Network for Studies on Pensions, Aging and Retirement

Netspar DISCUSSION PAPERS

Hong Li, Anja De Waegenaere and Bertrand Melenberg

Robust Longevity Risk Management

DP 12/2014-073

Robust longevity risk management

Hong Li[a,1], Anja De Waegenaere[a,b], Bertrand Melenberg[a,b,2]

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

[b]Netspar

December 31, 2014

Abstract

We consider longevity risk hedging problems, where survivor swaps are available as hedging instruments. As objective functions we consider the mean-variance and the mean-conditional-value-at-risk of the hedged liabilities, evaluated using an estimated probability law governing the mortality dynamics. To be robust against estimation inaccuracy, we optimize the worst-case value of the objective function, where the worst-case is with respect to a statistical confidence set around the estimated probability law. We derive reformulations of the worst-case optimization problems that can be solved easily. In the empirical analysis, we compare the performance of the worst case (robust) optimizations with the (non-robust) optimizations that ignore the estimation inaccuracy. We find that the robust optimizations perform better when the actual and estimated probability laws deviate, which is likely to happen in the presence of estimation inaccuracy.

Keywords: longevity risk, robust hedging, Kullback-Leibler divergence, mean-variance, conditional-value-at-risk

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

[2]The authors are grateful to Tortsen Kleinow, Johnny Li, Andrew Hunt, and the seminar participants at the 18th International Congress on Insurance: Mathematics and Economics, Longevity 10, and Nanjing University for their useful comments. All errors are our responsibility. This work is part of the research programme Managing Longevity Risk, which is financed by the Netherlands Organisation for Scientific Research (NWO).

1 Introduction

Pension plans and annuity providers (hereafter referred to as "insurer") are exposed to longevity risk, i.e., the risk due to unanticipated changes in the mortality rates of populations. During recent years, longevity risk has attracted increasing attention from the pension and insurance industry due to demographic and economic changes. As reported in Bor and Cowling (2013), 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. This increase of pension liabilities is partly 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 longevity risk is not taken into account appropriately.

In this paper, we consider the longevity risk management of an insurer who chooses not to buy-out, but to hedge the longevity risk using survivor swaps as hedging instruments. Compared to traditional longevity risk transfer approaches, such as buy-outs, hedging using mortality-linked derivatives[3] is cheaper, and may provide more flexibility to the insurer (Cairns et al. 2014). We assume that the insurer uses as objective function the mean-variance or the mean-conditional-value-at-risk of the hedged liabilities. To be able to evaluate the objective function the insurer uses an estimated probability law governing the mortality dynamics. However, the insurer might be concerned about the possible effects of estimation inaccuracy (due to sampling error or model misspecification), when using an estimated prob-

[3] Mortality-linked derivatives different from survivor swaps, such as longevity bond or q-forward, can alo easily be incorporated in our model.

ability law. For example, Garlappi et al. (2007) show that mean-variance portfolios of stock indexes may be highly sensitive with respect to parameter estimations, which may lower the portfolios Sharpe ratios in applications. Similar results are found by Glasserman and Xu (2013) using commodity futures. Estimation inaccuracy is an important concern when it comes to longevity risk management. In the past decades, the improvement in the life expectancy is accelerating in most of the western countries (van Berkum et al. 2013; Li et al. 2013). As a consequence, the estimations and point forecasts produced by mortality models based on linear extrapolation methods (Cairns et al. 2009; Cairns et al. 2011), probably the most widely used class of mortality models nowadays, may be rather sensitive to the calibration window. For instance, Cairns et al. (2006) obtain substantially different estimation results by fitting the Cairns-Blake-Dowd model to England & Wales males data with two different calibration windows. Therefore, relying only on an estimated mortality model, and hence ignoring the corresponding estimation inaccuracy, might result in a poor performance of one's hedging strategy if the actual probability distribution turns out to deviate from the estimated one.

To be robust against such estimation inaccuracy, we consider optimizing the worst-case value of the objective function, where the worst-case is with respect to a statistical confidence set around the estimated probability law. The confidence set is constructed using the Kullback-Leibler divergence, allowing reformulations of the worst-case optimization problems that can easily be solved, using Ben-Tal et al. (2013) and Ben-Tal et al. (2014). Optimizing the worst-case value of the objective function to deal with estimation inaccuracy is an application of "robust optimization." Robust optimization has been studied extensively in the past two decades, with successful applications to various fields, such as finance, statistics, learning, and engineering (Mulvey et al. 1995; Garlappi et al. 2007; Zhu and Fukushima 2009; Bertsimas et al. 2011). For a detailed introduction of the subject, see Ben-Tal et al. (2009).

We apply our robust optimization problem to Dutch male mortality data and compare the performance of the robust optimizations with their nominal counterparts, when the insurer optimizes her portfolio ignoring estimation inaccuracy. We consider the hedging of longevity risk of a few age classes, when survivor swaps for only one age class are available. In the robust optimization we choose our confidence set to deal with sampling error in particular, ignoring possible model misspecification.[4] We find that the robust optimizations outperform the nominal optimizations when estimation inaccuracies exist, i.e., when the actual probability distribution turns out to deviate from the estimated one.

We also show the impact of population basis risk on the effectiveness of the nominal and robust optimizations. There are currently two major categories of mortality-linked derivatives in the market: the customized contracts and the standardized ones (cf., Cairns et al. 2014). In our analysis we consider hedging using either customized or standardized survivor swaps. By trading customized contracts, the insurer receives payments that match exactly the actual mortality experience in her portfolio. However, such contracts are in general expensive and have very low liquidity. In contrast to customized contracts, the floating payments of standardized contracts are contingent on the mortality experience of a reference population, typically a national population. Standardized contracts are more transparent to investors and offer better liquidity. However, the insurer may not be able to

[4]Also taking into account model misspecification is straightforward, but would require the use of a much larger confidence set.

fully transfer her exposures due to population basis risk, i.e., the mismatch of the mortality experience in the reference population and the population in her portfolios. When using standardized instead of customized survivor swaps in our analysis, the presence of population basis risk indeed leads to worse objective function values for both the robust and the nominal optimizations, but the ordering in terms of performance of the nominal and robust optimizations remains unaffected.

Our paper contributes to the rapidly growing literature focusing on longevity risk management, by explicitly taking estimation inaccuracy in the optimization into account. Among the many existing studies,[5] only few deal with estimation inaccuracy. For example, Cairns (2013) derives robust hedging strategies with respect to the re-calibration of the mortality models at a later stage. Cox et al. (2013) derive an optimal portfolio choice given a fixed mortality law, and evaluate the effectiveness of the optimal portfolio choice when the true mortality law is only known to belong to a set of distributions. In contrast, our paper applies robust optimization to make the longevity risk management robust against estimation inaccuracy.

The remainder of this article is organized as follows. Section 2 describes the setup, the construction of the insurer's liabilities, and the survivor swaps. Section 3 presents the nominal and robust optimization problems. Section 4 reports the application of the robust and nominal optimizations to Dutch male mortality data. Finally we conclude in section 5. The appendix con-

[5]For example, in continuous-time settings, Dahl et al. (2008) and Barbarin (2008) consider quadratic risk minimization problems, and Wong et al. (2014) and Li (2014) consider dynamic minimum-variance and mean-variance hedging of longevity risk. In discrete-time settings, Dowd et al. (2011), Cairns (2013), and Cairns et al. (2014) study static value hedging problems, while Cairns et al. (2008) study a static cash flow hedging problem of longevity risk; Li and Luo (2012) propose a key-q duration concept to hedge longevity risk by trading standardized contracts contingent on a few key cohorts; Cox et al. (2013) propose a "Mean-variance + Conditional-value-at-risk" approach to optimize the mean-variance trade-off while controlling the downside risk.

tains the reformulations of the robust optimization problems that we used to solve these problems.

2 Liabilities and swaps

In this section we specify the cash flow of the insurer's liabilities and the considered survivor swaps. We consider two types of survivor swaps: customized swaps and standardized swaps. The floating legs of the customized swap are contingent on the actual mortality experience of the insurer's portfolio specific population, while the floating legs of the standardized swap are contingent on the mortality experience of the reference population. Let $K = \{rp, pp\}$ be the set of populations, where rp denotes the reference population and pp denotes the insurer's portfolio specific population. Denote by $p(t, x, k)$ the probability that a male aged x in year 0 in population k is alive in year t. We shall assume that $p(t, x, k)$ is observed in year t. However, before year t, $p(t, x, k)$ is an unobserved random quantity.

