# Robust longevity risk management

Hong Li[a,1], Anja De Waegenaere[a,b], Bertrand Melenberg[a,b]

[a]Department of Econometrics and Operations Research, Tilburg University
[b]Netspar

Preliminary draft

April 29, 2014

**Abstract**

We consider an annuity provider aiming to hedge her longevity risk using mortality-linked derivatives, such as survivor swaps. In particular, the annuity provider does not know the actual probability law governing the mortality dynamics in her portfolio, but instead considers a set of possible probability laws and optimizes her portfolio choice with respect to the worst-case scenario. We study the mean-variance and the conditional-value-at-risk formulations, and derive tractable reformulation of the robust optimization problems. The set of possible probability laws considered by the annuity provider is statistically given by the Kullback-Leibler divergence. We apply the robust optimization problem to Dutch male mortality data and compare its performance with optimization problems where the estimated probability law is assumed to be the true one. The robust optimization turns out to be superior as it yields better values of the objective functions and is more robust to the misspecification of the probability law. In the presence of basis risk, the situation of the insurer becomes worse, indicating that basis risk is an important factor affecting the longevity risk management. However, the robust optimization still outperforms the nominal optimization in this case.

**Keywords:** robust optimization, Kullback-Leibler divergence, mean-variance, conditional-value-at-risk

[1]Corresponding author. Email: h.li@tilburguniversity.edu.

# 1 Introduction

Pension plans and annuity providers (hereafter referred to as "insurer") are exposed to longevity risk, the risk that individuals live longer than anticipated. During recent years, longevity risk has gained greater attention from the pension and insurance industry due to demographic and economic changes. Bor and Cowling (2013) find that the total disclosed pension liabilities of the companies in the FTSE100 index increased by £31 billion in 2012, while the total deficit increased by £8 billion. The increase of pension liabilities is not only due to a low risk-free interest rate but also due to the unexpected increases in life expectancy. Moreover, a stricter regulatory environment, such as Solvency II (Olivieri and Pitacco 2009; Plat 2011), imposes more pressure on longevity risk management. Therefore, the financial sustainability of the insurer is at stake if the longevity risk is not controlled appropriately.

An insurer can choose to reduce her exposure to longevity risk by trading mortality-linked derivatives. For example, she can enter the survivor swap market as a fixed rate payer, so that she receives payoffs that increase with the improvement of realized life expectancy of her annuitants. In this way the insurer can offset (a part of) the unexpected increases in her liabilities. Mortality-linked derivatives can be divided into two categories. First, customized contracts are based on the actual mortality experience in the insurer's portfolio, so that the insurer might be able to transfer all of her longevity risk exposure to the capital market. However, such contracts are in general quite expensive and complex, and have very low liquidity. Second, standardized contracts are based on the mortality experience of a reference national population. Such standardized contracts are more transparent to investors and offer better liquidity. However, the insurer may not be able to fully transfer her exposures since standardized contracts might only be available for a subset of cohorts in her portfolio and, moreover, the insurer suffers from basis risk, i.e., the mismatch between the mortality experience in the reference population and her portfolios. For more examples of these two types of contracts, see, for example, Li and Luo (2012). We are particularly interested in standardized contracts since they are more realistic to insurers with smaller amounts of liabilities and also attractive to larger insurers who prefer better liquidities.

There are many studies involving longevity risk management in the literature. For example, Dahl et al. (2008) and Barbarin (2008) consider in a continuous time setting the longevity risk minimization problems with survivor swaps and longevity bonds, respectively. In discrete time settings, Dowd et al. (2011), Cairns (2013), and Cairns et al. (2014) study the static value hedging problem, while Cairns et al. (2008) study the static cash flow hedging problem of the longevity risk; Li and Luo (2012) propose a key-$q$ duration concept to hedge the insurers' longevity risk by trading standardized contracts contingent on a few key cohorts; Cox et al. (2013) propose a "MV + CVaR" approach to optimize the mean-variance trade-off while controlling the downside risk. However, in most of the aforementioned studies, the probability law governing the development of future mortality rates is assumed to be known by the insurers. The only exceptions are Cairns (2013), who derives robust hedging strategies with respect to the recalibration of mortality models, and Cox et al. (2013), who optimize against all mortality probability laws having the same moments.

The most widely used class of mortality models so far are based on linear extrapolation approaches (for a summary of some popular mortality models, see, for example, Cairns et al. (2009) and Cairns et al. (2011)), but the outcome of the linear extrapolation models are sensitive to the calibration window (Cairns et al. 2006; van Berkum et al. 2013; Li et al. 2013). Also, the improvement in life expectancy is speeding up during the last two decades in countries such as the Netherlands. As a consequence, the development of future mortality rates becomes even more uncertain. Therefore, the estimated probability law of the insurer may not very well represent the actual mortality developments. As we show in the numerical studies, a slight misspecification of the probability law may lead to substantial underperformance of the insurer's portfolio.

In this paper, we consider the longevity risk management of an insurer who does not know the actual probability law governing the mortality dynamics. Instead of treating the estimated probability measure as the true one, the insurer considers a set of probability laws which are "close" to the estimated law, and optimizes her portfolio choice with respect to the worst-case scenario. In particular, the set of probability laws is an approximate 95% confidence set determined by the Kullback-Leibler divergence. We consider the mean-variance and conditional-value-at-risk (CVaR) formulations, and derive tractable reformulations of the corresponding robust optimization problems. We apply our robust optimization problem to Dutch male mortality data and compare the performance of the robust optimization with their nominal counterparts, where the insurer optimizes under the assumption that the estimated probability law is the actual one. The robust optimizations turns out to be superior than the nominal optimizations as they yield better values of the objective functions and are more robust to the misspecification of the probability law. The inclusion of basis risk worsens the objective of the insurer by increasing the means and standard deviations of the optimal values of the objective functions, indicating that basis risk is an important risk factor affecting the insurer's longevity risk management. However, the robust optimization still outperforms the nominal ones in the presence of basis risk.

The remainder of this article is organized as follows. Section 2 describes out settings and the construction of the insurer's liability and the survivor swaps. Section 3 formulates the nominal and robust optimization problems.

Section 4 provides the tractable reformulation of the robust optimization problems. Section 5 reports the application of the robust and nominal optimization to Dutch male mortality data. Finally we conclude in section 6.

## 2 Liabilities and swaps

In this section we specify the cash flow of the liabilities of the insurer and of the survivor swaps. In this paper we consider two types of survivor swaps: the customized swaps and the standardized swaps. Without loss of generality, we assume that the only difference between a customized and a standardized swap contingent on the same cohort is that the payments from the former are related to the portfolio specific mortality experience of the insurer, while the payments from the latter are related to the reference population (such as a national population). Let $K = \{1, 2\}$ be the set of populations. For $k \in K$, denote by $k = 1$ the reference population, where the standardized swaps are based on, and $k = 2$ the portfolio specific population of the insurer. Denote by $p(t, x_j, k)$ the probability that an individual aged $x_j$ in year 0 in population $k$ is alive in year $t$.

Suppose that at time 0 an insurer has sold $n_i$ units of an annuity to a group of individuals aged $x_i$. Each of the annuities involves a commitment to pay 1 euro every year to the annuitant for the rest of his/her life. We assume that $n_i$-s are large enough so that the mortality experience of the insurer's portfolio can be well approximated by the mortality experience in the whole population in the absence of basis risk. Denote by $X$ the set of cohorts to which the annuities are sold, with $|X| = N$, and $r$ the fixed annual interest rate. $T$ is the terminal date such that no cash flow occurs after $T$. A convenient choice of $T$ is, for example, $T = 120 - \min_{x_j} X$. Denote by $y$ the $NT$ dimensional vector containing the (random) $p(t, x_j, k)$-s for all $t$, $i$, and $k$. The time 0 discounted (random) liability of the insurer can be written as

$$\tilde{L}(y) = \sum_{x_i \in X} n_i \sum_{t=1}^{T} \frac{p(t, x_j, 2)}{(1+r)^t}. \tag{2.1}$$

We assume that the insurer has the opportunity to hedge a portion of the uncertainty arising from her committed annuity payment by trading mortality-linked derivatives. One typical mortality-linked derivative is a survivor swap, proposed by, for example, Dowd et al. (2006) and Dawson et al. (2010). In the most basic setting, a survivor swap involves the exchange of a preset payment and a random mortality-dependent payment at a fixed frequency till maturity $\mu$. To be more precise, consider one unit of the survivor swap contingent on cohort $x_j$, $S(x_j, k)$.[2] Denote by $Fix(x_j, t, k)$

[2]The swap is a standardized swap in case $k = 1$ and a customized swap in case $k = 2$.

and $Flt(x_j, t, k)$ the preset and random payment, respectively, at year $t$ for $t \in \{1, 2, ..., \mu - 1, \mu\}$. The fixed rate payer pays $Fix(x_j, t, k) - Flt(x_j, t, k)$ to the floating rate payer if $Fix(x_j, t, k) - Flt(x_j, t, k) > 0$ and vice versa. Without loss of generality, we assume that $\mu = T$. With a slight abuse of notation, the time 0 discounted cash flows received by the fixed rate payer can be written as

$$S(x_j) = \sum_{t=1}^{T} \frac{Fix(x_j, t, k) - Flt(x_j, t, k)}{(1 + r)^t}. \tag{2.2}$$

Although the indexed mortality experience are different for these two types of swaps, the principle of valuation of the swap and the formulation of the hedging problem are the same. Therefore, in the sequel we denote the underlying mortality rates related to the survivor swaps by $p(t, x_j, k)$ for all $x_j$-s and $t$-s, where $k$ can be either 1 or 2. The effect of using different types of swaps will be discussed in the numerical study.

In a complete market setting with no arbitrage opportunities, we might be able to determine the value of the fixed rate and the floating rate by applying standard no-arbitrage arguments. However, this argument cannot be applied to the mortality-linked derivative market. Still at its infancy, the mortality-linked derivatives market is far from being complete and free from arbitrage opportunities. In fact, up to now, there are only few publicly traded mortality-linked derivatives that have been successfully issued (Dawson et al. 2010). Therefore, there is no obvious manner to determine the price of a survivor swap. One method of pricing a survivor swap is proposed by Dowd et al. (2006), who specify $Fix(x_j, t, k) = (1 + \tau_j)E_P[p(t, x_j, k)]$ and $Flt(x_j, t, k) = p(t, x_j, k)$. In other words, at year $t$, the preset payment is the best estimated $t$-year survival probability of the cohort $x_i$ times a number $(1 + \tau_j)$, while the random payment is the realized $t$-year survival probability. Denote by $V[p(t, x_j, k)]$ the present value of $p(t, x_j, k)$. If the value of $V[p(t, x_j, k)]$ is given, then one can determine the risk premium, $\tau_j$, such that the expected time 0 value (under the physical measure) of the swap is zero. Specifically, given a set of $V[p(t, x_j, k)]$-s, $\tau_j$ is then determined such that

$$\begin{aligned} V[S(x_j, k)] &= \sum_{t=1}^{T} \left( \frac{Fix(x_j, t)}{(1 + r)^t} - V[p(t, x_j, k)] \right) \\ &= \sum_{t=1}^{T} \left( \frac{(1 + \tau_j)E_P[p(t, x_j, k)]}{(1 + r)^t} - V[p(t, x_j, k)] \right) \\ &= 0 \end{aligned} \tag{2.3}$$

holds. Therefore, there is a relationship between $\tau_j$ and the $V[p(t, x_j, k)]$-s. The question now is how to determine the $V[p(t, x_j, k)]$ for all $t$ and $j$.

