analytical expressions for the moments of the values of the hedging instruments and the liability being hedged. These expressions are integrated with a polynomial programming goal model, from which a solution to the optimal hedge portfolio is identified.

While a static hedge is implemented in this paper, the proposed hedging strategy may be extended to a dynamic setting with a dynamic programming approach. For instance, if we were to develop a dynamic hedge that aims to mitigate the uncertainty associated with V_P in T^* years, for some integer $T^* > 1$, then the optimal value of $\vec{N}_t$ (where $t_0 \le t < t_0 + T^*$) would depend on the values of $\vec{N}_{t+1}, \vec{N}_{t+2}, \dots, \vec{N}_{t_0+T^*-1}$ (and their risk mitigation effects). A dynamic programming approach would allow us to involve the future hedge portfolios in the optimization.

As the focus of this paper is on developing hedging strategies, we have not paid much attention on the pricing problem. Admittedly, the q-forward pricing formula we apply (equation (5)) involves variance (of historical mortality improvements) only, and is therefore inadequate when higher moments also matter to investors acquiring longevity risk exposures. In future research, it would be interesting to investigate how higher moments may be incorporated into q-forward pricing, possibly drawing on the work of Affleck-Graves and McDonald (1989) and Sears (1985).

In addition to non-normality, mortality dynamics may also exhibit conditional heteroscedasticity. Recent studies have found that the GARCH effect is significant in the mortality dynamics of various national populations (Chai et al., 2013; Gao and Hu, 2009; Wang and Li, 2016). Another possible avenue of future research is to generalize our proposed hedging strategy to incorporate GARCH effects. Previous studies concerning the existence of moments of various GARCH processes (e.g. Ling and McAleer, 2002) are relevant to this suggested future work.

Declaration of competing interest

There is no competing interest.

Data availability

The data used in this paper is available from the public domain.

Acknowledgements

This work is supported by research grants from the Natural Sciences and Engineering Research Council of Canada (Discovery Grant RGPIN-2021-02409) and the Society of Actuaries Center of Actuarial Excellence Program.

Appendix A. Moments of the assumed multivariate skew-*t* distribution

In this Appendix, we study the moments of random vectors which follow the multivariate skew-t distribution introduced by Azzalini and Capitanio (2003), the distribution we use to model the innovation vector $\vec{\eta}(t)$ in the stochastic mortality models.

As in the main text, we use

$$\vec{Y} \sim \mathrm{St}_n(\vec{\xi}, \mathbf{\Omega}, \vec{\zeta}, \nu)$$

to represent an n-dimensional random vector which follows the assumed multivariate skew-t distribution. Recall that $\vec{\xi}$, $\mathbf{\Omega}$, $\vec{\zeta}$, and ν represent the location parameter, the dispersion parameter, the skewness parameter, and the degrees of freedom, respectively.

Azzalini and Capitanio (2003) derived the moments of $\vec{Y}$ for $\vec{\xi} = \vec{0}$ only, but the moments of $\vec{Y}$ for $\vec{\xi} \neq \vec{0}$ are required in the context of this research. In what follows, we first review the moments of $\vec{Y}$ when $\vec{\xi} = \vec{0}$, and then extend the results of Azzalini and Capitanio (2003) to obtain the moments of $\vec{Y}$ when $\vec{\xi} \neq \vec{0}$.

A.1. Moments when $\vec{\xi} = \vec{0}$

As pointed out by Azzalini and Capitanio (2003), $\vec{Y}$ can be expressed as

$$\vec{Y} = \vec{\xi} + W^{-\frac{1}{2}} \vec{Z}, \tag{15}$$

where $W = \chi_\nu^2/\nu$ with χ_ν^2 being a random variable which follows a chi-square distribution with ν degrees of freedom, and Z, which is independent of W, is a random vector that follows a skew-normal distribution with a location parameter of $\vec{0}$, a dispersion parameter of $\mathbf{\Omega}$, and a skewness parameter of $\vec{\zeta}$.

Let $\mathrm{E}^{(m)}(\vec{X})$ be the mth moment of a generic random vector $\vec{X}$. When $\vec{\xi} = \vec{0}$ we have

$$\mathrm{E}^{(m)}(\vec{Y}) = \mathrm{E}(W^{-m/2})\mathrm{E}^{(m)}(\vec{Z}),$$

where

$$\mathrm{E}(W^{-m/2}) = \frac{(\nu/2)^{m/2}\Gamma((\nu-m)/2)}{\Gamma(\nu/2)} \tag{16}$$

and Γ denotes the gamma function. Using the results of Azzalini and Capitanio (1999) and Genton et al. (2001), the first four moments of $\vec{Z}$ can be expressed as follows:

$$\mathrm{E}^{(1)}(\vec{Z}) = \mathrm{E}(\vec{Z}) = \sqrt{\frac{2}{\pi}}\vec{\delta}, \tag{17}$$

$$\mathrm{E}^{(2)}(\vec{Z}) = \mathrm{E}(\vec{Z}\vec{Z}') = \mathbf{\Omega}, \tag{18}$$

$$\mathrm{E}^{(3)}(\vec{Z}) = \mathrm{E}(\vec{Z}\vec{Z}'\otimes\vec{Z}') = \sqrt{\frac{2}{\pi}}\left(\vec{\delta}\otimes\mathbf{\Omega} + \mathrm{vec}(\mathbf{\Omega})\vec{\delta}' + (\mathbf{I}_n\otimes\vec{\delta})\mathbf{\Omega} - \vec{\delta}\otimes\vec{\delta}'\otimes\vec{\delta}\right) \tag{19}$$