Suppose that at time 0 an insurer has sold n_x units of an annuity to a group of individuals aged x. Each of the annuities involves a commitment to pay 1 euro every year to the annuitant for the rest of his/her life. Denote by X the set of ages of the cohorts to which the annuities are sold, with $|X| = N$. r is the fixed annual risk free interest rate and T is the last year during which a cash flow occurs. T is typically chosen to be, for example, $120 - \min\{x \in X\}$, i.e., a year after which the number of annuitants that survive is negligibly small, so that we can ignore the annuity payments afterwards.

We consider an insurer with a sufficiently large number of annuitants N, so that the micro-longevity risk (i.e., the risk that the actual survival fractions in population k deviate from the probabilities $p(t, x, k)$) is negligibly small relative to the macro-longevity risk (i.e., the risk that the realizations of $p(t,x,k)$ deviate from their best estimates). Given this assumption, the time 0 (random) discounted liability of the insurer (without hedging) can be written as

$$\tilde{L} = \sum_{x \in X} n_x \sum_{t=1}^{T} \frac{p(t,x,pp)}{(1+r)^t}. \tag{2.1}$$

At time 0, $\tilde{L}$ is random, since all $p(t,x,k)$ for $t \geq 1$ are random.

We consider the case where the insurer opts to hedge the uncertainty of her annuity payments using mortality-linked derivatives. One typical mortality-linked derivative is the survivor swap (Dowd et al. 2006; Dawson et al. 2010). In case of a survivor swap (random) mortality-dependent payments are exchanged for fixed payments. For example, consider a survivor swap contingent on cohort x with maturity T, for $k \in \{rp, pp\}$. The swap is standardized when $k = rp$ and customized when $k = pp$. Denote by $Fix(x,t,k)$ and $Flt(x,t,k)$ the preset and random payment at year t for $t \in \{1,2,...,T-1,T\}$, respectively. At date t the fixed rate payer pays $Fix(x,t,k) - Flt(x,t,k)$ to the counterparty if $Fix(x,t,k) - Flt(x,t,k) > 0$ and vice versa.

In a complete and liquid financial market, one would typically set the fixed rate such that the time 0 market price of the swap contract is equal to zero. However, the mortality-linked capital market is newly developed, not yet complete and possibly not free from arbitrage opportunities. At the current stage, there is no consensus yet how mortality-linked contracts should be priced (Bauer et al. 2010). Following Dowd et al. (2006) we let the fixed and floating payments be given by $Fix(x,t,k) = (1+\tau_x)E_P[p(t,x,k)]$ and $Flt(x,t,k) = p(t,x,k)$, respectively, where P is the physical measure. In other words, at year t, the preset payment is the time 0 best estimated t-year survival probability of the cohort x multiplied by a number $(1+\tau_x)$, while the random payment is the corresponding realized t-year survival probability. For the formulation and derivation of our optimization problem, we will work with a fixed, constant risk premium.[6] The time 0 (random) discounted cash flows received by the fixed rate payer, $S(x,k)$, can now be written as

$$S(x,k)=\sum_{t=1}^{T}\frac{(1+\tau_x)E_P[p(t,x,k)]-p(t,x,k)}{(1+r)^t}. \tag{2.2}$$

In an ideal situation, there are publicly traded survivor swaps for all cohorts $x\in X$. However, at the current stage, derivative products are available only for a few cohorts. The reason is that, since the mortality-linked capital market is newly developed, the liquidity of the derivative products is low and the issuance costs are high. Therefore, issuing only products that are contingent on some key cohorts may help to increase the liquidity of the mortality-linked products and lower the hedging costs of the insurer. To incorporate this fact into our framework, we assume that there is a set $X_S\subset X$ with $|X_S|=m\leq N$ such that tradeable survivor swaps exist only for the cohorts $x\in X_S$.

Aiming to hedge longevity risk, the insurer acts as a fixed rate payer. We denote by a_x the units of swaps for the cohort $x\in X_S$ held by the insurer at time 0 and we denote the hedge portfolio by $\boldsymbol{a}$, the $m\times 1$ vector of a_x-s. The insurer's hedged discounted liabilities are given by

$$\widetilde{L}+\sum_{x\in X_S}a_xS(x,k). \tag{2.3}$$

If $X_S=X$, and $k=pp$, the longevity risk can be fully hedged by choosing

[6] We will examine the impact of the risk premium on the insurer's hedging decision in the numerical study in Section 4.

$a_x = n_x$ for all $x \in X$, since then the hedged discounted liabilities are given by $\sum_{x \in X} n_x E_P(S(x, pp))$. However, in the general case, when $X_S \neq X$ and/or $k \neq pp$, the longevity risk cannot be fully hedged.

We let $\boldsymbol{Z}(k)$ be a $(m+1) \times 1$ random vector with the 1-st entry $\tilde{L}$ and the other entries $S(x, k)$, for $x \in X_S$, with the same ordering as in $\boldsymbol{a}$, cf. (2.3). The hedged liabilities can then be written as

$$L(\boldsymbol{Z}(k), \boldsymbol{a}) = (1\ \boldsymbol{a}')\boldsymbol{Z}(k). \tag{2.4}$$

We assume that the insurer only takes nonnegative positions of the swaps and that the time 0 discounted best estimated amount paid by the insurer does not exceed a constant d. In other words, we have the constraint

$$\boldsymbol{a} \in \mathbf{A}(k) = \{\boldsymbol{a} \in R_+^m | \sum_{x \in X_S} a_x E_P(S(x, k)) \leq d\}, \tag{2.5}$$

where R_+^m denotes the nonnegative orthant of R^m, i.e., the set of $x \in R^m$ with $x_i \geq 0$, $i = 1, 2, ..., m$.

3 Optimal longevity risk hedging

In this section we present the insurer's optimization problems, as seen from time 0, when the insurer aims to choose the hedging portfolio $\boldsymbol{a}$ optimally. At time 0 the hedged liabilities are random, cf. (2.4). As objective function we shall use the mean-variance and the mean-conditional-value-at-risk of the random hedged liabilities. To be able to calculate these objective functions we need the probability distribution $P_{\boldsymbol{Z}(k)}$ of $\boldsymbol{Z}(k)$, which depends on the random survival probabilities $p(t, x, k)$. We assume that the insurer does not know $P_{\boldsymbol{Z}(k)}$, but has to estimate it, using a model like the Lee and

Carter (1992)-model. We first present the nominal optimization problems, obtained when the insurer uses $\widehat{P}_{\boldsymbol{Z}(k)}$, the best estimate of $P_{\boldsymbol{Z}(k)}$, without taking into account estimation inaccuracy. Next, we present the robust counterparts to deal with estimation inaccuracy, which are obtained when the insurer optimizes the worst-case objective functions, where the worst-case is with respect to $P_{\boldsymbol{Z}(k)}$ in a confidence set around $\widehat{P}_{\boldsymbol{Z}(k)}$. We present the optimization problems for fixed $k \in \{rp, pp\}$. Therefore, we suppress the dependence on k, i.e., we write $\boldsymbol{Z} = \boldsymbol{Z}(k)$ and $\boldsymbol{A} = \boldsymbol{A}(k)$. Moreover, we simplify notation by writing $P = P_{\boldsymbol{Z}}$ and $\widehat{P} = \widehat{P}_{\boldsymbol{Z}}$.

3.1 Nominal optimization problems

In this subsection we present the nominal optimization problems, which are obtained when the insurer uses $\widehat{P}$, the best estimate probability distribution P of $\boldsymbol{Z}$, ignoring possible estimation inaccuracy. We consider two objective functions, namely mean-variance and mean-CVaR, where CVaR stands for the conditional value-at-risk. The mean-variance optimization problem, when the insurer uses $\widehat{P}$, can be written as

$$\min_{\boldsymbol{a}\in\mathbf{A}}\{E_{\widehat{P}}(L(\boldsymbol{Z},\boldsymbol{a})) + \lambda Var_{\widehat{P}}(L(\boldsymbol{Z},\boldsymbol{a}))\}, \tag{3.1}$$

were λ quantifies the trade-off between the expected value of the liabilities and the risk of the liabilities, where the risk is quantified by the variance.[7]

To present the mean-CVaR optimization problem, we first introduce CVaR. Given a confidence level α with $0 < \alpha < 1$ and a fixed P, the Value-

[7]In case of a risk neutral insurer, i.e., one who is only concerned about the expected value of her liabilities (i.e., $\lambda = 0$), the optimal solution is $a^* = 0$ if $\tau_x > 0$ for all x, since a long position of any swap would increase the insurer's expected liabilities. Thus, a risk neutral insurer would not hedge the longevity risk, even if she could, when there is a positive cost to do so, at least, when the insurer ignores estimation inaccuracy.