If the market would be compete and arbitrage opportunities are excluded, the no-arbitrage condition implies

$$V[p(t,x_j,k)] = \frac{E^Q[p(t,x_j,k)]}{(1+r)^t} \tag{2.4}$$

for all $x_j \in X$ and $t \in (1,2,...,T-1,T)$. In Equation (2.4), $Q$ is the unique risk neutral measure. However, completeness of the market dose not apply to the mortality-linked derivatives market. Therefore, $V[p(t,x_j,k)]$ should be determined such that it is acceptable to both parties of the contract. Dowd et al. (2006) and Dawson et al. (2010) propose a statistically based method to look for reasonable values of $V[Flt(x_j,t)]$-s. First, they fit a mortality model to the historical data and produce point forecasts of future mortality rates using the fitted model. Second, they distort the generated forecasts by some degree, to reflect the risk attitude of the investors. For example, if both parties believe that the $t$-year survival probability for cohort $x_j$ will always be 2% higher than anticipated, then $V[p(t,x_j,k)]$ would be $\frac{1.02p(t,x_j,k)}{(1+r)^t}$ for all $t$-s and we have $\tau_j = 2\%$ from Equation (2.3). Indeed, it is obvious that the risk premium will affect the insurer's decision of hedging. In later sections where we evaluate the effectiveness of the hedging, we will consider multiple reasonable values of the risk premium. However, for formulating our problem it suffices to assume a fixed risk premium. Therefore, for cohort $x_j$, the expected discounted cash flows of $S(x_j)$ at time 0 to the fixed rate payer is

$$\tau_j \sum_{t=1}^{T} \frac{E_P[p(t,x_j,k)]}{(1+r)^t}. \tag{2.5}$$

In an ideal situation, publicly traded survivor swaps are available for all cohorts $x_j \in X$. However, at the current stage, only for a few cohorts the corresponding derivative products are available. The reason is that, as the mortality-linked derivative market is not yet well developed, the liquidity of the derivative products are low and the issuance costs are high. Therefore, it is helpful at the current stage that the market concentrates liquidity on products that are contingent on a portion of the cohorts. To incorporate this fact, we assume that there is a set $X_S \in X$ with $|X_S| = m < N$ such that $S(x_j)$ exists only for the cohorts $x_j \in X_S$.

Aiming to hedge the longevity risks in her liabilities, the insurer acts as a fixed rate payer. Denote by $a_j$ the units of $S(x_j)$ the insurer holds at time 0, then her discounted random liabilities are

$$L(y,a,k) = \sum_{x_i \in X} n_i \sum_{t=1}^{T} \frac{p(t,x_j,2)}{(1+r)^t} + \sum_{x_j \in X_S} a_j \sum_{t=1}^{T} \frac{(1+\tau_j)E_P[p(t,x_j,k)] - p(t,x_j,k)}{(1+r)^t}, \tag{2.6}$$

where $k = 1, 2$ means the insurer hedges with standardized survivor swaps and customized swaps, respectively. In this way, the time 0 discounted original liabilities, $\tilde{L}(y)$, can be understood as $L(y, 0, k)$ (in this case $k$ does not matter), i.e., the liabilities when the insurer does not hold any swap.

In the sequel, we simplify the notation by letting $v(y, k)$ be a $m \times 1$ vector with the $j$-th entry $v_j(y, k) = \sum_{t=1}^{T} \frac{(1+\tau_j)E_P[p(t,x_j,k)]-p(t,x_j,k)}{(1+r)^t}$ and $a$ a $m \times 1$ vector with $j$-th entry $a_j$. The liabilities can then be written as

$$L(y, a, k) = \tilde{L}(y) + a'v(y, k). \tag{2.7}$$

Furthermore, we assume that the insurer will not hold negative positions of the swaps and she will pay a maximal amount $d$ to hedge her longevity risk. In other words, we have the constraint

$$a \in \mathbf{A}(k) \tag{2.8}$$

with

$$\mathbf{A}(k) = \{a \in R^m | a_i \geq 0,\ \forall i = \{1, 2, ..., m\} \text{ and } \sum_{x_j \in X_S} a_j \tau_j \sum_{t=1}^{T} \frac{E_P[p(t, x_j, k)]}{(1+r)^t} \leq d\}. \tag{2.9}$$

In (2.9), $E_P[.]$ denotes the expectation under the best estimated (physical) measure. In the numerical study we will look at both situations where $k = 1$ and 2.

# 3 Optimization problem

At time 0, the future survival probabilities in the two populations, $y$, are not yet revealed, but only known up to a distribution parameterized by $\theta$, i.e., $y \sim P_\theta$. Furthermore, the insurer does not know the value of $\theta$, but instead estimates $\theta$ using all information available at time 0. Denote by $\hat{\theta}$ the insurer's best estimate of $\theta$. In classical decision-making problems (only) under risk, the actual probability measure of the random quantities is assumed to be known to the decision maker (but not the econometrician). More precisely, the insurer assumes that $P_{\hat{\theta}}$ is the true distribution of $y$. For example, if our insurer is risk neutral, i.e., she is only concerned about the expected value of her liabilities, then her optimization problem is

$$\begin{aligned} \min_a \ & E_{\hat{\theta}}[L(y, a, k)] \\ s.t.\ & a \in \mathbf{A}(k) \end{aligned} \tag{3.1}$$

where $E_{\hat{\theta}}[.]$ is the expectation under $P_{\hat{\theta}}$ and $\mathbf{A}(k)$ is defined in (2.9). In this case the optimal solution is $a^* = 0$ if $\tau_i \geq 0$ for all $i$. The reason is that a positive position of any swaps would increase the insurer's expected liability.

In other words, a risk neutral insurer would not hedge any longevity risk, even if she could, when there is a positive cost to do so.

However, the result mentioned above will not necessarily happen if ambiguity enters the decision-making problem. In the presence of ambiguity, the actual probability measure is unknown to the decision maker (in the same way as it is unknown to the econometrician). In this case, we assume that the insurer treats $P_{\hat{\theta}}$ as the nominal measure, which is subject to estimation error. The insurer considers a set of measures, $\mathbf{P} = \{P_\theta | \theta \in \Theta\}$, where $\Theta$ is a set containing all values of $\theta$ which the insurer takes into account. Typically, one assumes that all measures in $\mathbf{P}$ should be absolute continuous with respect to $P_{\hat{\theta}}$. In other words, if $P_{\hat{\theta}}$ assigns zero probability to a certain event, then all measures in $\mathbf{P}$ assign also zero probability to that event.

Suppose that the insurer wants to minimize the expected discounted liability under the worst-case scenario, then the optimization problem becomes[3]

$$\begin{aligned} \inf_a \sup_\theta \quad & E_\theta[L(y, a, k)] \\ s.t. \quad & a \in \mathbf{A}(k) \\ & \theta \in \Theta. \end{aligned} \tag{3.2}$$

In this case, we do not necessarily have $a^* = 0$ even if all risk premiums are positive. In fact, without any assumption on $\Theta$, a solution to problem (3.2) may not even exist.

The maximin-type formulation in (3.2) is a special case of the multi-prior specification proposed by Gilboa and Schmeidler (1989). There are other popular formulations of decision-making problems under ambiguity, such as the variational preferences (Maccheroni et al. 2006) or the multiplier preferences (Hansen and Sargent 2001; Hansen and Sargent 2008). In the sequel, we shall refer to the maximin optimization as robust optimization, and the optimization problems where a fixed $P_\theta$ (such as the $P_{\hat{\theta}}$ in (3.1)) is taken as the true probability measure as the nominal optimization.

In this paper, we focus on two specifications: the mean-variance and the conditional-value-at-risk (CVaR).

### 3.1 Mean-variance

Since $\mathbf{A}(k)$ is compact, we can replace inf by min in the optimization problem. Therefore, for any $\theta$, the nominal mean-variance problem can be for-

[3] In this paper we assume that the expected $t$-year survival probabilities in $\mathbf{A}$ are calculated under $P_{\hat{\theta}}$, and are thus independent of $\Theta$. However, in general, one can allow $\mathbf{A}$ to depend on $\theta$ in (3.2) by replacing $E_{\hat{\theta}}[p(t, x_j, k)]$ by $E_\theta[p(t, x_j, k)]$. This generalization can be easily incorporated in our tractable formulation of robust optimizations introduced in the next section.

mulated as

$$\begin{aligned} \min_{a} \;\; & E_\theta[L(y,a,k)] + \lambda \mathrm{Var}_\theta[L(y,a,k)] \\ s.t. \;\; & a \in \mathbf{A}(k), \end{aligned} \tag{3.3}$$

where $\lambda$ is an exogenously given risk aversion parameter. For a compact parameter set, $\Theta$, the robust mean-variance problem is

$$\begin{aligned} \min_{a} \max_{\theta} \;\; & E_\theta[L(y,a,k)] + \lambda \mathrm{Var}_\theta[L(y,a,k)] \\ s.t. \;\; & a \in \mathbf{A}(k) \\ & \theta \in \Theta, \end{aligned} \tag{3.4}$$

where, similarly, the sup operator is replaced by max.

## 3.2 Conditional-Value-at-Risk

Given a confidence level $\alpha$ with $0 < \alpha < 1$ and a fixed $\theta$, the value-at-risk for $L(y,a,k)$ is defined as

$$\mathrm{VaR}_\alpha(a,\theta,k) = \min\{d \in R : \int_{L(y,a,k)\leq d} P_\theta(dy) \geq \alpha\}. \tag{3.5}$$

$\mathrm{VaR}_\alpha(a,\theta,k)$ is well defined if $E_\theta[|L(y,a,k)|] < \infty$ for any $a$ and $\theta$. From (2.6) and (2.9), we see that this assumption holds.[4] The CVaR for $L(y,a,k)$ is

$$\mathrm{CVaR}_\alpha(a,\theta,k) = \frac{1}{1-\alpha} \int_{L(y,a,k)\geq \mathrm{VaR}_\alpha(a,k)} L(y,a,k) P_\theta(dy), \tag{3.6}$$

and is also well defined.