and

$$\mathrm{E}^{(4)}(\vec{Z}) = \mathrm{E}(\vec{Z}\vec{Z}'\otimes\vec{Z}'\otimes\vec{Z}') = \mathrm{vec}(\mathbf{\Omega})'\otimes\mathbf{\Omega} + \mathbf{\Omega}(\mathbf{I}_n\otimes\mathrm{vec}(\mathbf{\Omega})) + \mathbf{\Omega}\mathbf{I}_n(\mathrm{vec}(\mathbf{\Omega})'\otimes\mathbf{I}_n)(\mathbf{I}_n\otimes\mathbf{U}), \tag{20}$$

where

$$\vec{\delta} = \frac{\mathbf{\Lambda}\bar{\mathbf{\Omega}}\vec{\zeta}}{\left(1+\vec{\zeta}'\bar{\mathbf{\Omega}}\vec{\zeta}\right)^{1/2}},$$

$\bar{\mathbf{\Omega}}$ is the corresponding correlation matrix of $\mathbf{\Omega}$,[13] $\mathbf{I}_n$ is an n-by-n identity matrix, and $\mathbf{U}$ is an n^2-by-n^2 matrix representing the permutation matrix associated with an n-by-n matrix.[14]

A.2. Moments when $\vec{\xi} \neq \vec{0}$

We now extend the results of Azzalini and Capitanio (2003) to obtain the first four moments of $\vec{Y}$ when $\vec{\xi} \neq \vec{0}$.

A.2.1. The first moment

It follows from equation (15) that the first moment of $\vec{Y}$ for $\vec{\xi} \neq \vec{0}$ can be expressed as

$$\mathrm{E}^{(1)}(\vec{Y}) = \vec{\xi} + \mathrm{E}(W^{-1/2})\mathrm{E}(\vec{Z}).$$

Substituting the expressions of $\mathrm{E}(W^{-1/2})$ and $\mathrm{E}(\vec{Z})$ into equations (16) and (17), we obtain

$$\mathrm{E}^{(1)}(\vec{Y}) = \vec{\xi} + \sqrt{\frac{\nu}{\pi}}\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}\vec{\delta}. \tag{21}$$

A.2.2. The second moment

Using equation (15), we can express the second moment of $\vec{Y}$ for $\vec{\xi} \neq \vec{0}$ as follows:

$$\begin{aligned}
&\mathrm{E}^{(2)}(\vec{Y})\\
&= \mathrm{E}\left[(\vec{Y}-\mathrm{E}(\vec{Y}))(\vec{Y}-\mathrm{E}(\vec{Y}))'\right]\\
&= \mathrm{E}\left[\left(W^{-1/2}\vec{Z}-\mathrm{E}(W^{-1/2})\mathrm{E}(\vec{Z})\right)\left(W^{-1/2}\vec{Z}-\mathrm{E}(W^{-1/2})\mathrm{E}(\vec{Z})\right)'\right]\\
&= \mathrm{E}\left[W^{-1}\vec{Z}\vec{Z}' - W^{-1/2}\mathrm{E}(W^{-1/2})\mathrm{E}(\vec{Z})\vec{Z}' - W^{-1/2}\mathrm{E}(W^{-1/2})\vec{Z}\mathrm{E}(\vec{Z})' + (\mathrm{E}(W^{-1/2}))^2\mathrm{E}(\vec{Z})\mathrm{E}(\vec{Z})'\right]\\
&= \mathrm{E}(W^{-1})\mathrm{E}(\vec{Z}\vec{Z}') - (\mathrm{E}(W^{-1/2}))^2\mathrm{E}(\vec{Z})\mathrm{E}(\vec{Z})'.
\end{aligned}$$

Substituting the expressions for $\mathrm{E}(W^{-1/2})$, $\mathrm{E}(W^{-1})$, $\mathrm{E}(\vec{Z})$ and $\mathrm{E}(\vec{Z}\vec{Z}')$ into equations (16), (17) and (18), we obtain

$$\mathrm{E}^{(2)}(\vec{Y}) = \frac{\nu}{\nu-2}\mathbf{\Omega} - \frac{\nu}{\pi}\left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\right)^2\frac{\mathbf{\Lambda}\bar{\mathbf{\Omega}}\vec{\zeta}\vec{\zeta}'\bar{\mathbf{\Omega}}\mathbf{\Lambda}}{1+\vec{\zeta}'\bar{\mathbf{\Omega}}\vec{\zeta}}. \tag{22}$$

[13] The (i,j)th element of $\bar{\mathbf{\Omega}}$ is calculated as $\omega_{i,j}/\sqrt{\omega_{i,i}\omega_{j,j}}$, where $\omega_{i,j}$ denotes the (i,j)th element in $\mathbf{\Omega}$.

[14] The permutation matrix associated with a generic matrix $\mathbf{X}$ is the matrix $\mathbf{U}$ that satisfies $\mathrm{vec}(\mathbf{X}') = \mathbf{U}\mathrm{vec}(\mathbf{X})$.