at-Risk (VaR) of $L(\boldsymbol{Z}, \boldsymbol{a})$ is defined as

$$\mathrm{VaR}_{\alpha,P}(L(\boldsymbol{Z}, \boldsymbol{a})) = \min\{d \in R | \int\limits_{L(\boldsymbol{z},\boldsymbol{a}) \leq d} L(\boldsymbol{z}, \boldsymbol{a}) dP(\boldsymbol{z}) \geq \alpha\}. \tag{3.2}$$

For the same confidence level α and the same P, CVaR is then defined as

$$\mathrm{CVaR}_{\alpha,P}(L(\boldsymbol{Z}, \boldsymbol{a})) = \frac{1}{1-\alpha} \int\limits_{L(\boldsymbol{z},\boldsymbol{a}) \geq \mathrm{VaR}_{\alpha,P}(L(\boldsymbol{Z},\boldsymbol{a}))} L(\boldsymbol{z}, \boldsymbol{a}) dP(\boldsymbol{z}). \tag{3.3}$$

The mean-CVaR optimization problem, when the insurer uses $\widehat{P}$, can then be written as

$$\min_{\boldsymbol{a} \in \mathbf{A}} \{E_{\widehat{P}}(L(\boldsymbol{Z}, \boldsymbol{a})) + \lambda \mathrm{CVaR}_{\alpha,\widehat{P}}(L(\boldsymbol{Z}, \boldsymbol{a}))\}, \tag{3.4}$$

were λ again quantifies the trade-off between the expected value of the liabilities and its risk, now quantified by the conditional value-at-risk.

In the sequel, we shall solve the optimization problems (3.1) and (3.4) under the assumption that $\boldsymbol{Z}$ has a discrete distribution, represented by the I-dimensional probability vector $\boldsymbol{\pi}$, with components π_i, $i = 1, \cdots, I$, and outcome space $\{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_I\}$, where $P(\boldsymbol{Z} = \boldsymbol{z}_i) = \pi_i$, $i = 1, 2, ..., I$. We denote by $\hat{\boldsymbol{\pi}}$ the best estimate of $\boldsymbol{\pi}$, so that $\widehat{P}(\boldsymbol{Z} = \boldsymbol{z}_i) = \hat{\pi}_i$, $i = 1, 2, ..., I$. In subsection 4.1 we show how $\hat{\boldsymbol{\pi}}$ can be determined using Lee and Carter (1992) via simulation and discretization.

3.2 Robust Counterparts

Next, we consider the optimization problems when estimation inaccuracy enters the decision-making problem. In this case, the insurer recognizes that her best estimates may be subject to estimation inaccuracy (sampling error or model misspecification), and considers a (compact) confidence set of probability distributions, $\hat{\Pi}$, around $\widehat{P}$. We assume that the insurer now optimizes the worst-case objective function, where the worst-case is with respect to $P \in \hat{\Pi}$. The resulting optimization problems are referred to as the robust counterparts of the original nominal optimization problems, cf. Ben-Tal et al. (2009). Robust counterparts are attractive alternatives to the nominal optimization problems, when there is uncertainty about the parameter(s) determining the objective function and/or constraints (in our case the parameter is the probability distribution): often, the robust counterpart performs better than the original nominal optimization problem in terms of the objective function, when this objective function is evaluated using a parameter (in our case a probability distribution) deviating from the nominal parameter (in our case the estimated probability distribution).[8]

The robust counterpart of (3.1) becomes

$$\min_{\boldsymbol{a}\in\mathbf{A}} \max_{P\in\hat{\Pi}} \{E_P(L(\boldsymbol{Z},\boldsymbol{a})) + \lambda Var_P(L(\boldsymbol{Z},\boldsymbol{a}))\}, \tag{3.5}$$

and the robust counterpart of (3.4) is

$$\min_{\boldsymbol{a}\in\mathbf{A}} \max_{P\in\hat{\Pi}} \{E_P(L(\boldsymbol{Z},\boldsymbol{a})) + \lambda \mathrm{CVaR}_{\alpha,P}(L(\boldsymbol{Z},\boldsymbol{a}))\}. \tag{3.6}$$

Before we can solve the robust counterparts, the structure of the uncertainty set, $\hat{\Pi}$, has to be specified. There are many popular structures of the uncertainty set in the literature.[9] We choose our uncertainty set to be

[8]See Ben-Tal et al. (2009). Of course, without estimation inaccuracy the nominal optimization problem will perform better.

[9]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 characterized ϕ divergence; Laeven and Stadje (2012 and 2013) study risk measures and dynamic portfolio choice problems

characterized by the Kullback-Leibler divergence (also known as relative entropy). This divergence is used 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 2012, 2013), macroeconomics (Hansen and Sargent 2001, 2008), and decision makings (Gollier 2004; Ben-Tal et al. 2013). Like in the nominal optimization problems, we shall solve the robust counterparts (3.5) and (3.6) under the assumption that $\boldsymbol{Z}$ has a discrete distribution, so that a probability distribution P is fully represented by the I-dimensional probability vector $\boldsymbol{\pi}$, with $P(\boldsymbol{Z} = \boldsymbol{z}_i) = \pi_i$, with as best estimates $\hat{\boldsymbol{\pi}}$, $i = 1, 2, ..., I$. In our context, the uncertainty set characterized by the Kullback-Leibler divergence can then be written as

$$\hat{\Pi} = \{\boldsymbol{\pi} \in R_+^I | \sum_{i=1}^{I} \pi_i = 1, \sum_{i=1}^{I} \pi_i \log(\frac{\pi_i}{\hat{\pi}_i}) \leq \rho\}, \tag{3.7}$$

for some $\rho > 0$. From (3.7), we see that the degree of divergence between the candidate probability distributions and the nominal one is determined by a single parameter, ρ. Hence, ρ can be interpreted as the degree of ambiguity aversion of the decision maker.

To complete the robust counterpart, we need to choose a specific value for ρ. We consider the case where the distribution P of $\boldsymbol{Z}$ belongs to a parametrized set of probability distributions, such as induced by the Lee and Carter (1992)-model. We choose ρ such that $\hat{\Pi}$ in (3.7) becomes an (approximate) confidence set around $\hat{\boldsymbol{\pi}}$ of at least level $(1-\beta)$. Let $\{P_\theta | \theta \in \Theta \subset R^e\}$ denote this parametrized set of probability distributions, with θ estimated by the Maximum Likelihood estimate $\hat{\theta}$. We only consider estimation inaccuracy, due to sampling error in $\hat{\theta}$. According to Ben-Tal et al. (2013) the

using uncertainty sets defined by relative entropy.

appropriate choice for ρ is then given by

$$\rho = \frac{\chi^2_{e,1-\beta}}{2N}, \tag{3.8}$$

where $\chi^2_{e,1-\beta}$ is the $1-\beta$ percent critical value of a χ^2 distribution with degrees of freedom e, and N is the sample size (used to estimate θ).[10] The reason is that in this case $\hat{\Pi}$, defined in (3.7), includes as subset all P_θ whose Kullback-Leibler divergence with respect to $P_{\hat{\theta}}$ is less than or equal to ρ. This set of P_θ-s turns out to be an approximate $(1-\beta)$-confidence set of P_θ. Since $\hat{\Pi}$ includes this approximate $(1-\beta)$-confidence set of P_θ, it is an (approximate) confidence set of P (represented by the probability vector $\boldsymbol{\pi}$) of *at least* confidence $(1-\beta)$. For a more detailed motivation of this choice we refer to Ben-Tal et al. (2013).

After specifying $\hat{\Pi}$ and ρ, we follow the method proposed in Ben-Tal et al. (2013) and Ben-Tal et al. (2014) to derive reformulations of the robust optimization problems (3.5) and (3.6) that allow an easy way to solve the optimization problems. These reformulations are presented in the Appendix.

4 Performance of the nominal optimization problems and their robust counterparts

In this section we consider an insurer whose portfolio includes annuitants of different age classes and who chooses a hedge portfolio of survivor swaps to hedge the corresponding longevity risk, using one of the optimization problems described in the previous section. We assume that the insurer models

[10] Also taking into account model misspecification would require a larger choice of ρ, such as $\rho = \frac{\chi^2_{I-1,1-\beta}}{2N}$, see Ben-Tal et al. (2013).

the mortality rates in the reference population by the Lee-Carter model (Lee and Carter 1992) and the mortality rates in the insurer's portfolio by the method proposed in Plat (2009).[11] The insurer uses mortality data to estimate the distribution of the mortality law and to construct the corresponding uncertainty set. In this section we evaluate the performance of both the nominal optimization problems, presented in the previous section, and their robust counterparts. We first specify the (empirical) nominal distribution of the mortality law and the uncertainty set determining the robust counterpart. Next, we specify the insurer's portfolio. Finally, we present the performance evaluation of the nominal and robust hedging strategies.