Rockafellar and Uryasev (2000) show that the minimization of the CVaR function over $a \in \mathbf{A}$ is equivalent to the minimization of the following function

$$F_\alpha(a,\xi,\theta,k) = \xi + \frac{1}{1-\alpha} \int [L(y,a,k) - \xi]^+ P_\theta(dy) \tag{3.7}$$

over $(a,\xi) \in \mathbf{A}(k) \times R$. Moreover, it is reasonable to assume that, while minimizing her CVaR, the insurer is also concerned about the expected value of the liabilities. In other words, the insurer's objective is to minimize

$$\text{Mean-CVaR}_\alpha(a,\theta,k) = E_\theta[L(y,a,k)] + \lambda \mathrm{CVaR}_\alpha(a,\theta,k), \tag{3.8}$$

[4] In fact, $E_P[|L(y,a,k)|] < \infty$ holds for any probability measure $P : \mathcal{B}([0,1]^{NT}) \to [0,1]$ and $a \in \mathbf{A}(k)$, where $\mathcal{B}([0,1]^{NT})$ is the Borel $\sigma$-algebra over $[0,1]^{NT}$. The reason is that every $p(t,x_j,k) \in [0,1]$ and $\mathbf{A}(k)$ is also compact.

where $\lambda$ is the exogenous risk aversion parameter. It can be shown that, for a given $\theta$, minimizing (3.8) over $a \in \mathbf{A}(k)$ is equivalent to the following nominal optimization problem

$$\begin{aligned} \min_{(a,\xi)} \quad & E_\theta[L(y,a,k)] + \lambda F_\alpha(a,\xi,\theta,k) \\ s.t. \quad & a \in \mathbf{A}(k) \\ & \xi \in R. \end{aligned} \tag{3.9}$$

The proof is provided in the Appendix. With a slight abuse of notation, we refer to the minimization of the "Mean-CVaR" problem as the CVaR optimization problem in the sequel. The robust optimization problem is given by

$$\begin{aligned} \min_{(a,\xi)} \max_{\theta} \quad & E_\theta[L(y,a,k)] + \lambda F_\alpha(a,\xi,\theta,k) \\ s.t. \quad & \theta \in \Theta \\ & a \in \mathbf{A}(k) \\ & \xi \in R. \end{aligned} \tag{3.10}$$

### 3.3 Discretization

Before we further explore the robust optimizations, we would like to first make sure that the nominal optimizations, i.e., problems (3.3) and (3.9), are tractable. The most difficult part in solving (3.3) and (3.9) is the calculation of the integrals. The reason is that $y$ is a high dimensional vector, and, moreover, $L(y,a,k)$ is a nonsmooth function of $y$ for each $k$. To tackle this problem, Rockafellar and Uryasev (2000) use Monte Carlo simulation method and approximate (3.7) as

$$\tilde{F}_\alpha(a,\xi,\theta,k) = \xi + \frac{1}{I(1-\alpha)} \sum_{i=1}^{I} [L(y_i,a,k) - \xi]^+, \tag{3.11}$$

where $y_i$ is the $i$-th draw of $y$ from the random sample according to the distribution $P_\theta$. The law of large number guarantees that $\tilde{F}_\alpha(a,\xi,\theta,k)$ will converge to $F_\alpha(a,\xi,\theta,k)$ in probability as $I$ goes to infinity. Moreover, Zhu and Fukushima (2009) approximate (3.7) by explicitly assuming that the distribution of $y$, $P_{\hat{\theta}}$, is discrete. As a result, they deal with

$$\tilde{G}_\alpha(a,\xi,\pi,k) = \xi + \frac{1}{(1-\alpha)} \sum_{i=1}^{I} \pi_i [L(y_i,a,k) - \xi]^+, \tag{3.12}$$

where $\pi$ is a $I \times 1$ probability vector of $\pi_i$-s.

In this paper, we discretize the continuous distribution $P_\theta$. First, write

$$L(y,a,k) = \tilde{L(y)} + v(y,k)'a = (\tilde{L(y)}\ v(y,k)')(1\ \ a)' \equiv z'(1\ \ a)' \equiv \bar{L}(z,a,k). \tag{3.13}$$

The first block of $z$ is the time 0 discounted original liability and the second block are the time 0 discounted cash flow from a unit of survivor swap contingent on the cohorts in $X_S$. The function $f: y \mapsto z$ induces a probability measure, $P_{\hat{z}}$, over $z$. We then discretize $P_{\hat{z}}$ by a $I \times 1$ vector of realizations, $\boldsymbol{z}$, and a $I \times 1$ probability vector, $\pi^0$.[5] The uncertainty set of probabilities vectors is denoted by $\Pi$.

Denote by $\bar{\boldsymbol{L}}(z,a,k)$ the $I \times 1$ vector where its $i$-th entry is $\bar{L}(z_i,a,k)$, the nominal optimization problem can be written as

$$\begin{aligned} \min_{a} \;\; & \pi^{0\prime}\bar{\boldsymbol{L}}(z,a,k) + \lambda\pi^{0\prime}(\bar{\boldsymbol{L}}(z,a,k) - \pi^{0\prime}\bar{\boldsymbol{L}}(z,a,k))^2 \\ s.t. \;\; & a \in \mathbf{A}(k), \end{aligned} \tag{3.14}$$

for the mean-variance specification, and

$$\begin{aligned} \min_{a,\xi} \;\; & \pi^{0\prime}\bar{\boldsymbol{L}}(z,a,k) + \lambda[\xi + \frac{1}{1-\alpha}\sum_{i=1}^{I}\pi_i^0(\bar{L}(z_i,a,k) - \xi)^+] \\ s.t. \;\; & a \in \mathbf{A}(k) \\ & \xi \in R, \end{aligned} \tag{3.15}$$

for the CVaR specification. In (3.15), the $(\bar{L}(z_i,a,k) - \xi)^+$ parts are nonlinear in $(a,\xi)$. To turn (3.15) into a linear programming problem, we rewrite it as

$$\begin{aligned} \min_{a,\xi} \;\; & \pi^{0\prime}\bar{\boldsymbol{L}}(z,a,k) + \lambda[\xi + \frac{1}{1-\alpha}\sum_{i=1}^{I}\pi_i^0 u_i] \\ s.t. \;\; & a \in \mathbf{A} \\ & \xi \in R \\ & u_i \geq \bar{L}(z_i,a,k) - \xi, \;\; \forall\, i \in \{1,2,...,I\} \\ & u_i \geq 0, \;\; \forall\, i \in \{1,2,...,I\}. \end{aligned} \tag{3.16}$$

Similarly, the robust optimization problem can be written as

$$\begin{aligned} \min_{a}\max_{\pi} \;\; & \pi'\bar{\boldsymbol{L}}(z,a,k) + \lambda\pi'(\bar{\boldsymbol{L}}(z,a,k) - \pi'\bar{\boldsymbol{L}}(z,a,k))^2 \\ s.t. \;\; & a \in \mathbf{A}(k) \\ & \pi \in \Pi, \end{aligned} \tag{3.17}$$

[5] The detailed procedures will be explained in next section.

for the mean-variance specification and

$$
\begin{aligned}
\min_{a,\xi}\max_{\pi}\quad & \pi'\bar{\boldsymbol{L}}(z,a,k)+\lambda[\xi+\frac{1}{1-\alpha}\sum_{i=1}^{I}\pi_i u_i] \\
s.t.\quad & a\in\mathbf{A} \\
& \xi\in R \\
& u_i \geq \bar{L}(z_i,a,k)-\xi,\ \ \forall\, i\in\{1,2,...,I\} \\
& u_i \geq 0,\ \ \forall\, i\in\{1,2,...,I\} \\
& \pi\in\Pi, \qquad (3.18)
\end{aligned}
$$

for the CVaR specification.

# 4 Tractable reformulation of the robust optimization problems

In this section we derive tractable reformulations of problems (3.17) and (3.18).

## 4.1 The construction of the uncertainty set

The structure of the uncertainty set, $\Pi$, is crucial when studying decision-making problems under ambiguity, since, on the one hand, it reflects the decision maker's attitude to ambiguity and, on the other hand, it determines the tractability of the optimization problem. There are many popular structures of the uncertainty set in the literature. For example, Zhu and Fukushima (2009) study the worst-case CVaR problem using box uncertainty, ellipsoidal uncertainty, and mixture distribution uncertainty; Ben-Tal et al. (2013) study robust optimization problems with objective functions that are linear in the uncertain variables using various uncertainty sets defined by a $\phi$ divergence; Laeven and Stadje (2013) and Laeven and Stadje (2012) study risk measures and dynamic portfolio choice problems using uncertainty sets defined by relative entropy.

Among the various structures, we are particularly interested in the uncertainty set characterized by the Kullback-Leibler divergence (also known as relative entropy), due to its popularity in statistics (Liese and Vajda 2006; Reid and Williamson 2011), insurance and financial mathematics (Föllmer and Schied 2004; Mania et al. 2005; Laeven and Stadje 2013; Laeven and Stadje 2012), macroeconomics (Hansen and Sargent 2001; Hansen and Sargent 2008), and decision makings (Gollier 2004; Ben-Tal et al. 2013).

The uncertainty set characterized by the Kullback-Leibler divergence can be written as

$$\Pi = \{\pi \in R^I | \pi_i \geq 0 \; \forall i, \; \sum_{i=1}^{I} \pi_i \log(\frac{\pi_i}{\pi_i^0}) \leq \rho\}. \tag{4.1}$$

From (4.1), we see that the degree of divergence between the candidate probability vectors and the nominal one is determined by a single parameter, $\rho$. Hence, $\rho$ can be interpreted as the degree of ambiguity aversion of the decision maker. Preferably, $\rho$ should be chosen in a statistically relevant way. Ben-Tal et al. (2013) argue that, if the true distribution of $z$, $\pi^z$, belongs to a parameterized set of probability distributions $\{\pi^\theta | \theta \in \Theta \subset R^d\}$, i.e., there exists a $\theta \in \Theta$ such that $\pi^z = \pi^{\theta_0}$, then

$$\rho = \frac{\chi^2_{d,1-\beta}}{2N} \tag{4.2}$$

is a reasonable choice as $\Pi$ in (4.1) is an approximate $(1-\beta)$-confidence set if $\pi^0$ is replaced by an appropriate estimate. In (4.2), $\chi^2_{d,1-\beta}$ is the $1-\beta$ percent critical value of a $\chi^2$ distribution with degree of freedom $d$, and $N$ is the sample size used to estimate $\theta$. For a more detailed motivation of this choice we refer to Ben-Tal et al. (2013). In this paper we take the value of $\rho$ as in (4.2).

After specifying $\Pi$ and $\rho$, we follow the method proposed in Ben-Tal et al. (2013) and Ben-Tal et al. (2012) to derive the tractable reformulations of the robust optimization problems (3.17) and (3.18). The problems can be reformulated as

$$\begin{aligned} \min_{a,\eta,\mathcal{K},\xi} \quad & \eta\rho + \xi + \eta \sum_{i=1}^{I} \pi_i^0 \exp\left(\frac{\lambda \bar{L}^2(z_i,a,k) + (\mathcal{K}+1)\bar{L}(z,a,k) - \xi}{\eta} - 1\right) + \frac{\mathcal{K}^2}{4\lambda} \\ s.t. \quad & a \in \mathbf{A}(k) \\ & \mathcal{K} \in R; \\ & \xi \in R; \\ & \eta > 0 \end{aligned} \tag{4.3}$$

for the mean-variance specification and

$$\begin{aligned} \min_{a,\xi,u,\zeta,\eta} \quad & \lambda\xi + \rho\zeta + \eta + \zeta \sum_{i=1}^{I} \pi_i^0 \exp\left(\frac{\bar{L}(z_i,a,k) + \frac{\lambda}{1-\alpha} u_i - \eta}{\zeta} - 1\right) \\ & a \in \mathbf{A}(k) \\ & \xi \in R \\ & \eta \in R \\ & u_i \geq \bar{L}(z_i,a,k) - \xi, \quad \forall i \in \{1,2,...,I\} \\ & u_i \geq 0, \quad \forall i \in \{1,2,...,I\} \\ & \zeta \geq 0 \end{aligned} \tag{4.4}$$

for the CVaR specification. For detailed derivations see Appendix.