A.2.3. The third moment

Using equation (15), we can express the third moment of $\vec{Y}$ for $\vec{\xi} \neq \vec{0}$ as follows:

$$\begin{aligned}
&\mathrm{E}^{(3)}\left(\vec{Y}-\mathrm{E}(\vec{Y})\right) \\
&= \mathrm{E}\left[(\vec{Y}-\mathrm{E}(\vec{Y}))(\vec{Y}-\mathrm{E}(\vec{Y}))' \otimes (\vec{Y}-\mathrm{E}(\vec{Y}))'\right] \\
&= \mathrm{E}\left[\left(W^{-\frac{1}{2}}\vec{Z}-\mathrm{E}(W^{-\frac{1}{2}})\mathrm{E}(\vec{Z})\right)\left(W^{-\frac{1}{2}}\vec{Z}-\mathrm{E}(W^{-\frac{1}{2}})\mathrm{E}(\vec{Z})\right)' \otimes \left(W^{-\frac{1}{2}}\vec{Z}-\mathrm{E}(W^{-\frac{1}{2}})\mathrm{E}(\vec{Z})\right)'\right] \\
&= \mathrm{E}\Big[\left(W^{-1}\vec{Z}\vec{Z}' - W^{-\frac{1}{2}}\mathrm{E}(W^{-\frac{1}{2}})\mathrm{E}(\vec{Z})\vec{Z}' - W^{-\frac{1}{2}}\mathrm{E}(W^{-\frac{1}{2}})\vec{Z}\mathrm{E}(\vec{Z})' + (\mathrm{E}(W^{-\frac{1}{2}}))^2\mathrm{E}(\vec{Z})\mathrm{E}(\vec{Z})'\right) \\
&\quad \otimes \left(W^{-\frac{1}{2}}\vec{Z}-\mathrm{E}(W^{-\frac{1}{2}})\mathrm{E}(\vec{Z})\right)'\Big] \\
&= \mathrm{E}\Big[W^{-\frac{3}{2}}\vec{Z}\vec{Z}' \otimes \vec{Z}' - W^{-1}\mathrm{E}(V^{-\frac{1}{2}})\left(\mathrm{E}(\vec{Z})\vec{Z}' \otimes \vec{Z}' + \vec{Z}\mathrm{E}(\vec{Z})' \otimes \vec{Z}' + \vec{Z}\vec{Z}' \otimes \mathrm{E}(\vec{Z})'\right) \\
&\quad + W^{-\frac{1}{2}}\left(\mathrm{E}(W^{-\frac{1}{2}})\right)^2\left(\mathrm{E}(\vec{Z})\mathrm{E}(\vec{Z})' \otimes \vec{Z}' + \mathrm{E}(\vec{Z})\vec{Z}' \otimes \mathrm{E}(\vec{Z})' + \vec{Z}\mathrm{E}(\vec{Z})' \otimes \mathrm{E}(\vec{Z})'\right) \\
&\quad - \left(\mathrm{E}(W^{-\frac{1}{2}})\right)^3\mathrm{E}(\vec{Z})\mathrm{E}(\vec{Z})' \otimes \mathrm{E}(\vec{Z})'\Big] \\
&= \mathrm{E}(W^{-\frac{3}{2}})\mathrm{E}(\vec{Z}\vec{Z}' \otimes \vec{Z}') + 2\left(\mathrm{E}(W^{-\frac{1}{2}})\right)^3\left(\mathrm{E}(\vec{Z})\mathrm{E}(\vec{Z})' \otimes \mathrm{E}(\vec{Z})'\right) \\
&\quad - \mathrm{E}(W^{-1})\mathrm{E}(W^{-\frac{1}{2}})\left(\mathrm{E}\left(\mathrm{E}(\vec{Z})\vec{Z}' \otimes \vec{Z}'\right) + \mathrm{E}\left(\vec{Z}\mathrm{E}(\vec{Z})' \otimes \vec{Z}'\right) + \mathrm{E}\left(\vec{Z}\vec{Z}' \otimes \mathrm{E}(\vec{Z})'\right)\right).
\end{aligned}$$

Noting that

$$\begin{aligned}
&\mathrm{E}\left(\mathrm{E}(\vec{Z})\vec{Z}' \otimes \vec{Z}'\right) = \mathrm{E}(\vec{Z}) \otimes \mathrm{vec}(\mathrm{E}(\vec{Z}\vec{Z}'))', \\
&\mathrm{E}\left(\vec{Z}\mathrm{E}(\vec{Z})' \otimes \vec{Z}'\right) = \mathrm{E}(\vec{Z}') \otimes \mathrm{E}(\vec{Z}\vec{Z}')
\end{aligned}$$

and

$$\mathrm{E}\left(\vec{Z}\vec{Z}' \otimes \mathrm{E}(\vec{Z})'\right) = \mathrm{E}(\vec{Z}\vec{Z}') \otimes \mathrm{E}(\vec{Z})',$$

we have the following simplified expression for the third moment of $\vec{Y}$ when $\vec{\xi} \neq \vec{0}$:

$$\begin{aligned}
&\mathrm{E}^{(3)}\left(\vec{Y}-\mathrm{E}(\vec{Y})\right) \\
&= \mathrm{E}(W^{-\frac{3}{2}})\mathrm{E}(\vec{Z}\vec{Z}' \otimes \vec{Z}') + 2\left(\mathrm{E}(W^{-\frac{1}{2}})\right)^3\left(\mathrm{E}(\vec{Z})\mathrm{E}(\vec{Z})' \otimes \mathrm{E}(\vec{Z})'\right) \\
&\quad - \mathrm{E}(W^{-1})\mathrm{E}(W^{-\frac{1}{2}})\left(\mathrm{E}(\vec{Z}) \otimes \mathrm{vec}(\mathrm{E}(\vec{Z}\vec{Z}'))' + \mathrm{E}(\vec{Z}') \otimes \mathrm{E}(\vec{Z}\vec{Z}') + \mathrm{E}(\vec{Z}\vec{Z}') \otimes \mathrm{E}(\vec{Z})'\right).
\end{aligned} \tag{23}$$

In the above expression, $\mathrm{E}(W^{-\frac{1}{2}})$, $\mathrm{E}(W^{-1})$ and $\mathrm{E}(W^{-\frac{3}{2}})$ can be obtained using equation (16), whereas $\mathrm{E}(\vec{Z})$, $\mathrm{E}(\vec{Z}\vec{Z}')$ and $\mathrm{E}(\vec{Z}\vec{Z}' \otimes \vec{Z}')$ can be obtained using equations (17), (18) and (19), respectively.