4.1 The nominal distribution and the uncertainty set

In this subsection we present the distribution of $\boldsymbol{Z}$, see (2.4). For given $k \in \{pp, rp\}$, $\boldsymbol{Z}$ depends on $p(t, x, k)$ for $t = 1, \cdots, T$ and $x \in X$. We first describe the distribution of $p(t, x, k)$ for $k = rp$ and $k = pp$, respectively. Then we describe the distribution of $\boldsymbol{Z}$, i.e., we describe the construction of $P(\boldsymbol{Z} = \boldsymbol{z}_i) = \pi_i$, $i = 1, 2, ..., I$. Finally, we present the uncertainty set $\hat{\Pi}$, see (3.7).

The reference population—First, we model the mortality process of the reference population, i.e., $p(t, x, rp)$ for $t = 1, \cdots, T$ and $x \in X$. Denote by $m(t, x, rp)$ the one-year crude death rate[12] in year t applying to the cohort whose age is x in year 0 in the reference population rp. The t-year survival

[11] We make these choices for illustrative purposes only. Other models can easily be incorporated in our approach as well. 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 modeling.

[12] Number of deaths over the corresponding exposure.

probabilities can be approximated by (Pitacco et al. 2009)

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

Let $\boldsymbol{y}_t$ bet the N-dimensional column vector with as components $\log(m(t, x, rp))$, $x \in X$. The Lee-Carter model (Lee and Carter 1992) models this vector $\boldsymbol{y}_t$ as[13]

$$\begin{aligned} \boldsymbol{y}_t =& \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\varepsilon}_t, \;\; \boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(0, \Sigma_\varepsilon) \\ \kappa_t =& d + \kappa_{t-1} + \omega_t, \;\; \omega_t \overset{iid}{\sim} N(0, \sigma_\omega^2), \end{aligned} \tag{4.1}$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ are vectors of parameters, $\boldsymbol{\varepsilon}_t$ and ω_t are mutually independent i.i.d shocks ($\boldsymbol{\varepsilon}_t$ N-dimensional and ω_t one-dimensional), assumed to be normally distributed, and Σ_ε and σ_ω^2 are the covariance matrix of $\boldsymbol{\varepsilon}_t$ and the variance of ω_t, respectively. The κ process captures the common time varying trend of the central death rates, and it is modeled by a random walk process with drift term d and volatility σ_ω. Moreover, we assume Σ_ε to be a diagonal matrix.

The portfolio specific population—Next, we model the mortality process of the portfolio specific population, i.e., $p(t, x, pp)$ for $t = 1, \cdots, T$ and $x \in X$. First, denote by $p(t, t-1, x, k)$ the probability that an individual aged x in year 0 in population k survives in year t, given that he/she is alive in year $t - 1$.[14] The $(t-1)$-conditional one year death probability can then be written as $q(t, t-1, x, k) = 1 - p(t, t-1, x, k)$.

[13] In general the κ process appearing in (4.1) 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.

[14] Thus, $p(t, t-1, x, k)$ is the $(t-1)$-conditional one year survival probability, while $p(t, x, k)$ is the (0-conditional) t year survival probability.

Denote by $f(t,t-1,x) \equiv \frac{q(t,t-1,x,pp)}{q(t,t-1,x,rp)}$ the mortality factor of the reference population over the portfolio population, given x and t. Let $\boldsymbol{f}_t$ be the N-dimensional column vector with as components $f(t,t-1,x)$, $x \in X$. Plat (2009) fits the mortality factor by a one factor linear model[15]

$$
\begin{aligned}
\boldsymbol{f}_t =& \iota + \boldsymbol{w}\vartheta_t + \boldsymbol{\varepsilon}_t^f \\
\vartheta_t =& \delta + \omega_t^f,
\end{aligned}
\tag{4.2}
$$

where ι is a $N \times 1$ vector of ones, $\boldsymbol{w}$ is a $N \times 1$ weighting vector with as entries $w_x = 1 - \frac{x-\underline{x}}{\bar{x}-\underline{x}}$, $x \in X$, and $\bar{x}$ and $\underline{x}$ are the oldest and youngest cohort in X, respectively, and $(\vartheta_t)_t$ is modeled as a ARIMA(0,0,0) process with mean δ. Moreover, $\boldsymbol{\varepsilon}_t^f \overset{iid}{\sim} N(0, \Sigma_{\varepsilon^f})$ and $\omega_t^f \overset{iid}{\sim} N(0, \sigma^2_{\omega^f})$. We call (4.2) the mortality factor model. According to (4.2), we have the following relation between the reference population and the portfolio's population

$$
p(t,x,pp) = \prod_{s=1}^{t} [1 - (1 - p(s,s-1,x,rp)) f(s,s-1,x)]. \tag{4.3}
$$

Data—To fit the mortality process of the reference population by the Lee-Carter model, we use Dutch male mortality data from 1980 to 2009.[16] For the mortality process of the portfolio population, we make use of Plat (2009), who fits model (4.2) to Dutch male mortality data of a collective pension portfolio of Dutch insurers containing about 100,000 male policyholders aged 65 or older (and using the Dutch males population as reference populaton).

[15] Similar to the Lee-Carter model, (4.2) can be generalized to a multi-factor model, with the state variable ϑ_t following a general $ARIMA(p,d,q)$ process. However, the author fits the form (4.2) to the data and we adopt the same form in this paper.

[16] The data is downloaded from the Human Mortality Database (http://www.mortality.org/).

In particular, the estimation of the state equation in (4.2) is

$$\vartheta_t = -0.2497 + \omega_t^f, \quad \hat{\sigma}_{\omega^f} = 0.0625. \tag{4.4}$$

The nominal distribution—So far, we have specified the estimated distribution of $p(t, x, k)$, for $t \geq 1$, $x \in X$, and for given $k \in \{pp, rp\}$. In the optimization problems we use as nominal distribution the estimated induced distribution of $\boldsymbol{Z} = \boldsymbol{Z}(k)$, for given $k \in \{pp, rp\}$. We obtain this nominal distribution, $\widehat{P}$, with $\widehat{P}(\boldsymbol{Z} = \boldsymbol{z}_i) = \hat{\pi}_i$, $i = 1, \cdots, I$, as follows. First, we simulate M realizations of $p(t, x, k)$, for $t \geq 1$, $x \in X$, and for $k \in \{pp, rp\}$, using the two estimated models (4.1) and (4.2). Specifically, we first simulate the log of central death rates of the reference population, and transfer them to the corresponding t-year survival probabilities. The t-year survival probabilities of the portfolio population can be obtained using (4.3). Based on these realizations of $p(t, x, k)$, we construct the corresponding realizations of $\boldsymbol{Z}$.

Next, we divide the range of the M realizations of $\boldsymbol{Z}$ into I subsets. The values of $\boldsymbol{z}_i$ are obtained by taking the average of all realizations of $\boldsymbol{Z}$ that fall into the corresponding subset i and the values of $\hat{\pi}_i$ are the frequencies of the realizations of $\boldsymbol{Z}$ falling into the corresponding subset i. We use $M = 50,000$ and $I = 1,000$.

Since the nominal distribution $\widehat{P}$ is derived from the two estimated models (4.1) and (4.2), we actually have $\widehat{P} = P_{\hat{\theta}}$, with $\hat{\theta}$ the estimated parameters of the two models.

The uncertainty set—Our uncertainty set is given by (3.7), with ρ given by (3.8), with e equal to the dimension of θ, the vector of parameters appearing in the two models (4.1) and (4.2). In principle, all these parameters are subject to estimation errors. However, as stated in Cairns (2013), most uncertainty of the (single population) mortality forecasting comes from the estimation of the drift term in the κ process. Also, Lee and Carter (1992) only take the uncertainty from the κ process into account when calculating the confidence intervals of the forecast mortality rates. Therefore, we proceed as if only d and σ_ω in (4.1), and only δ in (4.2) are estimated with possible estimation inaccuracy. Thus, we proceed as if $\theta = (d, \sigma_\omega^2, \delta)'$, where θ is assumed to be estimated by an appropriate estimator, so that $e = 3$ in (3.8), and $\hat{\Pi}$, given by (3.7), follows given this value of ρ. When there is no population basis risk, the estimation of δ in (4.2) does not play a role, so that we have $\theta = (d, \sigma_\omega^2)'$, with $e = 2$, in this case.