### 4.2 Properties of the robust counterpart

Before continuing to the numerical studies, we take a deeper look at the optimization problems to gain some intuition. By looking at the Lagrangian functions of the two problems, we can obtain closed form representations of the worst-case probabilities. For the CVaR case, denote by $(a^*, \xi^*, u^*, \zeta^*, \eta^*)$ the solution to (4.4). The worst-case probability is characterized by $\pi_i^* = \pi_i^0 \exp(\frac{L(z_i,a^*,k)+\frac{\lambda}{1-\alpha}u_i^*-\eta^*}{\zeta^*} - 1)$ for all $i$-s. Similarly, for the Mean-Variance case, denote by $(a^*, \eta^*, \mathcal{K}^*, \xi^*)$ the solution to (4.3), the worst-case probability is characterized by $\pi_i^* = \pi_i^0 \exp(\frac{\lambda L^2(z_i,a^*,k)+(\mathcal{K}^*+1)L(z_i,a^*,k)-\xi^*}{\eta^*} - 1)$. The explicit derivation of the above result is reported in the Appendix.

Therefore, we see that the robust optimization is actually the nominal optimization with the nominal probability vector replaced by the worst-cast counterpart.

### 4.3 The nominal distribution

In order to specify $\pi^0$, we need to fit a parametric mortality model to the observed mortality data. During the past twenty years, various mortality models for single or multiple populations have been proposed in the literature. For example, see Lee and Carter (1992), Cairns et al. (2006), and Cairns et al. (2009) for single population, and Li and Lee (2005), Cairns et al. (2011), Dowd et al. (2011), and Plat (2009) for multi-population mortality modelling. In particular, the insurer believes that the mortality rates in the reference population ($k = 1$) are generated by the Lee-Carter model (Lee and Carter 1992) and the mortality rates in her portfolio ($k = 2$) is generated by the method proposed in Plat (2009).[6] Nevertheless, the insurer recognizes that the estimations of the parameters in these models are only approximations to the true values, and thus not fully reliable.

#### 4.3.1 The reference population

Denote by $m(s, x_j, 1)$ the crude death rate in year $s$ of the cohort aged $x_j$ in year 0 in the reference population, then we have the well known approximation (Pitacco et al. 2009)

$$p(t, x_j, 1) = \exp(-\sum_{s=1}^{t} m(s, x_j, 1)).$$

[6] We make this assumption only for illustrating how our model works. In reality one may have different beliefs. For example, one may compare a number of different models, and choose the one which yields the highest BIC while fitted to the observed data as the candidate model.

The Lee-Carter model (Lee and Carter 1992) models the log of the crude death rate as[7]

$$\begin{aligned}\log(\boldsymbol{m}_t) =& \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(0, \Sigma_\varepsilon) \\ \kappa_t =& d + \kappa_{t-1} + \omega_t, \quad \omega_t \overset{iid}{\sim} N(0, \sigma_\omega^2).\end{aligned} \tag{4.5}$$

In (4.5), $\log(\boldsymbol{m}_t)$ is a vector containing the log of the age specific central death rates at year $t$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ are vectors of parameters, $\boldsymbol{\varepsilon}_t$ and $\omega_t$ are mutually independent i.i.d shocks, and $\sigma_\omega^2$ and $\Sigma_\varepsilon$ are the variance and covariance matrix. The $\kappa$ process capture the common time varying trend of the central death rates, and it is modeled by a random walk with drift process with drift term $d$ and volatility $\sigma_\omega$. Moreover, we assume $\Sigma_\varepsilon$ to be a diagonal matrix.

### 4.3.2 The portfolio specific population

First, denote by $p(t, t-1, x_j, k)$ the probability that an individual in population $k$, aged $x_j$ in year 0, and alive in year $t-1$ survives in year $t$. In this way, $p(t, t-1, x_j, k)$ is the $(t-1)$-conditional one year survival probability, while $p(t, x_j, 1)$ is the (0-conditional) $t$ year survival probability. The $(t-1)$-conditional one year death probability can thus be written as $q(t, t-1, x_j, k) = 1 - p(t, t-1, x_j, k)$.

Denote by $q^f(t, t-1, x_j) \equiv \frac{q(t,t-1,x_j,2)}{q(t,t-1,x_j,1)}$ the mortality factor of the reference population over the portfolio population for all $x_j$ and $t$, Plat (2009) fits the mortality factor by a one factor linear model[8]

$$\begin{aligned}\boldsymbol{q}_t^f =& 1 + \boldsymbol{w}\vartheta_t + \boldsymbol{\varepsilon}_t^f \\ \vartheta_t =& \delta + \omega_t^f.\end{aligned} \tag{4.6}$$

In (4.6), $\boldsymbol{q}_t^f$ is a $N \times 1$ vector with its $j$-th entry $q^f(t, t-1, x_j)$, $\boldsymbol{w}$ is a $N \times 1$ weighting vector with the $j$-th entry $w_j = 1 - \frac{x_j - x}{\bar{x} - \underline{x}}$, where $\bar{x}$ and $\underline{x}$ are the oldest and youngest cohort in $X$, respectively. $\vartheta_t$ is a parameter to be estimated, and $(\vartheta_t)_t$ is modeled as a ARIMA(0,0,0) process with mean $\delta$. Moreover, $\boldsymbol{\varepsilon}_t^f \overset{iid}{\sim} N(0, \Sigma_{\varepsilon^f})$ and $\omega_t^f \overset{iid}{\sim} N(0, \sigma_{\omega^f}^2)$. We call (4.6) the mortality factor model. According to (4.6), we have the following relation

$$p(t, x_j, 2) = \prod_{s=1}^{t} (1 - p(s, s-1, x_j, 1) q^f(s, s-1, x_j)). \tag{4.7}$$

[7]In general the $\kappa$ process in (4.5) can be modeled as any ARIMA(p,d,q) process. However, the authors state that a ARIMA(0,1,0) serves as a reasonable choice. ARIMA(0,1,0) is also the most widely used specification of the Lee-Carter model in the literature of mortality modelling.

[8]Similar to the Lee-Carter introduced in (4.5), model (4.6) can be generated to a multi-factor model, with the state variable $\vartheta_t$ following a general $ARIMA(p, d, q)$ process. However, the author fits the form (4.6) to the data, so we adopt the same form in this paper.

### 4.3.3 Construction of the nominal distribution

The nominal distribution, $\pi^0$, can be obtained as follows.

1. Obtain $\hat{\theta}$ by fitting the Lee-Carter model (4.5) and the mortality factor model (4.6) to observed data.

2. Simulate $M$ realizations of $z$ using the two models with the values of the parameters given by $\hat{\theta}$. Specifically, for each realization, we first simulate the log of central death rates, and transfer them to the $t$-year survival probabilities. The $t$-year survival probabilities in the portfolio population can be obtained using (4.7). Finally, we map these survival probabilities to $z$.

3. Divide the range of the $M$ realizations of $z$ into $I$ subsets, and calculate the frequency of the $M$ simulated $z$-s that fall into every subset. $\pi_i^0$ equals the frequency of the subset $i$ for all $i$-s.

4. The corresponding $z_i$ is obtained by taking the average of all $z$-s that fall into subset $i$. Denote by $\boldsymbol{z}$ the vector of $z_i$-s.

5. The value of $\rho$ can be determined by (4.2) by setting $d = |\theta|$ and $N$ the sample size used to estimate $\theta$.

In principle, we can assume that all parameters in (4.5) and (4.6) are estimated with errors. However, as stated in Cairns (2013), most uncertainty of the (single population) mortality forecasting comes from the estimation of the drift term of the $\kappa$ process. Also, Lee and Carter (1992) only take the uncertainty from the $\kappa$ process into account when calculating the forecast confidence intervals. Therefore, we assume that all parameters except for $d$ and $\sigma_\omega$ are estimated without error in (4.5). Moreover, we assume that only $\delta$ in (4.6) is estimated with error for the same reason. Our assumption leads to $\theta = (d, \sigma_\omega^2, \delta)'$ and $d = 3$.

# 5 Application to the mortality data

We evaluate the performance of our robust optimization problems by applying them to real mortality data. In this paper, we use historical mortality data for the Dutch males as the reference population.[9] Furthermore, Plat (2009) fits model (4.6) to the mortality data for the Dutch males population and an insurance portfolio, which is a collection of collective pension portfolios of the Dutch insurers containing about 100,000 male policyholders aged 65 or older. Therefore, we use the estimation results from Plat (2009) as

[9] The data we use is downloaded from Human Mortality Database (http://www.mortality.org/).

an illustration in our numerical study.[10] In particular, the estimation of the state equation in (4.6) is

$$\vartheta_t = -0.2497 + \omega_t^f, \ \ \hat{\sigma}_{\omega^f} = 0.0625. \tag{5.1}$$

In the numerical study, we assume that the insurer sells annuities to 5 Dutch male cohorts, aged from 64 to 68 in the year 2009. The annuities will be in effect in year 2010, i.e., when the cohorts of the annuitants become 65 to 69. Only the cohort aged 64 has the corresponding survivor swaps available, the payments of which also start from year 2010. In other words, $X = \{64, 65, ..., 68\}$ and $X_S = \{64\}$. Without loss of generality, we assume that all $n_i$-s, i.e., the number of annuitant in each cohort, are normalized to 1. The maturity of all annuities and the survivor swaps is 30 years. In other words, no cash flows happen when the oldest cohort reaches age 99. Moreover, the discount rate is assumed to be constant at 4%. The insurer uses historic data from 1980 to 2009 to estimate the Lee-Carter model. Finally, we assume the risk aversion parameter to be 5 for both the Mean-Variance and CVaR specifications, and $M = 50,000$ and $I = 1000$ when calculating the discrete nominal distribution.

## 5.1 The optimal $a$-s

As mentioned before, since the number of publicly traded mortality-linked derivatives is very small, we do not have enough information to price the survivor swaps properly. Nevertheless, the optimal amount of swaps, $a^*$-s, in the insurer's portfolio apparently depends on its price. Therefore, we solve the optimization problems for a range of different risk premiums. Dawson et al. (2010) consider the risk premium of a survivor swap contingent on $x_i = 65$ with maturity 50 years to be around 10% (the annual discount rate is assumed to be flat at 3% and mortality data in England and Wales is used). Since we consider survivor swaps of a shorter maturity and a higher discount rate,[11] we consider risk premiums $\tau \in \{0, 100, 200, 300, 400, 500\}$ basis points. Moreover, we report the results for both cases where the insurer hedges using standardized and customized swaps ($k = 1$ and $k = 2$ in the optimization problems, respectively). It is obvious that when $k = 2$, i.e., there is no basis risk, the estimation of $\delta$ in (4.6) does not play a role. Therefore, we let $d = 2$ in this case.