A.2.4. The fourth moment

The derivation for the fourth moment is similar to that for the third moment, and is therefore not shown for the sake of space. The fourth moment of $\vec{Y}$ for $\vec{\xi} \neq \vec{0}$ is presented below:

$$\begin{aligned}
&\mathrm{E}^{(4)}\left(\vec{Y}-\mathrm{E}(\vec{Y})\right) \\
&= \mathrm{E}\left[(\vec{Y}-\mathrm{E}(\vec{Y}))(\vec{Y}-\mathrm{E}(\vec{Y}))' \otimes (\vec{Y}-\mathrm{E}(\vec{Y}))' \otimes (\vec{Y}-\mathrm{E}(\vec{Y}))'\right] \\
&= \mathrm{E}(W^{-2})\mathrm{E}^{(4)}(\vec{Z}) - \mathrm{E}(W^{-\frac{3}{2}})\mathrm{E}(W^{-\frac{1}{2}})K_1 + \mathrm{E}(W^{-1})(\mathrm{E}(W^{-\frac{1}{2}}))^2K_2 - 3(\mathrm{E}(W^{-\frac{1}{2}}))^4K_3,
\end{aligned} \tag{24}$$

where

$$\begin{aligned}
K_1 &= \mathrm{E}(\vec{Z})\mathrm{vec}(\mathrm{E}^{(3)}(\vec{Z}))' + \mathrm{E}(\vec{Z})' \otimes \mathrm{E}^{(3)}(\vec{Z}) + \mathrm{E}^{(3)}(\vec{Z})\left(\mathbf{I}_d \otimes \mathrm{E}(\vec{Z})' \otimes \mathbf{I}_d\right) + \mathrm{E}^{(3)}(\vec{Z}) \otimes \mathrm{E}(\vec{Z})', \\
K_2 &= \mathrm{E}(\vec{Z})\mathrm{E}(\vec{Z})' \otimes \mathrm{vec}(\mathrm{E}(\vec{Z}\vec{Z}'))' + \left(\mathrm{E}(\vec{Z})\mathrm{E}(\vec{Z})'\right)(\mathrm{E}(\vec{Z}' \otimes \mathbf{I}_d \otimes \vec{Z}')) + \mathrm{vec}(\mathrm{E}(\vec{Z})\mathrm{E}(\vec{Z})')' \otimes \mathrm{E}(\vec{Z}\vec{Z}') \\
&\quad + \mathrm{E}(\vec{Z}) \otimes \mathrm{vec}(\mathrm{E}(\vec{Z}\vec{Z}'))' \otimes \mathrm{E}(\vec{Z})' + \mathrm{E}(\vec{Z}\vec{Z}')\left(\mathrm{E}(\vec{Z})' \otimes \mathbf{I}_d \otimes \mathrm{E}(\vec{Z})'\right) + \mathrm{E}(\vec{Z}\vec{Z}') \otimes \mathrm{E}(\vec{Z})' \otimes \mathrm{E}(\vec{Z})'
\end{aligned}$$

and

$$K_3 = \mathrm{E}(\vec{Z})\mathrm{E}(\vec{Z})' \otimes \mathrm{E}(\vec{Z})' \otimes \mathrm{E}(\vec{Z})'.$$

In the above expressions, $\mathrm{E}(W^{-\frac{1}{2}})$, $\mathrm{E}(W^{-1})$, $\mathrm{E}(W^{-\frac{3}{2}})$ and $\mathrm{E}(W^{-2})$ can be obtained using equation (16), whereas $\mathrm{E}(\vec{Z})$, $\mathrm{E}(\vec{Z}\vec{Z}')$, $\mathrm{E}(\vec{Z}^{(3)})$ and $\mathrm{E}(\vec{Z}^{(4)})$ can be obtained using equations (17) to (20), respectively.

Appendix B. Partial derivatives of L and $H(j,t)$

In this appendix, we derive the partial derivatives of L and $H(j,t)$, $j=1,\dots,m$, with respect to $\vec{\eta}(t+1)$. These partial derivatives, which are expressed in terms of the parameters in the general state-space representation of stochastic mortality models, are utilized in the approximation of $\mathbf{M}_t$, $\mathbf{V}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$ as discussed in Section 5.3.