4.2 Insurer's portfolio

We assume that the insurer's portfolio consists of five Dutch male cohorts aged 64, 65, 66, 67, and 68 in 2009. The annuities start paying out in 2010, i.e., when the cohorts of the annuitants become 65 to 69, respectively. Only the cohort aged 64 has a corresponding survivor swap available in the market, the payments of which start also from 2010. In other words, $X = \{64, 65, ..., 68\}$ and $X_S = \{64\}$. We write $\tau = \tau_{64}$. We normalize all n_x-s, i.e., the number of annuitant in each cohort, to $n_x = 1$. The maturity of all annuities and of the survivor swap is 30 years. In other words, we assume that no cash flows happen after the oldest cohort reaches age 99. The risk free interest rate is assumed to be constant at $r = 4\%$. We set the risk aversion parameter equal to 5, i.e., $\lambda = 5$ for both the Mean-Variance and CVaR specifications. Finally, the maximum amount that the hedger wishes to pay for hedging, i.e., d in (2.5), is assumed to be $E_{\widehat{P}}(\tilde{L})$.

As mentioned before, there is not yet a standard practice of pricing survivor swaps at the current stage. However, the optimal amount of swaps purchased by the insurer apparently depends on its price. To illustrate this effect, 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 $x = 65$ 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,[17] we consider risk premiums $\tau \in \{0, 100, 200, 300, 400, 500\}$ basis points.

4.3 Comparison of the nominal and robust optimizations

In this subsection we compare the nominal and robust optimal hedging strategies. First, we compare these hedging strategies, considering both the use of standardized and customized swaps. Next, we illustrate the performance of these hedging strategies, when the actual probability distribution P deviates from the nominal, estimated $\widehat{P}$, which is likely to happen when estimation inaccuracy plays a role.

Figure 1 shows the optimal hedging strategies (a) for the robust (diamonds) and the nominal (asterisks) optimizations. The left panels display the optimal a-s without basis risk, and the right panels display the results with basis risk. In all cases the optimal amounts of the swaps decrease as the risk premium increases. In particular, for CVaR the nominal optimal a 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, holding other factors equal. These results are intuitive, since the swap becomes less attractive as its price increases, and

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

the hedge effectiveness decreases when basis risk is introduced.

Next, we evaluate the performance of the robust and the nominal optimization when the true underlying probability distribution P differs from the best-estimate distribution $\widehat{P} = P_{\hat{\theta}}$, with $\hat{\theta} = (\widehat{d}, \widehat{\sigma}^2_\omega, \widehat{\delta})$. To do so, we consider a range of different hypothetical "true" probability distributions $P = P_\theta$, allowing for $\theta \neq \hat{\theta}$. For each of these hypothetical true distributions, we evaluate the performance of the robust ($\boldsymbol{a}^*_r$) and the nominal ($\boldsymbol{a}^*_n$) optimal values of $\boldsymbol{a}$ by determining the value of the mean-variance and the mean-CVaR objective functions, i.e., we calculate with respect to the true hypothetical distribution P

$$E_P(L(\boldsymbol{Z}, \boldsymbol{a})) + \lambda Var_P(L(\boldsymbol{Z}, \boldsymbol{a})), \tag{4.5}$$

in the mean-variabce case and

$$E_P(L(\boldsymbol{Z}, \boldsymbol{a})) + \lambda \mathrm{CVaR}_{\alpha,P}(L(\boldsymbol{Z}, \boldsymbol{a})). \tag{4.6}$$

in the mean-CVaR case, for $\boldsymbol{a} = \boldsymbol{a}^*_r$ and $\boldsymbol{a} = \boldsymbol{a}^*_n$, with $P = P_\theta$.

To generate the hypothetical true distributions for the reference population and for the portfolio-specific population, we let the drift terms in (4.1) and in (4.2) differ from their best-estimate values $\widehat{d}$ and $\widehat{\delta}$. Moreover, we inflate the variance of ω_t in (4.1) by a factor b. Specifically, the hypothetical true distributions P_θ are generated from

$$\begin{aligned} \log(\boldsymbol{m}_t) =& \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\varepsilon}_t, \;\; \boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(0, \Sigma_\varepsilon) \\ \kappa_t =& (\hat{d} + \varepsilon_d) + \kappa_{t-1} + \omega_t, \;\; \omega_t \overset{iid}{\sim} N(0, b\widehat{\sigma}^2_\omega), \end{aligned} \tag{4.7}$$

and

$$\begin{aligned}\boldsymbol{f}_t =& 1 + \boldsymbol{w}\vartheta_t + \varepsilon_t^f, \quad \boldsymbol{\varepsilon}_t^f \overset{iid}{\sim} N(0, \Sigma_{\varepsilon^f})\\ \vartheta_t =& (\hat{\delta} + \varepsilon_\delta) + \omega_t^f, \quad \omega_t^f \overset{iid}{\sim} N(0, \sigma^2_{\omega^f}).\end{aligned} \tag{4.8}$$

with $\varepsilon_d \sim N(0, \sigma_d^2)$, $\varepsilon_\delta \sim N(0, \sigma_\delta^2)$, and $b \in \{1, 2, 3, 4\}$. For σ_d^2 and σ_δ^2 we use the (estimated asymptotic) variances of $\widehat{d}$ and $\widehat{\delta}$, respectively.[18] The choice of b-values is based on the Bayesian approach from Li et al. (2013) (with the Lee-Carter specification applied to the Dutch males mortality data from 1970 to 2009), who find that the 95% quantile of the posterior distribution of σ_ω^2 is around $3.21\hat{\sigma}_\omega^2$.[19]

For each $b \in \{1, 2, 3, 4\}$, we construct 200 hypothetical distributions by taking 200 random draws ε_d^h from $N(0, \sigma_d^2)$ and ε_δ^h from $N(0, \sigma_\delta^2)$. This yields $200 \times 4 = 800$ hypothetical distributions for each population $k \in \{pp, rp\}$. For each of these 800 distributions, we determine the value of the mean-variance and the mean-CVaR objective functions in (4.5) and (4.6) for the optimal robust and non-robust number of swaps, using $P = P_\theta$ as true distribution. Because we have six risk premiums, we repeat this procedure for each of these values, so that in total $6 \times 800 = 4800$ comparisons are made for each k. Due to the large number of comparisons, we report our results on a (τ, b, k) basis. In particular, for each (τ, b, k), we report the mean and the standard deviation of the 200 optimal values of the objective function.

Figures 2 and 3 report the mean of the optimal values of the mean-

[18] These are $\widehat{\sigma}_d^2 = \frac{\widehat{\sigma}_\omega^2}{30}$ and $\widehat{\sigma}_\delta^2 = \frac{\widehat{\sigma}^2_{\omega^f}}{30}$.

[19] Moreover, Börger et al. (2011) model the mortality development for multiple populations with a stochastic trend model and find that, in order to generate wide enough forecast confidence intervals to include all the extreme historic mortality developments, they have to blow up the volatility of their κ_1 process by 2. This κ_1 process in their stochastic trend model is comparable to the κ process in our Lee-Carter model.

variance and the mean-CVaR specifications, respectively, and Figure 4 and 5 show the corresponding standard deviations of the optimal values. The plus signs stand for the robust case and the asterisks for the nominal case. The left panels display results without basis risk ($k = pp$), and the right panels display results with basis risk ($k = rp$). We see that, both with and without population basis risk, the robust optimization yields smaller mean optimal values than the nominal optimization for all specifications, except when $b = 1$ or $\tau = 0$.[20] In addition, we see that the robust optimization produces smaller standard deviations than the nominal optimization, except for some τ-s with $b = 4$ without basis risk, and $b = 3$ and 4 with population basis risk for the mean-variance specification.[21] For the mean-CVaR specification, all standard deviations produced by the robust optimization are smaller than the ones produced by the nominal optimization. Moreover, as b or τ increases, the difference between the robust and nominal mean optimal objective values also increases. We test the significance of this difference for each (b, τ, k) combination, and find that most differences are significant at the 95% level.[22] Furthermore, we see that the means, as well as the standard deviations, of the optimal values for each (b, τ) combination increase when population basis risk is included, indicating that the inclusion of this risk indeed worsens the insurer's situation.

The results indicate 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

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

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

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

deviations. 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 specification and 0.26 for the mean-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 (i.e., have smaller standard deviations) in some cases with $b = 3, 4$, they are uniformly dominated by their robust counterparts (i.e., have always larger values). Finally, the inclusion of population basis risk clearly lowers the hedging quality of the insurer by increasing the means and standard deviations of the optimal values of the objective functions, but the robust optimization still performs better than the nominal optimization in this case.