The result is shown in Figure 1. All plots show the optimal $a$-s for the robust (diamonds) and the nominal (asterisks) optimizations. Left-hand plots are the optimal $\alpha$-s without basis risk. Right-hand plots are with basis

[10] Other datasets can be naturally incorporated to our model. Also, other multi-population mortality model can be used if detailed portfolio specific mortality data is available.

[11] Dowd et al. (2006) show that, as determined by their method, the magnitude of the risk premium decreases as the discount rate increases.

risk. We see that in all cases the optimal amounts of the swaps decrease as the risk premium increases. In particular, the nominal optimal $a$ for CVaR becomes 0 when $\tau \geq 3\%$ and the robust optimal $a$ becomes 0 as $\tau \geq 4\%$, both with and without basis risk. Also, in the presence of basis risk, the optimal $a$ becomes smaller, *ceteris paribus*. These results are intuitive, since the swap becomes less attractive as its price increases, and the hedge effectiveness decreases when basis risk is introduced.

A more interesting observation is that the robust optimal $a$-s are lower than the nominal ones when the risk premium is low, but the nominal optimal $a$-s decrease more rapidly as the risk premium increases. The reason is that the robust optimization always optimize with respect to the worst-cast scenario. We provide the intuition of this observation via a simplified mean-variance specification.

In this simplified example, we assume there is no basis risk. With a slight abuse of notation, write the liability in (2.7) as $\tilde{L} + av(\tau)$, then the optimal $a$ for a nominal mean-variance optimization problem ignoring the constraint $a \in \mathbf{A}$ is

$$a_n^* = -\frac{\hat{v}(\tau)}{2\hat{\sigma}_v^2} - \frac{\hat{\sigma}_{Lv}}{\hat{\sigma}_v^2}, \tag{5.2}$$

and the optimal $a$ for the robust mean-variance optimization is

$$a_r^* = -\frac{v^*(\tau)}{2{\sigma^2}_v^*} - \frac{\sigma_{Lv}^*}{{\sigma^2}_v^*}. \tag{5.3}$$

For both specifications, we have

$$\begin{aligned}
\frac{\partial a}{\partial \sigma_v^2} &= \frac{v(\tau)}{2(\sigma_v^2)^2} + \frac{\sigma_{Lv}}{(\sigma_v^2)^2} \\
\frac{\partial a}{\partial \sigma_{Lv}} &= -\frac{1}{\sigma_v^2} \\
\frac{\partial a}{\partial v(\tau)} &= -\frac{1}{2\sigma_v^2}
\end{aligned} \tag{5.4}$$

In words, the value of the swaps to the insurer depends on three parts: the covariance of the swap and the liability ($\sigma_{Lv}$), the increase of the liability ($v(\tau)$) when buying a swap, and the extra volatility introduced by the swap ($\sigma_v^2$). The objective function is increasing in all three terms. Therefore, in the worst-case scenario, these three terms are larger than their nominal counterparts in the worst-case scenario. When $\tau = 0$, $\hat{v}(\tau) = 0$ and $v^*(\tau)$ is small, so the first equation in (5.4) is likely to be negative or very close to zero when evaluated at $v^*(0)$. Therefore, $a_r^*$ is smaller than $a_n^*$. As $\tau$ increase, both $\hat{v}(\tau)$ and $v^*(\tau)$ increase, so both $a_r^*$ and $a_n^*$ decrease. However, since ${\sigma_v^2}^*$ is bigger than $\hat{\sigma}_v^2$, the increase of $v(\tau)$ has a smaller effect on $a_r^*$. Moreover, the first equation in (5.4) becomes positive as $\tau$ increases, so a bigger $\sigma_v^2$ also increases the optimal $a$. Combining the two effects, the decrease of $a_r^*$

is slower than $a_n^*$. To see this pattern more clearly, we zoom into the range $\tau \in [0, 100]$ basis points and plot the optimal $a$-s for multiple risk premiums in this interval in Figure 2.

### 5.2 Comparison of the robust and nominal optimizations

The comparison between the nominal and the robust optimizations can be interpreted as two insurers aiming to hedge the longevity risk they are exposed to. Having identical liabilities, and estimating the future mortality experience by the Lee-Carter model, the first insurer takes the estimated Lee-Carter model and Equation (5.1) as the real data generating process, while the second insurer recognizes that these estimations may be wrong and optimizes against the possible misspecification. We compare the performance of the two insurers in two settings: with and without basis risk.

First of all, we look how the robust and nominal optimizations perform in a world without model misspecification, i.e., where the estimated parameters are identical to the true values. In this case we compare the values of the nominal objective function evaluated at $a_r^*$ and $a_n^*$, namely

$$\begin{aligned} F_{l,MV}^*(\theta, \tau, k) =& E_\theta[L(y, a_l^*(\tau), k)] + \lambda \text{Var}_\theta[L(y, a_l^*(\tau), k)] \\ F_{l,CVaR}^*(\theta, \tau, k) =& E_\theta[L(y, a_l^*(\tau), k)] + \lambda \text{CVar}_\theta[L(y, a_l^*(\tau), k)], \end{aligned} \tag{5.5}$$

where $l = r, n$. $a_l^*(\tau)$ indicates the dependence of the optimal $a$-s on the risk premium. The results with $k = 2$ (no basis risk) are reported in Table 1.[12] In this case, the nominal optimization performs slightly better than the robust optimization for all risk premiums. This result is not surprising, since when the true world is known without uncertainty, robust optimization will always lead to overconservative results.

| $\tau$ | 0 | 100 | 200 | 300 | 400 | 500 |
|---|---|---|---|---|---|---|
| $F_{r,MV}^*$ | 57.66402 | 58.20186 | 58.69394 | 59.13820 | 59.53151 | 59.88296 |
| $F_{n,MV}^*$ | 57.66400 | 58.19569 | 58.67115 | 59.08530 | 59.44337 | 59.73895 |
| $F_{r,CVaR}^*$ | 346.84 | 350.05 | 353.02 | 355.54 | 353.09 | 353.05 |
| $F_{n,CVaR}^*$ | 346.80 | 350.02 | 352.81 | 353.06 | 353.09 | 353.05 |

Table 1: The optimal values of $F_{l,MV}^*$ and $F_{l,CVaR}^*$ when there is no model misspecification (without basis risk).

Now we turn to the more interesting, and practically relevant, case where model misspecification exists. It is well known that the life expectancy has increased substantially during the past few decades (Plat 2011; van Berkum et al. 2013; Li et al. 2013). The most widely used class of mortality models at the moment is based on a linear extrapolation approach, and thus

[12] Results with $k = 1$ are similar.

the estimation of these models is very sensitive to the calibration windows (Pitacco et al. 2009; Cairns et al. 2006). Moreover, Cairns et al. (2009) and Cairns et al. (2011) compare several popular mortality models, and find that no model performs uniformly better than other models. To sum up, there is no consensus about the real data generating process of the mortality rates. In this study, we evaluate the performance when the real data generating process is not the estimated Lee-Carter model and (5.1), but the same model structures with different parameter values.

We consider a range of different hypothetically true data generating processes, and evaluate the performance of the robust and the nominal optimization under all these distributions. As mentioned above, we take $\theta = (d, \sigma_\omega^2, \delta)$. In particular, the hypothetically true data generating processes for the reference population are assumed to be parameterized by the modified Lee-Carter specification

$$\begin{aligned} \log(\boldsymbol{m}_t) =& \hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\beta}}\kappa_t + \boldsymbol{\varepsilon}_t, \;\; \boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(0, \hat{\Sigma}_\varepsilon) \\ \kappa_t =& d + \kappa_{t-1} + \omega_t, \;\; \omega_t \overset{iid}{\sim} N(0, b\hat{\sigma}_\omega^2), \end{aligned} \tag{5.6}$$

and for the mortality factor model

$$\begin{aligned} \boldsymbol{q}_t^f =& 1 + \boldsymbol{w}\vartheta_t + \varepsilon_t^f, \;\; \boldsymbol{\varepsilon}_t^f \overset{iid}{\sim} N(0, \hat{\Sigma}_{\varepsilon^f}) \\ \vartheta_t =& \delta + \omega_t^f, \;\; \omega_t^f \overset{iid}{\sim} N(0, \hat{\sigma}_{\omega^f}^2). \end{aligned} \tag{5.7}$$

Specifically, we sample $d$ and $\delta$ from their asymptotic distributions, $N(\hat{d}, \hat{\sigma}_d^2)$ and $N(\hat{\delta}, \hat{\sigma}_\delta^2)$, with $\hat{\sigma}_d^2 = \frac{\hat{\sigma}_\omega^2}{30}$ and $\hat{\sigma}_\delta^2 = \frac{\hat{\sigma}_{\omega^f}^2}{30}$. The choice of $\sigma_\omega^2$ follows related findings in literature. Börger et al. (2011) model the mortality development for multiple populations with a stochastic trend model. They find that, in order to generate wide enough forecast confidence intervals to include all the extreme historic mortality developments, they have to blow up volatility of their $\kappa_1$ process by 2. The $\kappa_1$ process in their stochastic trend model is comparable to the $\kappa$ process in the Lee-Carter model. Also, we apply the Bayesian approach from Li et al. (2013) with the Lee-Carter specification to the Dutch males mortality data from 1970 to 2009, and find that the 95% quantile of the posterior distribution of $\sigma_\omega^2$ is around $3.21\hat{\sigma}_\omega^2$. Therefore, we propose 4 possible values of $b$, i.e., $b = \{1, 2, 3, 4\}$.

To sum up, the construction of the hypothetically true distribution is as follows.

1. Let $b = 1$.

2. Construct 200 distributions. For the $k$-th distribution, draw $d^k$ from $N(\hat{d}, \hat{\sigma}_d^2)$ and $\delta^k$ from $N(\hat{\delta}, \hat{\sigma}_\delta^2)$.

3. The $k$-th hypothetical distribution with $b = 1$ is characterized by (5.6) and (5.7) with $b = 1$, $d = d^k$, and $\delta = \delta^k$.

4. Repeat step 2 and 3 for each $b$-s.

Therefore, for each $k$, we have $200 \times 4 = 800$ hypothetical distributions. Moreover, since we have different $a_r^*$ and $a_n^*$-s for each risk premium, in total $6 \times 800 = 4800$ comparisons are made for each $k$. Due to the large number of comparisons, we report our result on a $(\tau, b, k)$ basis. In particular, for each $(\tau, b, k)$, we report the mean and the standard deviation of the 200 optimal values of the nominal objective function.

Figures 3 and 4 report the mean of the optimal values of the Mean-Variance and the CVaR specifications, respectively. The plus signs stand for the robust case and the asterisks for the nominal case. Left-hand plots are without basis risk. Right-hand plots are with basis risks. We see that, both with and without basis risk, the robust optimization yields smaller mean optimal values than the nominal optimization for all specifications except for when $b = 1$ or $\tau = 0$.[13] Moreover, as $b$ or $\tau$ increase, the difference between the robust and nominal optimization also increases. We test the significance of the difference between the robust optimal value and the nominal optimal value for each $(b, \tau, k)$ combination, and find that most differences are significant at 95% level.[14] Furthermore, we see that the mean of the optimal values are higher, *ceteris paribus*, when the basis risk is included, indicating that the inclusion of the basis risk indeed worsens the insurer's situation. However, the differences are not big.