B.1. The partial derivative of L with respect to $\vec{\eta}(t+1)$

Recall that

$$L=\sum_{u=1}^{T_L}\left(e^{-ru}\prod_{s=1}^{u}\tilde{p}(x_L+s-1,t_0+s)\right).$$

It is important to note that $\tilde{p}(x_L+s-1,t_0+s)$ in the expression above does not depend on $\vec{\eta}(t+1)$ for any $t_0+s<t+1$ (i.e., $s<t-t_0+1$). Accordingly, the first partial derivative of L with respect to $\vec{\eta}(t+1)$ can be written as

$$\frac{\partial L}{\partial\vec{\eta}(t+1)}=\sum_{u=t-t_0+1}^{T_L}e^{-ru}\frac{\partial}{\partial\vec{\eta}(t+1)}\left(\prod_{s=1}^{u}\tilde{p}(x_L+s-1,t_0+s)\right),$$

for $t=t_0,t_0+1,\dots,t_0+T_L-1$, where

$$\begin{aligned}&\frac{\partial}{\partial\vec{\eta}(t+1)}\left(\prod_{s=1}^{u}\tilde{p}(x_L+s-1,t_0+s)\right)\\&=\sum_{s=t-t_0+1}^{u}\left(\frac{\partial\tilde{p}(x_L+s-1,t_0+s)}{\partial\vec{\eta}(t+1)}\times\prod_{\substack{w=1\\w\neq s}}^{u}\tilde{p}(x_L+w-1,t_0+w)\right)\\&=\sum_{s=t-t_0+1}^{u}\left(\frac{\partial\tilde{p}(x_L+s-1,t_0+s)}{\partial\tilde{y}(x_L+s-1,t_0+s)}\times\frac{\partial\tilde{y}(x_L+s-1,t_0+s)}{\partial\vec{\eta}(t+1)}\times\prod_{\substack{w=1\\w\neq s}}^{u}\tilde{p}(x_L+w-1,t_0+w)\right).\end{aligned}\tag{25}$$

Here, $y(x,t)$ is the age-x-related element in the observation vector $\vec{y}(t)$ in the general state space representation of stochastic mortality models, and $\tilde{y}(x,t)$ is the value of $y(x,t)$ when the observation error $\epsilon(x,t)$ is excluded (being set to zero).

Equation (25) involves the partial derivative of $\tilde{p}(x_L+s-1,t_0+s)$ with respect to $\tilde{y}(x_L+s-1,t_0+s)$ and the partial derivative of $\tilde{y}(x_L+s-1,t_0+s)$ with respect to $\vec{\eta}(t+1)$. The former derivative depends on the specification of the observations in the model. If $y(x,t)=\ln(m(x,t))$ (e.g., the Lee-Carter model), then

$$\frac{\partial\tilde{p}(x_L+s-1,t_0+s)}{\partial\tilde{y}(x_L+s-1,t_0+s)}=(-1)\times\tilde{p}(x_L+s-1,t_0+s)\times\tilde{m}(x_L+s-1,t_0+s),$$

and if $y(x,t)=\ln(q(x,t)/(1-q(x,t)))$ (e.g., the Cairns-Blake-Dowd model), then

$$\frac{\partial\tilde{p}(x_L+s-1,t_0+s)}{\partial\tilde{y}(x_L+s-1,t_0+s)}=(-1)\times\tilde{p}(x_L+s-1,t_0+s)\times\tilde{q}(x_L+s-1,t_0+s).$$

The latter derivative can be derived by first expressing $\tilde{y}(x_L+s-1,t_0+s)$ in terms of the parameters in the general state-space mortality model as

$$\tilde{y}(x_L+s-1,t_0+s)=d(x_L+s-1)+\mathbf{B}(x_L+s-1,\cdot)\,\vec{\alpha}(t_0+s),\tag{26}$$

where $\mathbf{B}(x,\cdot)$ and $d(x)$ represent the row of $\mathbf{B}$ and the element of $\vec{d}$ that are related to age x, respectively, and then rewriting the state vector $\vec{\alpha}(t_0+s)$ in terms of the innovation vectors from time $t+1$ to t_0+s as

$$\vec{\alpha}(t_0+s)=\mathbf{A}^{t_0+s-t}\vec{\alpha}(t)+\left(\sum_{w=1}^{t_0+s-t}\mathbf{A}^{t_0+s-t-w}\left(\vec{c}+\vec{\eta}(t+w)\right)\right).\tag{27}$$

These two steps allow us to express $\tilde{y}(x_L+s-1,t_0+s)$ as follows:

$$\tilde{y}_t(x_L+s-1,t_0+s)=h+\mathbf{B}(x_L+s-1,\cdot)\left(\sum_{w=1}^{t_0+s-t}\mathbf{A}^{t_0+s-t-w}\vec{\eta}(t+w)\right),\tag{28}$$

for some constant h that is free of $\vec{\eta}(t+1)$, which in turn implies that

$$\frac{\partial\tilde{y}(x_L+s-1,t_0+s)}{\partial\vec{\eta}(t+1)}=\frac{\partial\mathbf{B}(x_L+s-1,\cdot)\left(\sum_{w=1}^{t_0+s-t}\mathbf{A}^{t_0+s-t-w}\vec{\eta}(t+w)\right)}{\partial\vec{\eta}(t+1)}=\left(\mathbf{B}(x_L+s-1,\cdot)\mathbf{A}^{t_0+s-t-1}\right)',$$

for $s\ge t-t_0+1$.