5 Conclusion

In this article we study a robust longevity risk management problem for an insurer with committed annuity payments, who uses survivor swaps as hedging instrument. We consider as objective functions the mean-variance and the mean-conditional-value-at-risk of the hedged liabilities. The insurer recognizes that the best estimated probability law affecting her liabilities may be subject to estimation inaccuracy, and optimizes her portfolio with respect to the worst-case scenario, where the worst case is with respect to the set of possible mortality laws determined by the Kullback-Leibler divergence.

We apply the robust optimization problems to Dutch male data and compare their performance with the corresponding nominal ones, which ignore estimation inaccuracy. We construct various realistic settings, where the estimated mortality distribution deviates from the actual one, 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 mortality law is further away from the insurer's estimate. The inclusion of basis risk lowers the insurer's hedge quality by increasing the means and standard deviations of the optimal values of the objective functions, which indicates that basis risk is an important risk factor affecting the insurer's longevity risk management. But whether basis risk is present or not does not affect the ordering of the performance of the robust optimization versus the nominal one.

Appendix

In this appendix we derive reformulations of the robust optimization problems (3.5) and (3.6).

Mean-variance—Denote by $\boldsymbol{L}(\boldsymbol{z}, \boldsymbol{a})$ the $I \times 1$ vector with the i-th entry $L(\boldsymbol{z}_i, \boldsymbol{a})$ and by $\boldsymbol{\iota}$ an $I \times 1$ vector of ones. The problem (3.5) can be reformulated as

$$\begin{aligned}
\min_{\boldsymbol{a},d} \max_{\boldsymbol{\pi}} \quad & d \\
s.t. \quad & \boldsymbol{\pi}'\boldsymbol{L}(\boldsymbol{z}, \boldsymbol{a}) + \lambda\boldsymbol{\pi}'(\boldsymbol{L}(\boldsymbol{z}, \boldsymbol{a}) - \boldsymbol{\pi}'\boldsymbol{L}(\boldsymbol{z}, \boldsymbol{a})\boldsymbol{\iota})^{\cdot 2} \leq d, \\
& \boldsymbol{a} \in \mathbf{A}, \boldsymbol{\pi} \in \hat{\Pi}, d \in R,
\end{aligned} \tag{.1}$$

where the superindex .2 means taking squares componentwise. Denote by $F(\boldsymbol{a}, \boldsymbol{h}) = h'\boldsymbol{L}(\boldsymbol{z}, \boldsymbol{a}) + \lambda\boldsymbol{h}'(\boldsymbol{L}(\boldsymbol{z}, \boldsymbol{a}) - \boldsymbol{h}'\boldsymbol{L}(\boldsymbol{z}, \boldsymbol{a})\boldsymbol{\iota})^{\cdot 2}$ and, for any set U,

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

Denote by

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

where $F_*(\boldsymbol{a}, \boldsymbol{\nu}) \equiv \inf_{\boldsymbol{h} \in R^I}\{\boldsymbol{h}'\boldsymbol{\nu} - F(\boldsymbol{a}, \boldsymbol{h})\}$ and $\delta^*(\boldsymbol{\nu}|\hat{\Pi}) \equiv \sup_{\boldsymbol{h} \in \hat{\Pi}} \boldsymbol{h}'\boldsymbol{\nu}$ are the conjugate function of $F(\boldsymbol{a}, \boldsymbol{h})$ (w.r.t. $\boldsymbol{a}$) and $\delta(\boldsymbol{\pi}|\hat{\Pi})$, respectively. Following Theorem 2 in Ben-Tal et al. (2014), for any fixed d, a $\boldsymbol{a} \in \mathbf{A}$ satisfies the first three constraints in problem (.1) if and only if there is a $\boldsymbol{\nu} \in R^I$ and $\boldsymbol{a}$

satisfying

$$\delta^*(\boldsymbol{\nu}|\hat{\Pi}) - F_*(\boldsymbol{a}, \boldsymbol{\nu}) \leq d. \tag{.3}$$

$\delta^*(\boldsymbol{\nu}|\hat{\Pi})$ can be reformulated as $\max_{\boldsymbol{h}\in\hat{\Pi}} \ \boldsymbol{h}'\boldsymbol{\nu}$ since $\hat{\Pi}$ is compact. Moreover, $\hat{\boldsymbol{\pi}}$ is regular since it is an element of $\hat{\Pi}$.[23] Therefore, following Ben-Tal et al. (2013), we have that $\boldsymbol{a} \in \mathbf{A}$ and $\boldsymbol{\nu} \in R^I$ satisfy (.3) with uncertainty region $\hat{\Pi}$ if and only if there exist $\eta \in R$ and $\xi > 0$ such that

$$\eta\rho + \xi + \eta \sum_{i=1}^{I} \hat{\pi}_i \exp(\frac{\nu_i - \xi}{\eta} - 1) - F_*(\boldsymbol{a}, \boldsymbol{\nu}) \leq d. \tag{.4}$$

Denote by $\boldsymbol{L} \equiv \boldsymbol{L}(\boldsymbol{z}, \boldsymbol{a})$ and $\boldsymbol{L}^2 \equiv \boldsymbol{L}^{\cdot 2}$. For any $\boldsymbol{a} \in \mathbf{A}$, $F_*(\boldsymbol{a}, \boldsymbol{\nu})$ can be reformulated as

$$\inf_{\boldsymbol{h}\in R^I} \boldsymbol{h}'\boldsymbol{\nu} - \boldsymbol{h}'(\boldsymbol{L} + \lambda\boldsymbol{L}^2) + \lambda\boldsymbol{h}'\boldsymbol{L}\boldsymbol{L}'\boldsymbol{h}. \tag{.5}$$

Write $G(\boldsymbol{h}) \equiv \boldsymbol{h}'\boldsymbol{\omega} + \lambda\boldsymbol{h}'\boldsymbol{L}\boldsymbol{L}'\boldsymbol{h}$ with $\boldsymbol{\omega} = \boldsymbol{\nu} - \boldsymbol{L} - \lambda\boldsymbol{L}^2$. Since $\boldsymbol{\nu}$ can take any value, $\boldsymbol{\omega}$ is also a free vector. Therefore, we can decompose $\boldsymbol{\omega}$ as

$$\boldsymbol{\omega} = \mathcal{K}\boldsymbol{L} + \boldsymbol{c}, \tag{.6}$$

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

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

Then $\inf_{\boldsymbol{h}} G(\boldsymbol{h}) = -\infty$ (using $\delta \to -\infty$). In this case, given some $d \in R$, constraint (.4) cannot be satisfied for any $\eta \in R$ and $\xi > 0$.

[23]For the definition of regularity of a vector, see Definition 1 in Ben-Tal et al. (2014).

If $\boldsymbol{c} = 0$, we can write $\boldsymbol{h} = \delta \boldsymbol{L} + \boldsymbol{d}$ with $\delta \in R$ and $\boldsymbol{d} \in N(\boldsymbol{L})$ such that

$$G(\delta \boldsymbol{L} + \boldsymbol{d}) = \mathcal{K}\delta \boldsymbol{L}'\boldsymbol{L} + \lambda\delta^2(\boldsymbol{L}'\boldsymbol{L})^2, \tag{.7}$$

which is a quadratic function of δ. Minimizing (.7) with respect to δ yields $\delta^* = -\frac{\mathcal{K}}{(\boldsymbol{L}'\boldsymbol{L})}$ and $G^* = -\frac{\mathcal{K}^2}{4\lambda}$. Therefore, using that in this case $\boldsymbol{\omega} = \mathcal{K}\boldsymbol{L}$, and thus $\boldsymbol{v} = \lambda \boldsymbol{L}^2 + (\mathcal{K}+1)\boldsymbol{L}$, we arrive at the following reformulation of (.1):

$$\begin{aligned} \min_{\boldsymbol{a},\eta,\mathcal{K},\xi} \;& \eta\rho + \xi + \eta \sum_{i=1}^{I} \hat{\pi}_i \exp\!\left(\frac{\lambda L^2(\boldsymbol{z}_i, \boldsymbol{a}) + (\mathcal{K}+1)L(\boldsymbol{z}_i, \boldsymbol{a}) - \xi}{\eta} - 1\right) + \frac{\mathcal{K}^2}{4\lambda} \\ s.t. \;& \boldsymbol{a} \in \mathbf{A}, \mathcal{K} \in R, \xi \in R, \eta > 0. \end{aligned} \tag{.8}$$

Mean-CVaR—We shall use the reformulation of the CVaR given by Rockafellar and Uryasev (2002), which is given by[24]

$$\text{CVaR}_{\alpha,P}(L(\boldsymbol{Z}, \boldsymbol{a})) = \min_{\xi \in R}\{\xi + \frac{1}{(1-\alpha)} E_P([L(\boldsymbol{Z}, \boldsymbol{a}) - \xi]^+)\}. \tag{.9}$$

where $[y]^+ = \max\{y, 0\}$. The $[L(\boldsymbol{z}_i, \boldsymbol{a}) - \xi]^+$ parts are nonlinear in $(\boldsymbol{a}, \xi)$. To turn (3.4) into a linear programming problem, we reformulate it as

$$\begin{aligned} \min_{\boldsymbol{a},\xi} \;& \hat{\boldsymbol{\pi}}'\boldsymbol{L}(\boldsymbol{z}, \boldsymbol{a}) + \lambda[\xi + \frac{1}{1-\alpha} \sum_{i=1}^{I} \hat{\pi}_i u_i] \\ s.t. \;& \boldsymbol{a} \in \mathbf{A}, \xi \in R, \\ & u_i \geq L(\boldsymbol{z}_i, \boldsymbol{a}) - \xi, \;\; \forall\, i \in \{1, 2, ..., I\} \\ & u_i \geq 0, \;\; \forall\, i \in \{1, 2, ..., I\}. \end{aligned} \tag{.10}$$

[24]The CVaR exists if $E_P|L(\boldsymbol{Z}, \boldsymbol{a})| < \infty$. This condition holds if $0 \leq \tau_x < \infty$ for all $x \in X_S$.