Figure 5 and 6 report the standard deviations of the optimal values. We see that the robust optimization produces smaller standard deviations than the nominal optimization except for some $\tau$-s with $b = 4$ without basis risk and $b = 3, 4$ with basis risk for the Mean-Variance specification.[15] For the CVaR specification, all standard deviations produced by the robust optimization are smaller than the ones produced by the nominal optimization in all cases. Again, when the basis risk is included, the standard deviations of the optimal values increase, *ceteris paribus*.

The result indicates that the robust optimization yields better objective function values in most situations. Although the robust optimization yields overconservative results when $b = 1$ or $\tau = 0$, it yields smaller standard deviations. As discussed above, a data generating process with $b = 1$ is very likely to underestimate the uncertainty in the development of the mortality rates. Also, $\tau = 0$ is a very unrealistic situation. In fact, when the nominal optimization performs better than the robust optimization, the average difference of the mean optimal value is around 0.04 for the Mean-Variance

[13] The difference when $\tau = 0, 100$ is relatively small and not obvious in the Figures. Also, when $\tau = 400, 500$, the robust and nonrobust optimal $a$ are both zero for the CVaR specification, so their performances are the same in this cases.

[14] The exceptions are the mean-variance specification with $(b, \tau, k) = (1, 0, 2)$ and both specifications with $(b, \tau, k) = (1, 0, 1)$.

[15] Also, the difference is relatively small when $\tau = 0$, and the performance of the robust and nonrobust CVaR are the same when $\tau = 400, 500$.

specification and 0.26 for the CVaR specification. However, when the robust optimization performs better, these differences are 1.84 and 2.53, respectively. Moreover, although the optimal values of the objective functions obtained from the nominal Mean-Variance optimization are less sensitive to the change of real data generating process (have smaller standard deviations) in some cases with $b = 3, 4$, they are uniformly dominated by their robust counterparts (have always larger values). Finally, the inclusion of basis risk clearly worsen the wellbeing of the insurer by increasing the means and standard deviations of the optimal values of the objective functions.

# 6 Conclusion

In this article we study the robust longevity risk management problem of an insurer's annuity portfolio. In particular, the insurer does not know the probability law governing the future survival probabilities of her annuitants, and optimizes her portfolio with respect to the worst-case scenario. We consider the mean-variance and the conditional-value-at-risk formulation, and derive the corresponding tractable robust optimization problems. The set of possible probability laws considered by the insurer is statistically given by the Kullback-Leibler divergence.

We apply the robust optimization problem to Dutch male data and compare the performance of the robust optimization with the nominal optimization, in which case the estimated probability law is assumed to be the true law. We construct various realistic settings, and find that the robust optimization yields better results than the nominal optimization in almost all scenarios. Moreover, the degree of the outperformance is higher when the real probability law is further away from the insurer's estimate. Finally, the inclusion of basis risk worsen the wellbeing of the insurer clearly by increasing the means and standard deviations of the optimal values of the objective functions, which indicates that the basis risk is an important risk factor affecting the insurer's longevity risk management. However, the robust optimization still outperforms the nominal ones in the presence of basis risk.

In this study we only consider static robust risk management. It would be interesting to develop dynamic robust risk management strategy in future research.

# Appendices

## A Derivation of the CVaR optimization problem.

In this section we proof that minimizing (3.9) over $a \in \mathbf{A}$ is equivalent to the optimization problem (3.9).

Denote by

$$G_\alpha(a, \xi, \theta, k) = E_\theta[L(y, a, k)] + \lambda F_\alpha(a, \xi, \theta, k). \tag{A.1}$$

According to Theorem 1 in Rockafellar and Uryasev (2000), we have, for any $\theta \in \Theta$,

$$\text{CVaR}_\alpha(a, \theta) = \min_{\xi \in R} F_\alpha(a, \xi, \theta). \tag{A.2}$$

From (A.2), we have

$$\begin{aligned}\text{Mean-CVaR}_\alpha(a, \theta) =& E_\theta[L(y, a, k)] + \lambda \text{CVaR}_\alpha(a, \theta, k)\\ =& E_\theta[L(y, a, k)] + \lambda \min_{\xi \in R} F_\alpha(a, \xi, \theta, k)\\ =& \min_{\xi \in R} G_\alpha(a, \xi, \theta, k).. \end{aligned} \tag{A.3}$$

Therefore, $\min_{\xi \in R} G_\alpha(a, \xi, \theta, k)$ is convex in $a$, and minimizing $G_\alpha(a, \xi, \theta, k)$ with respect to $(a, \xi)$ is equivalent to first minimize with respect to $\xi$ given a fixed $a$ and then minimize with respect to $a$. As a result, we have

$$\begin{aligned}\min_{a \in \mathbf{A}} \text{Mean-CVaR}_\alpha(a, \theta, k) =& \min_{a \in \mathbf{A}} \min_{\xi \in R} G_\alpha(a, \xi, \theta, k)\\ =& \min_{(a,\xi) \in \mathbf{A}(k) \times R} G_\alpha(a, \xi, \theta, k). \end{aligned} \tag{A.4}$$

## B Tractable reformulation of the robust mean-variance problem.

In this section we derive tractable reformulation of the robust optimization problems (3.17) and (3.18).

### B.1 Mean-variance

The problem (3.17) can be reformulated as

$$\begin{aligned}\min_a \max_\pi \quad & d\\ s.t. \quad & \pi' \bar{\boldsymbol{L}}(z, a, k) + \lambda \pi' (\bar{\boldsymbol{L}}(z, a, k) - \pi' \bar{\boldsymbol{L}}(z, a, k))^2 \le d\\ & a \in \mathbf{A}(k)\\ & \pi \in \Pi. \end{aligned} \tag{B.1}$$

Denote by $F(a,\pi,k) = \pi'\bar{\boldsymbol{L}}(z,a,k) + \lambda\pi'(\bar{\boldsymbol{L}}(z,a,k) - \pi'\bar{\boldsymbol{L}}(z,a,k))^2$ and, for any set $U$,

$$\delta(u|U) = \begin{cases} 0 \text{ if } u \in U; \\ \infty \text{ otherwise.} \end{cases}$$

Following Ben-Tal et al. (2012), we have

$$\begin{aligned} \mathcal{F}(a,\pi,k) &\equiv \max_{\pi\in\Pi} F(a,\pi,k) \\ &= \max_{\pi\in R^I}\{F(a,\pi,k) - \delta(\pi|\Pi)\} \\ &= \min_{\nu\in R^I}\{\delta^*(\nu|\Pi) - F_*(a,\nu,k)\}, \end{aligned} \tag{B.2}$$

where $F_*(a,\nu,k) \equiv \inf_{\pi\in R^I}\{\pi'\nu - F(a,\pi,k)\}$ and $\delta^*(\nu|\Pi) \equiv \sup_{\pi\in\Pi}\pi'\nu$ are the conjugate function of $F(a,\pi,k)$ (w.r.t. $a$) and $\delta(\pi|\Pi)$, respectively. Therefore, a $a \in \mathbf{A}(k)$ satisfies problem (B.1) if and only if there is a $\nu \in R^I$ and $a$ satisfying

$$\delta^*(\nu|\Pi) - F_*(a,\nu,k) \le d. \tag{B.3}$$

### B.1.1 Derivation of $\delta^*(\nu|\Pi)$

We have $\delta^*(\nu|\Pi) = \sup_{\pi\in\Pi}\pi'\nu$, i.e.,

$$\begin{aligned} &\max_{\pi} \ \pi'\nu \\ &s.t. \ \pi \in \Pi, \end{aligned} \tag{B.4}$$

since $\Pi$ is compact. We can see that $\pi^0 \in \Pi$, thus $\Pi$ is regular. Therefore, following Ben-Tal et al. (2013), we have $a \in \mathbf{A}(k)$, $v(y,k) \in R^m$ which satisfy (B.3) with uncertainty region $\Pi$ if and only if there exist $a \in \mathbf{A}(Y,K)$, $v(y,k) \in R^m$, $\eta \in R$ and $\xi > 0$ such that

$$\eta\rho + \xi + \eta\sum_{i=1}^{I}\pi_i^0 \exp(\frac{\nu_i - \xi}{\eta} - 1) - F_*(a,\nu,k) \le d. \tag{B.5}$$

### B.1.2 Derivation of $F_*(a,\nu,k)$

To simplify notation, denote by $L \equiv \bar{\boldsymbol{L}}(z,a,k)$ and $L^2 \equiv \bar{\boldsymbol{L}}^2(z,a,k)$, where $\bar{\boldsymbol{L}}^2(z,a,k)$ is the component-wise square of $\bar{\boldsymbol{L}}(z,a,k)$. For any $a \in \mathbf{A}(k)$, $F_*(a,\nu,k)$ can be reformulated as

$$\begin{aligned} &\min_{\pi} \ \pi'\nu - \pi'(L + \lambda L^2 + \lambda\pi' LL'\pi) \\ &s.t. \ \pi \in R^L \end{aligned} \tag{B.6}$$

Write $G(\pi) \equiv \pi'\omega + \lambda\pi' LL'\pi$ with $\omega = \nu - L + \lambda L^2$. Since $\omega$ is a free vector, we can write it as

$$\omega = \mathcal{K}L + c, \tag{B.7}$$

where $\mathcal{K} \in R$ and $c \in N(L) \equiv \{x \in R^I | L'x = 0\}$. It follows that if $c \neq 0$, we can choose $\pi = \delta c$ for some $\delta \in R$ such that

$$\begin{aligned} G(\pi) =& \mathcal{K}\delta c' L + \delta c' c + \lambda \delta^2 c' L L' c \\ =& \delta c' c. \end{aligned}$$

Then $\min_\pi G(\pi) = -\infty$ with $\delta = -\infty$. On the other hand, if $c = 0$, we can choose $\pi = \delta c + d$ with $\delta \in R$ and $d \in N(L)$ such that

$$G(\pi) = \mathcal{K}\delta L + \lambda \delta^2 (L'L)^2, \tag{B.8}$$

which is quadratic function of $\delta$. Minimize (B.8) w.r.t. $\delta$ yields $G^* = -\frac{\mathcal{K}^2}{4\lambda}$ and $\delta^* = -\frac{\mathcal{K}L}{(L'L)^2}$. Therefore, we should that $\omega = \mathcal{K}L$, and thus $v = \lambda L^2 + (\mathcal{K}+1)L$, with $\mathcal{K} \in R$.

### B.2 CVaR

The uncertain vector, $\pi$, is linear in (3.18). Thus, the derivation of the tractable reformulation of (3.18) is a direct application of Theorem 1 in Ben-Tal et al. (2013).