B.2. The partial derivative of $H(j,t)$ with respect to $\vec{\eta}(t+1)$

Recall that

$$H(j,t) = e^{-r\times(t-t_0+T_j)}(q^f(x_j,t+T_j) - \tilde{q}(x_j,t+T_j)),$$

for $j = 1,\dots,m$. Using the chain rule, the partial derivative of $H(j,t)$ with respect to $\vec{\eta}(t+1)$ can be expanded as follows:

$$\begin{aligned}\frac{\partial H(j,t)}{\partial \vec{\eta}(t+1)} &= \frac{\partial}{\partial \vec{\eta}(t+1)}\left(e^{-r\times(t-t_0+T_j)}\left(q^f(x_j,t+T_j) - \tilde{q}(x_j,t+T_j)\right)\right) \\ &= -e^{-r\times(t-t_0+T_j)} \times \frac{\partial \tilde{q}(x_j,t+T_j)}{\partial \vec{\eta}(t+1)} \\ &= -e^{-r\times(t-t_0+T_j)} \times \frac{\partial \tilde{q}(x_j,t+T_j)}{\partial \tilde{y}(x_j,t+T_j)} \times \frac{\partial \tilde{y}(x_j,t+T_j)}{\partial \vec{\eta}(t+1)}.\end{aligned} \tag{29}$$

In the above expression, the partial derivative of $\tilde{q}(x_j,t+T_j)$ with respect to $\tilde{y}(x_j,t+T_j)$ depends on the specification of the observations in the model. We have

$$\frac{\partial \tilde{q}(x_j,t+T_j)}{\partial \tilde{y}(x_j,t+T_j)} = \tilde{p}(x_j,t+T_j) \times \tilde{m}(x_j,t+T_j)$$

if $y(x,t) = \ln(m(x,t))$ and

$$\frac{\partial \tilde{q}(x_j,t+T_j)}{\partial \tilde{y}(x_j,t+T_j)} = \tilde{p}(x_j,t+T_j) \times \tilde{q}(x_j,t+T_j)$$

if $y(x,t) = \ln(q(x,t)/(1-q(x,t))$. Similar to the steps described in equations (26) to (28), we can express $\tilde{y}(x_j,t+T_j)$ in terms of the innovation vectors from time $t+1$ to $t+T_j$, and then differentiate the expression with respect to $\vec{\eta}(t+1)$ to obtain

$$\frac{\partial \tilde{y}(x_j,t+T_j)}{\vec{\eta}(t+1)} = \frac{\partial\left(\mathbf{B}(x_j,\cdot)\sum_{w=1}^{T_j}\mathbf{A}^{T_j-w}\vec{\eta}(t+w)\right)}{\partial \vec{\eta}(t+1)} = \left(\mathbf{B}(x_j,\cdot)\mathbf{A}^{T_j-1}\right)'.$$

Appendix C. Value-at-risk minimization

In this appendix, we describe the analytical solution provided by Liu and Li (2021) that solves the Value-at-Risk minimization prescribed in expression (12). Following the set-up established in the main text, we consider minimizing the Value-at-Risk of the hedged position 1-year after time t_0 when the hedge is established. The objective function can be expressed as follows:

$$\min_{N_{1,t_0},\dots,N_{4,t_0}} \mathrm{VaR}_\gamma\left(V_L(t_0+1) - \sum_{j=1}^{4}(N_{j,t_0}\times V_H(t_0+1,j,t_0))\middle|\mathcal{F}_{t_0}\right) \tag{30}$$

When the innovation vector follows a multivariate skew-t distribution, the optimization above is not straightforward to solve. To overcome this challenge, Liu and Li (2021) approximate the values of the liability and the hedge instruments using a first-order Taylor expansion, so that the objective function is rewritten as

$$\min_{N_{1,t_0},\dots,N_{4,t_0}} \mathrm{VaR}_\gamma\left(V_l(t_0+1) - \sum_{j=1}^{4}(N_{j,t_0}\times V_h(t_0+1,j,t_0))\middle|\mathcal{F}_{t_0}\right), \tag{31}$$

where $V_l(t_0+1)$ and $V_h(t_0+1,j,t_0)$ are the first-order Taylor approximations of $V_L(t_0+1)$ and $V_H(t_0+1,j,t_0)$, respectively. Using the expressions of $V_l(t_0+1)$ and $V_h(t_0+1,j,t_0)$ provided in Section 5.2, the (approximate) value of the hedge portfolio can be expressed as

$$\begin{aligned}& V_l(t_0+1) - \sum_{j=1}^{4}(N_{j,t_0}\times V_h(t_0+1,j,t_0)) \\ =\ & \hat{L} - \sum_{j=1}^{4} N_{j,t_0}\times\hat{H}_j + \sum_{k=1}^{n}\left(D_k(0,t_0+1) - \sum_{j=1}^{4}N_{j,t_0}\times D_k(j,t_0+1)\right)\times\eta_k(t_0+1)\,, \\ =\ & \hat{L} - \sum_{j=1}^{4} N_{j,t_0}\times\hat{H}_j + \vec{D}'_{t_0+1}\vec{\eta}_{t_0+1}\end{aligned} \tag{32}$$

where $\vec{\eta}_{t_0+1}$ is the innovation vector at time t_0+1, which follows an n-dimensional skew-t distribution $\mathrm{St}_n(\vec{\xi},\Omega,\vec{\zeta},\nu)$, and

$$\vec{D}_{t_0+1} = \begin{pmatrix} D_1(0,t_0+1) - \sum_{j=1}^{4}N_{j,t_0}\times D_1(j,t_0+1) \\ \vdots \\ D_n(0,t_0+1) - \sum_{j=1}^{4}N_{j,t_0}\times D_n(j,t_0+1)\end{pmatrix}$$

is the vector of the first-order partial derivatives of the hedged position with respect to the innovations.