The corresponding robust optimization problem is given by

$$\begin{aligned}
\min_{\boldsymbol{a},\xi} \max_{\boldsymbol{\pi}} \quad & \boldsymbol{\pi}'\boldsymbol{L}(\boldsymbol{z},\boldsymbol{a}) + \lambda[\xi + \frac{1}{1-\alpha}\sum_{i=1}^{I} \pi_i u_i] \\
s.t. \quad & \boldsymbol{a} \in \mathbf{A}, \xi \in R, \boldsymbol{\pi} \in \hat{\Pi}, \\
& u_i \geq L(\boldsymbol{z}_i,\boldsymbol{a}) - \xi, \;\; \forall\, i \in \{1,2,...,I\} \\
& u_i \geq 0, \;\; \forall\, i \in \{1,2,...,I\}.
\end{aligned} \tag{.11}$$

The uncertain vector, $\boldsymbol{\pi}$, is linear in (.11). The derivation of the reformulation of (.11) that we use is therefore a direct application of Theorem 1 in Ben-Tal et al. (2013). We find as reformulation

$$\min_{\boldsymbol{a},\xi,u,\zeta,\eta} \quad \lambda\xi + \rho\zeta + \eta + \zeta \sum_{i=1}^{I} \hat{\pi}_i \exp\left(\frac{L(\boldsymbol{z}_i,\boldsymbol{a}) + \frac{\lambda}{1-\alpha}u_i - \eta}{\zeta} - 1\right)$$

$$\boldsymbol{a} \in \mathbf{A}, \xi \in R, \eta \in R, \zeta \geq 0, \tag{.12}$$

$$u_i \geq L(\boldsymbol{z}_i,\boldsymbol{a}) - \xi, \;\; \forall i \in \{1,2,...,I\}$$

$$u_i \geq 0, \;\; \forall i \in \{1,2,...,I\}. \tag{.13}$$

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.

Bauer, D., M. Börger, and J. Ruß (2010). On the pricing of longevity-linked securities. *Insurance: Mathematics and Economics 46*(1), 139–149.

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.-P. Vial (2014). Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 1–35.

Ben-Tal, A., L. El Ghaoui, and A. Nemirovski (2009). *Robust optimization*. Princeton University Press.

Bertsimas, D., D. B. Brown, and C. Caramanis (2011). Theory and applications of robust optimization. *SIAM review 53*(3), 464–501.

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.

Garlappi, L., R. Uppal, and T. Wang (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies 20*(1), 41–81.

Glasserman, P. and X. Xu (2013). Robust portfolio control with stochastic factor dynamics. *Operations Research 61*(4), 874–893.

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. (2014). Dynamic hedging of longevity risk: the effect of trading frequency. *Available at SSRN 2532422*.

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.

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

Mulvey, J. M., R. J. Vanderbei, and S. A. Zenios (1995). Robust optimization of large-scale systems. *Operations research 43*(2), 264–281.

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 Eco-*

nomics 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 (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance 26*(7), 1443–1471.

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

Wong, T. W., M. C. Chiu, and H. Y. Wong (2014). Time-consistent mean–variance hedging of longevity risk: Effect of cointegration. *Insurance: Mathematics and Economics 56*, 56–67.

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

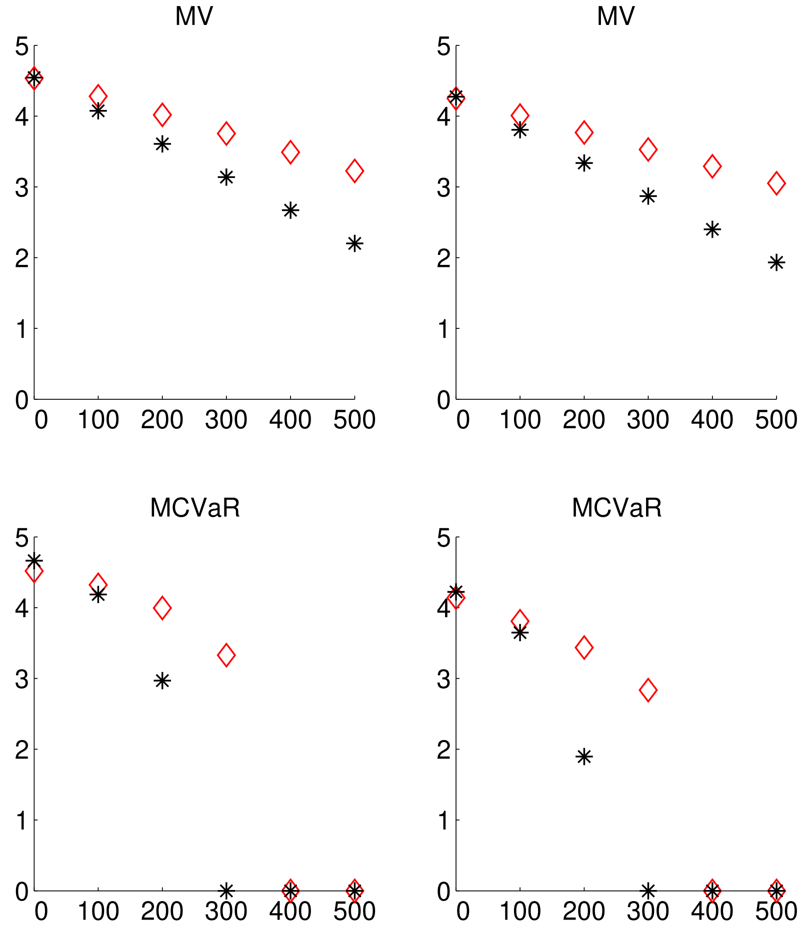

Figure 1: The optimal amount of swaps purchased by the insurer as a function of the risk premium in basis points. In each plot the diamonds are the optimal a-s for the robust optimizations and the asterisks are the optimal a-s for the nominal optimizations. The left panels are without basis risk and the right panel with basis risks. The upper panels show the mean-variance objective values, and the lower ones the mean-CVaR objective values.

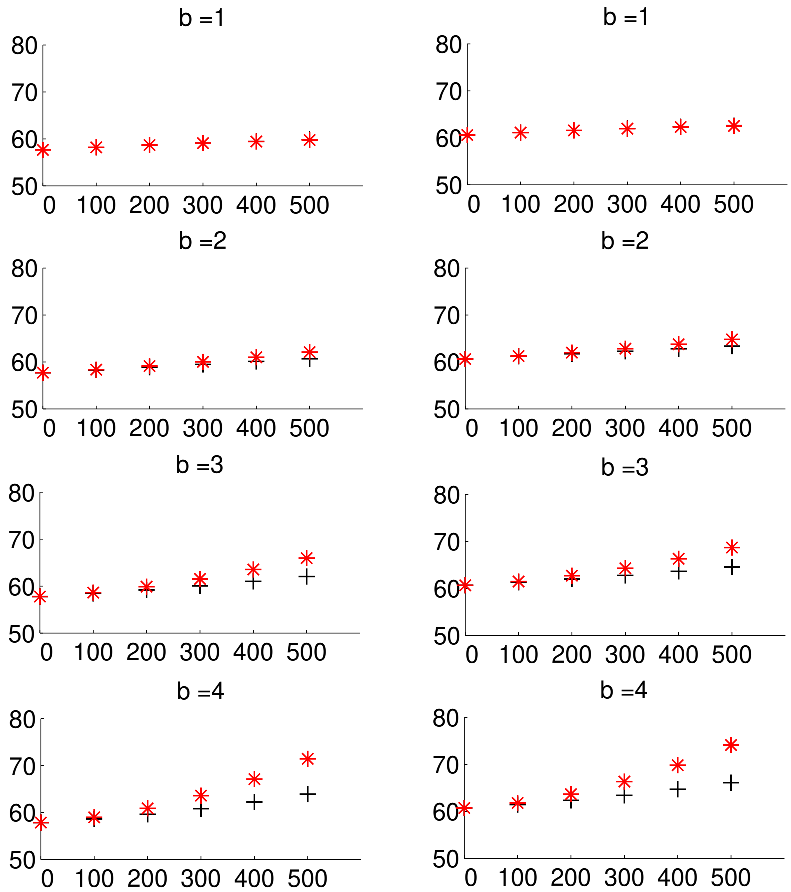

Figure 2: The mean of the optimal values of the objective functions for the Mean-Variance case as a function of the risk premium in basis points. The plus signs denote the robust means and the asterisks denote the nominal means. The left panels display the results without basis risk, and the right panels display results with basis risks.