## C Derivation of the worst-case probability

In this section we give an explicit derivation of the solutions to the robust and nominal optimization problems and the characterizations of the worst-case probability vectors for both the Mean-Variance and the CVaR specification. First, recall that the structure of $\mathbf{A}(k)$ and $z$ is given by

$$\mathbf{A} = \{a \in R^m | a_i \geq 0,\ \forall i = \{1, 2, ..., m\} \text{ and } \sum_{x_j \in X_S} a_j \tau_j \sum_{t=1}^{T} \frac{E_{P_{\hat{\theta}}[p(t,x_j,k)]}}{(1+r)^t} \leq d\}$$

and $z = [L(y), v_1(y,k), v_2(y,k), ..., v_m(y,k)]'$. Denote by $c$ a $m \times 1$ vector with the $j$-th entry $\tau_j \sum_{t=1}^{T} \frac{E_{P_{\hat{\theta}}}[p(t,x_j,k)]}{(1+r)^t}$ and $\boldsymbol{v} = [v(y_1)', v(y_2)', ..., v(y_{I-1})', v(y_I)']'$.

Take the CVaR specification for example, the Lagrangian function of the robust problem (4.4) (with subscript $r$) and nominal problem (3.9) (with subscript $n$) is

$$\mathcal{L}_r = \lambda\xi + \rho\zeta + \eta + \zeta \sum_{i=1}^{I} \pi_i^0 \exp\left(\frac{\bar{L}(z_i, a, k) + \frac{\lambda}{1-\alpha} u_i - \eta}{\zeta} - 1\right) + h'(u - \bar{\boldsymbol{L}}(z, a, k) - \xi) + J(d - a'c) \tag{C.1}$$

and

$$\mathcal{L}_n = \pi^{0\prime} \bar{\boldsymbol{L}}(z, a, k) + \lambda(\xi + \frac{1}{1-\alpha} \sum_{i=1}^{I} \pi_i^0 u_i) + h'(u - \bar{\boldsymbol{L}}(z, a, k) - \xi), \tag{C.2}$$

where $J$ and $h$ are the Lagrange multipliers. Denote by $\pi_i = \pi_i^0 \exp(\frac{\bar{L}(z_i,a,k)+\frac{\lambda}{1-\alpha}u_i-\eta}{\zeta} - 1)$ and $\pi = [\pi_1, \pi_2, ..., \pi_I]'$. The first order derivatives of (C.1) and (C.2) can be written as

$$\begin{aligned}
\frac{\partial \mathcal{L}_r}{\partial a} &= (\pi - h)'\boldsymbol{v} - Jc' \\
\frac{\partial \mathcal{L}_r}{\partial \xi} &= \lambda - h'e \\
\frac{\partial \mathcal{L}_r}{\partial u} &= \frac{\lambda}{1-\alpha}\pi + h \\
\frac{\partial \mathcal{L}_r}{\partial \zeta} &= \rho - \sum_{i=1}^{I} \pi_i^0 \exp(\frac{\bar{L}(z_i,a,k) + \frac{\lambda}{1-\alpha}u_i - \eta}{\zeta} - 1)(\frac{\bar{L}(z_i,a,k) + \frac{\lambda}{1-\alpha}u_i - \eta}{\zeta} - 1) \\
&= \rho - \sum_{i=1}^{I} \pi_i \log(\frac{\pi_i}{\pi_i^0}) \\
\frac{\partial \mathcal{L}_r}{\partial \eta} &= 1 - \sum_{i=1}^{I} \pi_i \qquad \text{(C.3)}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \mathcal{L}_n}{\partial a} &= \pi^{0\prime}\boldsymbol{v} - h'\boldsymbol{v} - Jc' \\
\frac{\partial \mathcal{L}_n}{\partial \xi} &= \lambda - h'e \\
\frac{\partial \mathcal{L}_n}{\partial u} &= \frac{\lambda}{1-\alpha}\pi^0 + h, \quad \text{for } i = 1, 2, ..., I, \qquad \text{(C.4)}
\end{aligned}$$

respectively.

Denote by $(a^*, \xi^*, u^*, \zeta^*, \eta^*)$ the solution to the robust CVaR problem and $\pi_i^* = \pi_i^0 \exp(\frac{\bar{L}(z_i,a^*,k)+\frac{\lambda}{1-\alpha}u_i^*-\eta^*}{\zeta^*} - 1)$ for all $i$-s. From (C.3), we see that $\frac{\partial \mathcal{L}_r}{\partial \eta} = 0$ and $\frac{\partial \mathcal{L}_r}{\partial \zeta} = 0$ implies

$$\begin{aligned}
\sum_{i=1}^{I} \pi_i^* &= 1 \\
\sum_{i=1}^{I} \pi_i^* \log(\frac{\pi_i^*}{\pi_i^0}) &= \rho. \qquad \text{(C.5)}
\end{aligned}$$

Also, $\pi^*$ is a nonnegative vector by construction. Thus, $\pi^*$ is a probability vector on the boundary of $\Pi$. Furthermore, comparing the first three equations in (C.3) and (C.4), we see that the first order conditions for $(\alpha, \xi, u)$ are identical in the two optimizations except for the nominal probability, $\pi^0$, is replaced by the worst case probability, $\pi^*$, when the optimal solution is achieved. In this case, $\exp(\frac{\bar{L}(z_i,a^*,k)+\frac{\lambda}{1-\alpha}u_i^*-\eta^*}{\zeta^*} - 1)$ is the worst-case adjustment attached to the $i$-th entry of the nominal probability.

Similar results hold for the mean variance specification. In particular, denote by $\pi_i = \pi_i^0 \exp(\frac{\lambda \bar{L}^2(z_i,a,k)+(k+1)\bar{L}(z_i,a,k)-\xi}{\eta} - 1)$. The worst-case probability vector is $\pi^* = (\pi_1^*, \pi_2^*, ..., \pi_I^*)$ with $\pi_i^* = \pi_i^0 \exp(\frac{\lambda \bar{L}^2(z_i,a^*,k)+(k^*+1)\bar{L}(z_i,a^*,k)-\xi^*}{\eta^*} - 1)$.

# References


Barbarin, J. (2008). Heath–jarrow–morton modelling of longevity bonds and the risk minimization of life insurance portfolios. *Insurance: Mathematics and Economics 43*(1), 41–55.

Ben-Tal, A., D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science 59*(2), 341–357.

Ben-Tal, A., D. Den Hertog, and J. Vial (2012). Deriving robust counterparts of nonlinear uncertain inequalities.

Bor, D. and C. Cowling (2013). The ftse 100 and their pension disclosures. Technical report, JLT Pension Capital Strategies.

Börger, M., D. Fleischer, and N. Kuksin (2011). Modeling mortality trend under modern solvency regimes. *ASTIN Bulletin*, 1–38.

Cairns, A. J. (2013). Robust hedging of longevity risk. *Journal of Risk and Insurance 80*(3), 621–648.

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

Cairns, A. J., D. Blake, and K. Dowd (2008). Modelling and management of mortality risk: a review. *Scandinavian Actuarial Journal 2008*(2-3), 79–113.

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

Cairns, A. J., D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, A. Ong, and I. Balevich (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., K. Dowd, D. Blake, and G. D. Coughlan (2014). Longevity hedge effectiveness: A decomposition. *Quantitative Finance 14*(2), 217–235.

Cairns, A. J. B., D. Blake, K. Dowd, G. D. Coughlan, and M. Khalaf-Allah (2011). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin-Actuarial Studies in non LifeInsurance 41*(1), 29.

Cox, S. H., Y. Lin, R. Tian, and L. F. Zuluaga (2013). Mortality portfolio risk management. *Journal of Risk and Insurance 80*(4), 853–890.

Dahl, M., M. Melchior, and T. Møller (2008). On systematic mortality risk and risk-minimization with survivor swaps. *Scandinavian Actuarial Journal 2008*(2-3), 114–146.

Dawson, P., K. Dowd, A. J. Cairns, and D. Blake (2010). Survivor derivatives: A consistent pricing framework. *Journal of Risk and Insurance 77*(3), 579–596.

Dowd, K., D. Blake, A. Cairns, and G. Coughlan (2011). Hedging pension risks with the age-period-cohort two-population gravity model. In *Seventh International Longevity Risk and Capital Markets Solutions Conference*.

Dowd, K., D. Blake, A. J. Cairns, and P. Dawson (2006). Survivor swaps. *Journal of Risk and Insurance 73*(1), 1–17.

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

Föllmer, H. and A. Schied (2004). Stochastic finance, volume 27 of de gruyter studies in mathematics.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of mathematical economics 18*(2), 141–153.

Gollier, C. (2004). *The economics of risk and time*. MIT press.

Hansen, L. P. and T. J. Sargent (2001). Robust control and model uncertainty. *American Economic Review*, 60–66.

Hansen, L. P. and T. J. Sargent (2008). *Robustness*. Princeton university press.

Laeven, R. J. and M. Stadje (2012). Robust portfolio choice and indifference valuation. *preprint*.

Laeven, R. J. and M. Stadje (2013). Entropy coherent and entropy convex measures of risk. *Mathematics of Operations Research 38*(2), 265–293.

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

Li, H., A. De Waegenaere, and B. Melenberg (2013). The choice of sample size for mortality forecasting: A bayesian learning approach.

Li, J. S.-H. and A. Luo (2012). Key q-duration: A framework for hedging longevity risk. *Astin Bulletin 42*(02), 413–452.

Li, N. and R. Lee (2005). Coherent mortality forecasts for a group of populations: An extension of the lee-carter method. *Demography 42*(3), 575–594.

Liese, F. and I. Vajda (2006). On divergences and informations in statistics and information theory. *Information Theory, IEEE Transactions on 52*(10), 4394–4412.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica 74*(6), 1447–1498.

Mania, M., M. Schweizer, et al. (2005). Dynamic exponential utility indifference valuation. *The Annals of Applied Probability 15*(3), 2113–2143.

Olivieri, A. and E. Pitacco (2009). Stochastic mortality: the impact on target capital. *Astin Bulletin 39*(02), 541–563.

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

Plat, R. (2009). Stochastic portfolio specific mortality and the quantification of mortality basis risk. *Insurance: Mathematics and Economics 45*(1), 123–132.

Plat, R. (2011). One-year value-at-risk for longevity and mortality. *Insurance: Mathematics and Economics 49*(3), 462–470.

Reid, M. D. and R. C. Williamson (2011). Information, divergence and risk for binary experiments. *The Journal of Machine Learning Research 12*, 731–817.

Rockafellar, R. T. and S. Uryasev (2000). Optimization of conditional value-at-risk. *Journal of risk 2*, 21–42.

van Berkum, F., K. Antonio, M. Vellekoop, and K. Leuven (2013). Structural changes in mortality rates.

Zhu, S. and M. Fukushima (2009). Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research 57*(5), 1155–1168.