Equation (32) is essentially a linear transformation of a multivariate skew-t distribution, which according to Azzalini and Capitanio (1999 and 2003), follows a skew-t distribution with modified parameters. In Equation (32), $\hat{L} - \sum_{j=1}^{4} N_{j,t_0} \times \hat{H}_j$ is a constant which represents the expectation of the (approximate) hedged position and

$$\vec{D}'_{t_0+1}\vec{\eta}_{t_0+1} \sim \mathrm{St}_1(\vec{\xi}^*, \Omega^*, \zeta^*, \nu), \tag{33}$$

where

$$\vec{\xi}^* = \vec{D}'_{t_0+1}\vec{\xi}$$
$$\Omega^* = \vec{D}_{t_0+1}\Omega\vec{D}'_{t_0+1}$$

and

$$\zeta^* = \frac{(\Omega^*)^{-1}\vec{D}'_{t_0+1}\omega\zeta}{\sqrt{1 + \zeta'(\omega - \omega\vec{D}_{t_0+1}(\omega^*)^{-1}\vec{D}'_{t_0+1}\omega)\zeta}}.$$

The optimization problem is then converted into finding the smallest 100γ-th percentile of a skew-t distribution, subject to the two constraints mentioned in Section 4. The constrained optimization can be solved using the “SN” package in R.

References

Affleck-Graves, J., McDonald, B., 1989. Nonnormalities and tests of asset pricing theories. The Journal of Finance 44 (4), 889–908.
Ahmadi, S.S., Gaillardetz, P., 2014. Two factor stochastic mortality modeling with generalized hyperbolic distribution. Journal of Data Science 12 (1), 1–18.
Azzalini, A., Capitanio, A., 1999. Statistical applications of the multivariate skew normal distribution. Journal of the Royal Statistical Society, Series B, Statistical Methodology 61 (3), 579–602.
Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. Journal of the Royal Statistical Society, Series B, Statistical Methodology 65 (2), 367–389.
Bakshi, G., Kapadia, N., Madan, D., 2003. Stock return characteristics, skew laws, and the differential pricing of individual equity options. The Review of Financial Studies 16 (1), 101–143.
Blake, D., Cairns, A., Coughlan, G., Dowd, K., MacMinn, R., 2013. The new life market. The Journal of Risk and Insurance 80 (3), 501–558.
Cairns, A.J., 2011. Modelling and management of longevity risk: approximations to survivor functions and dynamic hedging. Insurance: Mathematics and Economics 49 (3), 438–453.
Cairns, A.J., Blake, D., Dowd, K., 2006. A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. The Journal of Risk and Insurance 73 (4), 687–718.
Cairns, A.J., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., Balevich, I., 2009. A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. North American Actuarial Journal 13 (1), 1–35.
Cairns, A.J., Dowd, K., Blake, D., Coughlan, G.D., 2014. Longevity hedge effectiveness: a decomposition. Quantitative Finance 14 (2), 217–235.
Callan Institute, 2017. 2017 investment management fee survey. https://www.callan.com.
Chai, C.M., Siu, T.K., Zhou, X., 2013. A double-exponential GARCH model for stochastic mortality. European Actuarial Journal 3 (2), 385–406.
Chan, W.-S., Li, J.S.-H., Li, J., 2014. The cbd mortality indexes: modeling and applications. North American Actuarial Journal 18 (1), 38–58.
Duan, J.-C., 2016. Local-momentum autoregression and the modeling of interest rate term structure. Journal of Econometrics 194 (2), 349–359.
Fung, M.C., Peters, G.W., Shevchenko, P.V., 2017. A unified approach to mortality modelling using state-space framework: characterisation, identification, estimation and forecasting. Annals of Actuarial Science 11 (2), 343–389.
Fung, M.C., Peters, G.W., Shevchenko, P.V., 2018. Cohort effects in mortality modelling: a bayesian state-space approach. Annals of Actuarial Science, 1–36.
Gao, Q., Hu, C., 2009. Dynamic mortality factor model with conditional heteroskedasticity. Insurance: Mathematics and Economics 45 (3), 410–423.
Genton, M.G., He, L., Liu, X., 2001. Moments of skew-normal random vectors and their quadratic forms. Statistics & Probability Letters 51 (4), 319–325.
Giacometti, R., Bertocchi, M., Ortobelli Lozza, S., 2009. Impact of different distributional assumptions in forecasting Italian mortality rates. Investment Management and Financial Innovations 6 (2), 65–72.
Goliński, A., Zaffaroni, P., 2016. Long memory affine term structure models. Journal of Econometrics 191 (1), 33–56.
Harvey, C.R., Liechty, J.C., Liechty, M.W., Müller, P., 2010. Portfolio selection with higher moments. Quantitative Finance 10 (5), 469–485.
Healy, M., 1968. Multivariate normal plotting. Journal of the Royal Statistical Society. Series C. Applied Statistics 17 (2), 157–161.
Jondeau, E., Rockinger, M., 2006. Optimal portfolio allocation under higher moments. European Financial Management 12 (1), 29–55.
Justel, A., Peña, D., Zamar, R., 1997. A multivariate Kolmogorov-Smirnov test of goodness of fit. Statistics & Probability Letters 35 (3), 251–259.
Kim, S., Kim, H., 2016. A new metric of absolute percentage error for intermittent demand forecasts. International Journal of Forecasting 32 (3), 669–679.
Lai, K.K., Yu, L., Wang, S., 2006. Mean-variance-skewness-kurtosis-based portfolio optimization. In: Computer and Computational Sciences, 2006. IMSCCS'06. First International Multi-Symposiums on, vol. 2. IEEE, pp. 292–297.
Lee, R.D., Carter, L.R., 1992. Modeling and forecasting US mortality. Journal of the American Statistical Association 87 (419), 659–671.
Lee, T.K., Tse, Y.K., 1991. Term structure of interest rates in the Singapore Asian dollar market. Journal of Applied Econometrics 6 (2), 143–152.
Leung, M.T., Daouk, H., Chen, A.-S., 2001. Using investment portfolio return to combine forecasts: a multiobjective approach. European Journal of Operational Research 134 (1), 84–102.
Li, H., De Waegenaere, A., Melenberg, B., 2017. Robust mean–variance hedging of longevity risk. The Journal of Risk and Insurance 84 (S1), 459–475.
Li, J., 2013. A Poisson common factor model for projecting mortality and life expectancy jointly for females and males. Population Studies 67 (1), 111–126.
Li, J.S.-H., Hardy, M.R., 2011. Measuring basis risk in longevity hedges. North American Actuarial Journal 15 (2), 177–200.
Li, J.S.-H., Luo, A., 2012. Key q-duration: a framework for hedging longevity risk. ASTIN Bulletin: The Journal of the IAA 42 (2), 413–452.
Li, J.S.-H., Zhou, R., Hardy, M., 2015. A step-by-step guide to building two-population stochastic mortality models. Insurance: Mathematics and Economics 63, 121–134.
Ling, S., McAleer, M., 2002. Stationarity and the existence of moments of a family of GARCH processes. Journal of Econometrics 106 (1), 109–117.
Ling, S., Tong, H., 2011. Score based goodness-of-fit tests for time series. Statistica Sinica, 1807–1829.
Liu, Y., Li, J.S.-H., 2016. It's all in the hidden states: a longevity hedging strategy with an explicit measure of population basis risk. Insurance: Mathematics and Economics 70, 301–319.
Liu, Y., Li, J.S.-H., 2017. The locally linear Cairns–Blake–Dowd model: A note on delta–nuga hedging of longevity risk. ASTIN Bulletin: The Journal of the IAA 47 (1), 79–151.
Liu, Y., Li, J.S.-H., 2018. A strategy for hedging risks associated with period and cohort effects using q-forwards. Insurance: Mathematics and Economics 78, 267–285.
Liu, Y., Li, J.S.-H., 2021. An efficient method for mitigating longevity value-at-risk. North American Actuarial Journal 25 (sup1), S309–S340.
Michaelson, A., Mulholland, J., 2014. Strategy for increasing the global capacity for longevity risk transfer: developing transactions that attract capital markets investors. The Journal of Alternative Investments 17 (1), 18–27.
Mills, T.C., 1995. Modelling skewness and kurtosis in the London stock exchange FTSE index return distributions. The Statistician, 323–332.