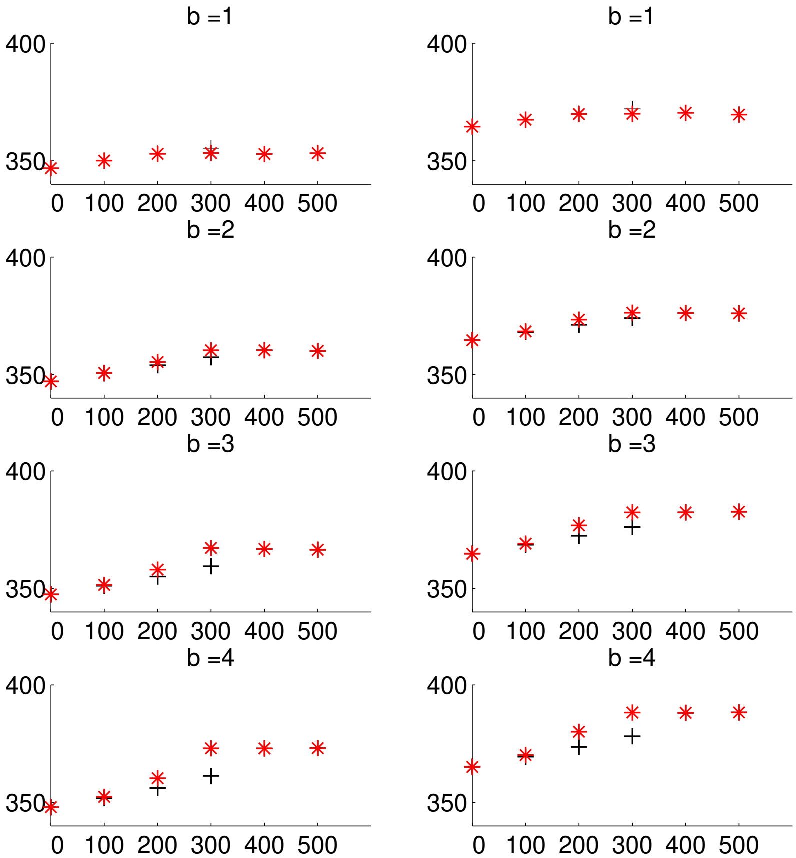

Figure 3: The mean of the optimal values of the objective functions for the CVaR case as a function of the risk premium in basis points. The plus signs denote the robust means and the asterisks denote the nominal means. The left panels display the results without basis risk, and the right panels display results with basis risks.

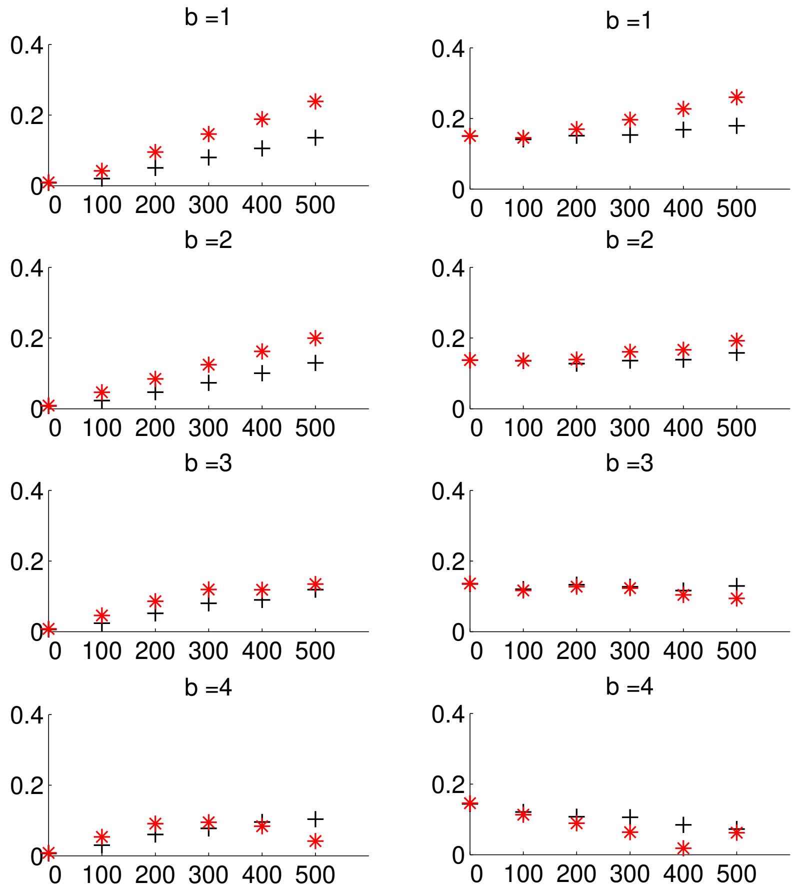

Figure 4: The standard deviation of the optimal values of the objective functions for the Mean-Variance case as a function of the risk premium in basis points. The plus signs denote the robust means and the asterisks denote the nominal means. The left panels display the results without basis risk, and the right panels display results with basis risks.

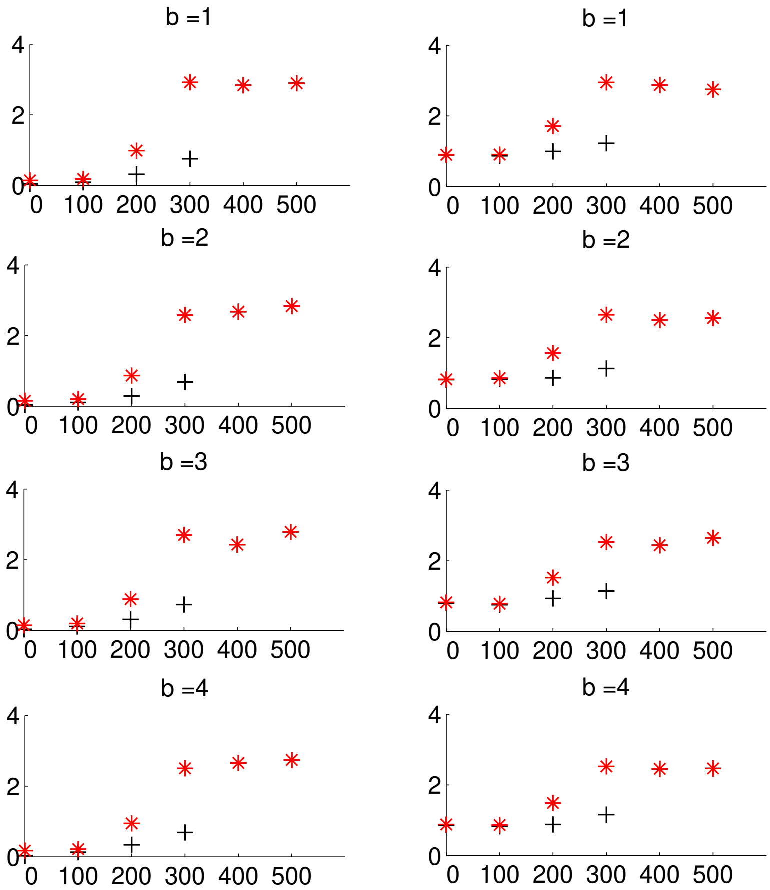

Figure 5: The standard deviation of the optimal values of the objective functions for the CVaR case as a function of the risk premium in basis points. The plus signs denote the robust means and the asterisks denote the nominal means. The left panels display the results without basis risk, and the right panels displays results with basis risks.