# Figures

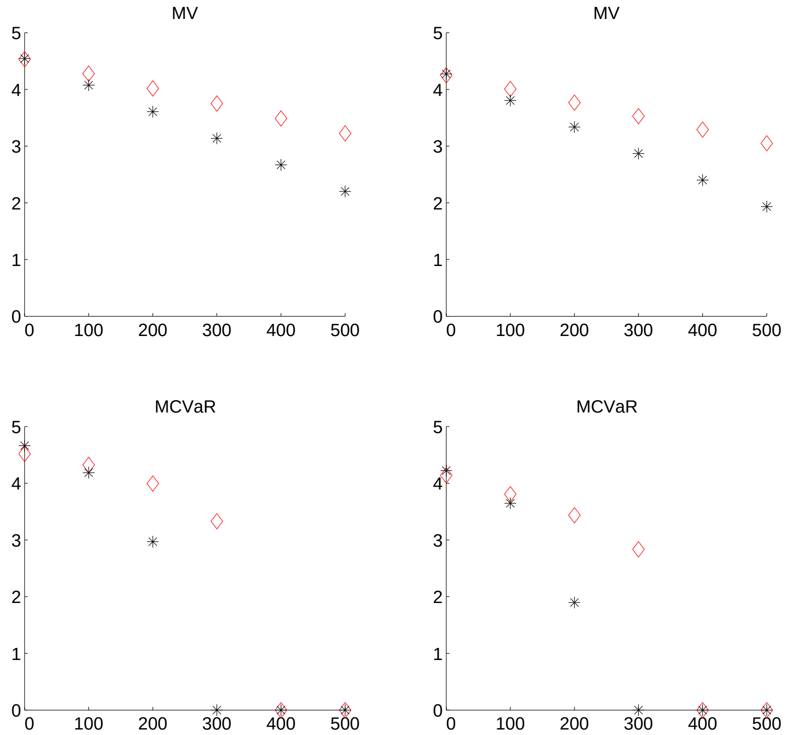


Figure 1: The optimal amount of swaps purchased by the insurer under each risk premium. All plots show the optimal $a$-s for the robust (diamonds) and the nominal (asterisks) optimizations. Left-hand plots are the optimal $\alpha$-s without basis risk. Right-hand plots are with basis risks. The horizontal axis is the risk premium measure in basis points.

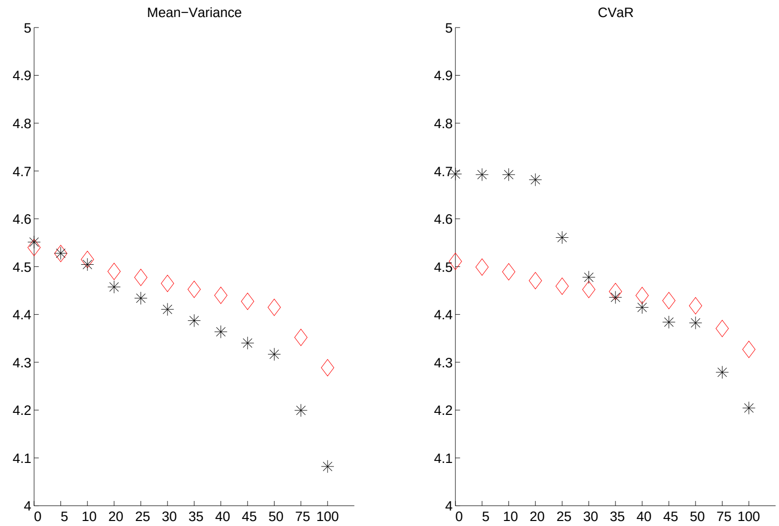


Figure 2: The optimal amount of swaps purchased by the insurer under risk premiums in $[0, 100]$ basis points (without basis risk). The left figure reports the optimal $a$-s for the Mean-Variance specification, and the right figure for the CVaR specification. The horizontal axis is the risk premium measure in basis points.

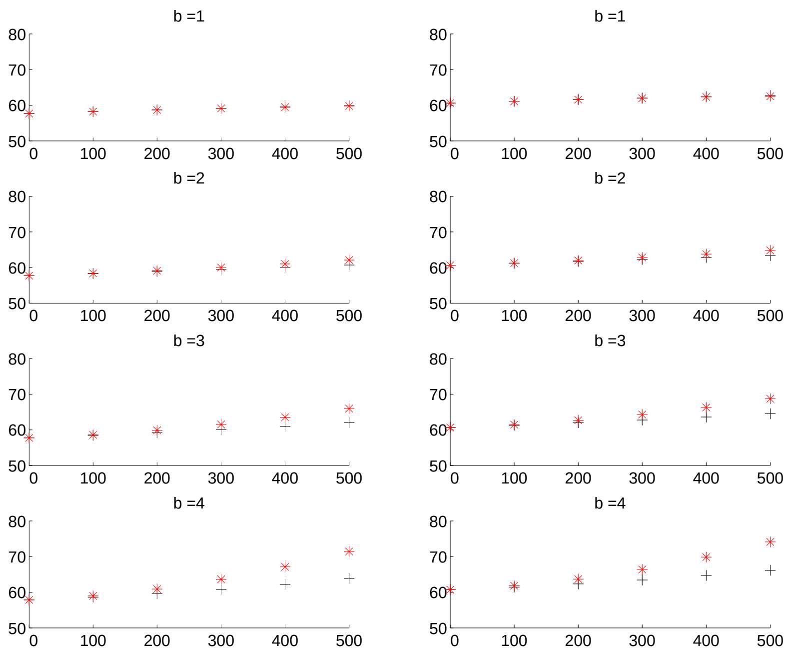


Figure 3: The mean of the optimal values of the objective functions for the Mean-Variance case. The plus signs denote the robust means and the asterisks denote the nominal means. Left-hand plots are without basis risk. Right-hand plots are with basis risks. The horizontal axis represents the risk premiums measured in basis points.

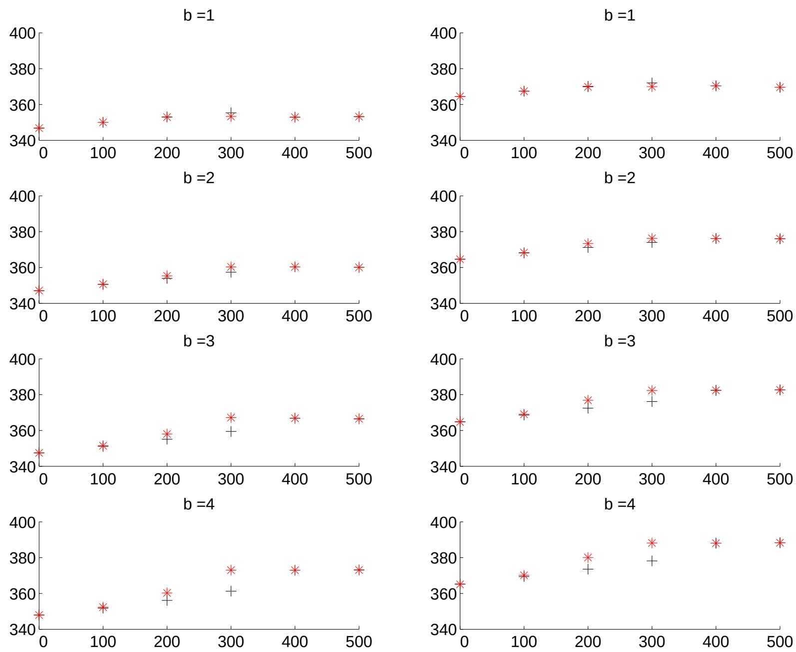


Figure 4: The mean of the optimal values of the objective functions for the CVaR case. The plus signs denote the robust means and the asterisks denote the nominal means. Left-hand plots are without basis risk. Right-hand plots are with basis risks. The horizontal axis represents the risk premiums measured in basis points.

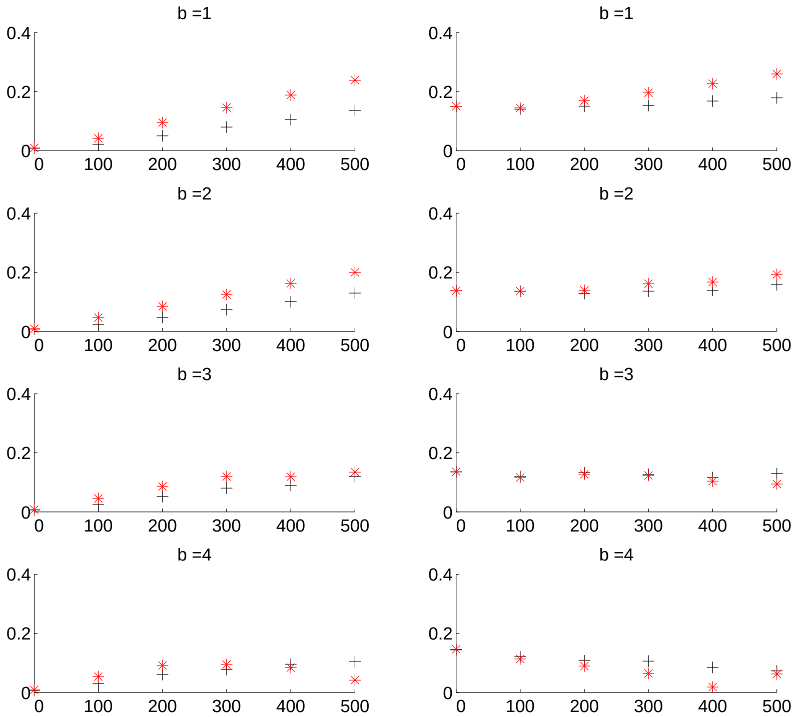


Figure 5: The standard deviation of the optimal values of the objective functions for the Mean-Variance case. The plus signs denote the robust means and the asterisks denote the nominal means. Left-hand plots are without basis risk. Right-hand plots are with basis risks. The horizontal axis represents the risk premiums measured in basis points.

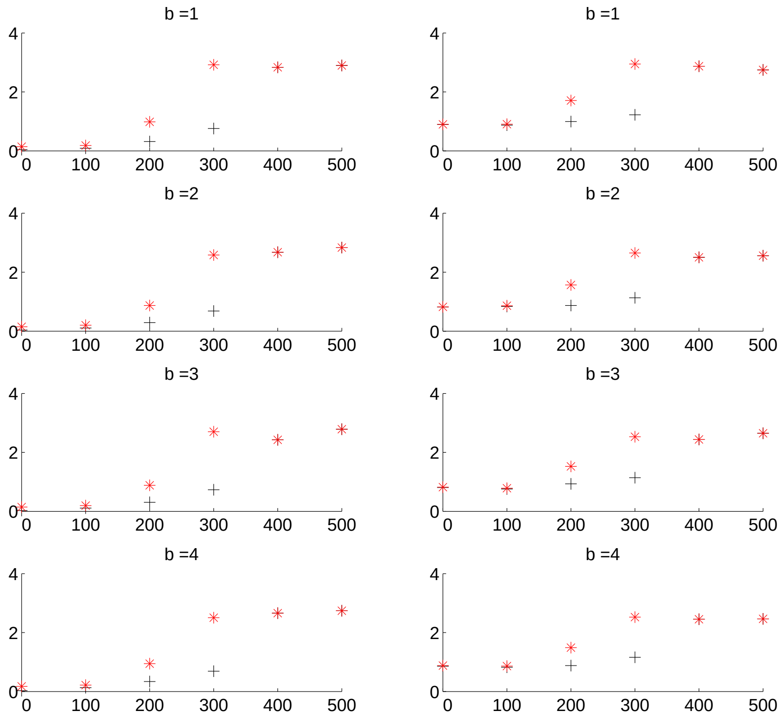


Figure 6: The standard deviation of the optimal values of the objective functions for the CVaR case. The plus signs denote the robust means and the asterisks denote the nominal means. Left-hand plots are without basis risk. Right-hand plots are with basis risks. The horizontal axis represents the risk premiums measured in basis points.