Peiro, A., 1999. Skewness in financial returns. Journal of Banking & Finance 23 (6), 847–862.
Plat, R., 2011. One-year value-at-risk for longevity and mortality. Insurance: Mathematics and Economics 49 (3), 462–470.
Prakash, A.J., Chang, C.-H., Pactwa, T.E., 2003. Selecting a portfolio with skewness: recent evidence from US, European, and Latin American equity markets. Journal of Banking & Finance 27 (7), 1375–1390.
Renshaw, A.E., Haberman, S., 2006. A cohort-based extension to the Lee–Carter model for mortality reduction factors. Insurance: Mathematics and Economics 38 (3), 556–570.
Sears, R.S., 1985. Asset pricing, higher moments, and the market risk premium: a note. The Journal of Finance 40 (4), 1251–1253.
Sun, Q., Yan, Y., 2003. Skewness persistence with optimal portfolio selection. Journal of Banking & Finance 27 (6), 1111–1121.
Tsay, R.S., 2010. Analysis of Financial Time-Series, 3rd edition. Wiley, New York.
Tse, Y.K., 1998. An empirical analysis of the stochastic behaviour of short-term interest rates in Singapore. Asian Economic Journal 12 (1), 23–34.
Tuljapurkar, S., Li, N., Boe, C., 2000. A universal pattern of mortality decline in the g7 countries. Nature 405 (6788), 789.
Wang, C.-W., Huang, H.-C., Liu, I.-C., 2011. A quantitative comparison of the Lee-Carter model under different types of non-gaussian innovations. The Geneva Papers on Risk and Insurance. Issues and Practice 36 (4), 675–696.
Wang, C.-W., Huang, H.-C., Liu, I.-C., 2013. Mortality modeling with non-gaussian innovations and applications to the valuation of longevity swaps. The Journal of Risk and Insurance 80 (3), 775–798.
Wang, Z., Li, J.S.-H., 2016. A DCC-GARCH multi-population mortality model and its applications to pricing catastrophic mortality bonds. Finance Research Letters 16, 103–111.
Wilmoth, J.R., 1993. Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. Technical report. Department of Demography, University of California, Berkeley.
Xue, X., Zhang, J., Weng, C., 2018. Mean-variance hedging with basis risk. Applied Stochastic Models in Business and Industry. https://doi.org/10.1002/asmb.2380.
Zhang, J., Tan, K.S., Weng, C., 2017. Optimal hedging with basis risk under mean–variance criterion. Insurance: Mathematics and Economics 75, 1–15.
Zhou, K.Q., Li, J.S.-H., 2017. Dynamic longevity hedging in the presence of population basis risk: a feasibility analysis from technical and economic perspectives. The Journal of Risk and Insurance 84 (S1), 417–437.