Annals of Actuarial Science, Vol. 11, part 2, pp. 286–314. © Institute and Faculty of Actuaries 2017
doi:10.1017/S174849951700001X
First published online 3 April 2017

Demographic risk in deep-deferred annuity valuation

Min Ji
Department of Mathematics, Towson University, Towson, MD, USA

Rui Zhou*
Warren Centre for Actuarial Studies and Research, University of Manitoba, Winnipeg, MB, Canada R3T 5V4

Abstract
A deep-deferred annuity is a deferred annuity where payments start very late in life, i.e. well after the normal retirement age. This annuity has received much attention lately as it was made accessible to 401(k) plans in the United States in 2014. By transferring the risk of outliving retirement savings at high ages to annuity providers, deep-deferred annuities provide annuitants with enhanced later-life financial security. However, the valuation of this annuity suffers from high uncertainty because the mortality data at high ages are sparse and possibly unreliable. In this paper, we use risk ratio to measure demographic risk in the valuation. Demographic risk is decomposed into the following four components: (1) mortality tail curve risk, (2) mortality improvement model risk, (3) parameter risk in mortality tail curves, and (4) parameter risk in mortality improvement rate models. Our quantitative analysis aims to provide insights into the development and risk management of deep-deferred annuities.

Keywords
High age mortality; Mortality improvement rate model; Deep-deferred annuity; Model risk; Parameter risk

1. Introduction

According to the Employee Benefit Research Institute's 2014 Retirement Confidence Survey, only 18% of workers are confident that they will have enough money for a comfortable retirement while many are concerned about outliving their retirement savings. A deep-deferred annuity, an innovative annuity product proposed by Milevsky (2005), can potentially alleviate this concern. Also known as advanced life deferred annuity or longevity annuity, a deep-deferred annuity is a deferred annuity which is usually purchased just before or after retirement but does not begin to make regular payments until annuitants reach a certain high age, typically 85.

This type of annuity is designed to enhance later-life financial security in a cost-effective way. It has a pure survivorship benefit, i.e. there is no death benefit if the policyholder dies before the annuity start age. Gong & Webb (2010) showed that deep-deferred annuities provide a substantial proportion of longevity risk protection offered by immediate annuities at a much lower cost. Blake & Turner (2014) stated that the extra financial security from deep-deferred annuities helps retirees better manage their post-retirement assets. The purchaser of deep-deferred annuities only needs to plan for

*Correspondence to: Rui Zhou, Warren Centre for Actuarial Studies and Research, University of Manitoba, 181 Freedman Crescent, Winnipeg, MB, Canada R3T 5V4. Tel: (1)204-474-9783; E-mail: rui.zhou@umanitoba.ca

a fixed period from the retirement date to the date at which the annuity benefits start, instead of planning for the uncertain lifespan.

With the rapid mortality improvement we have been experiencing recently, there is a high likelihood of retirees underestimating their life expectancy and outliving their retirement assets. Deep-deferred annuities can potentially relieve this problem and thus play an important role in our social security system. In fact, the US Department of Treasury and the Internal Revenue Service approved the usage of longevity annuities in the 401(k) and Individual Retirement Account (IRA) markets in 2014. However, only a few insurance companies in the United States offer this annuity, and no insurance companies in the United Kingdom currently provide this product. Blake & Turner (2014) argued that the current low supply of deep-deferred annuities in the private sector is primarily attributable to the systematic longevity risk, especially at very high ages. In particular, the increased difficulty of assessing longevity risk due to the long deferral period and lack of appropriate longevity risk-hedging tools will lead to a large increase of reserve requirement for this type of annuity.

The biggest challenge in assessing longevity risk in deep-deferred annuities is how to model and forecast high age mortalities. This question has remained open despite numerous past attempts. The sparseness and low reliability of mortality data at high ages exacerbate the problem, thereby leading to concerns of model risk and parameter risk in annuity valuation. In this paper, we aim to quantify these two risks and identify their driving factors. In addition, we use immediate annuity as a comparative tool and show that it has very different risk profiles than deep-deferred annuity.

The conventional mortality modelling approaches, such as the log-bilinear structure (Lee & Carter, 1992; Brouhns *et al.*, 2002; Renshaw & Haberman, 2003), directly model the mortality rates over time and then extrapolate the pattern using a simple time series method. These approaches work well when the focus is on relatively low ages at which mortality data are accurate. However, for deep-deferred annuities, we are more concerned with high ages where data are sparse. Since it is difficult to validate models using a small amount of data, we are more interested in how model assumptions affect the valuation of deep-deferred annuities. Therefore, we opt for an approach that is comprised of two parts: a base mortality curve and a mortality improvement model. This approach allows us to adapt different assumptions of high age mortality curves and mortality improvement patterns, and thus isolate their impacts on annuity valuation.

The base mortality curve determines how mortality changes with age, and – more importantly for deep-deferred annuities – how we close the mortality table. There have been different views of the trajectory of mortality at high ages. Researchers (Perks, 1932; Beard, 1963, 1971; Kannisto, 1992) have described the mortality pattern by logistic models, and suggested that exponential growth of mortality with age is followed by slower rates of mortality increase at higher ages. Gompertz (1825) and Greenwood & Irwin (1939) observed a late-life mortality plateau, i.e. level or declining mortality at high ages. Nusbaum *et al.* (1996) and Weitz & Fraser (2001) have built models to incorporate and explain this phenomenon. However, Gavrilov & Gavrilova (2011) have pointed out that evidence supporting mortality decelerations at high ages was caused by population heterogeneity, and the authors concluded that the Gompertz law is a good fit for the population studied up to the age of 106. To date, there is no consensus on mortality tail shape. In this paper, we consider various mortality curves and assess the model risk resulting from this assumption.

Mortality improvement rates have been modelled directly with parametric structures in Mitchell *et al.* (2013) and Haberman and Renshaw (2012). In this paper, we consider both the Mitchell *et al.* (2013)

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

model and an indirect method which obtains the improvement rates based on the Lee–Carter model. Proposed by Lee & Carter (1992), this model has been well-recognised for its parsimony and easy interpretation. It is often regarded as the benchmark in mortality modelling.

Another important determinant of annuity value is gender. Male and female mortalities are different but highly correlated. It is necessary to model them jointly using a two-population mortality model, such that their correlations can be adequately captured. Failure to do so may lead to underestimation of potential risk in the annuity. Multi-population mortality models have been studied by several researchers (Cairns *et al*., 2011; Dowd *et al*., 2011; Li & Hardy, 2011; Zhou *et al*., 2014; Li *et al*., 2015*a*, 2015*b*). We follow Zhou *et al*. (2014) and consider two versions of two-population Lee–Carter model: one with non-divergence constraints and one without.

With the non-divergence constraints, the mortality forecasts of the two populations being modelled will not diverge indefinitely over time. These constraints have been used by Li & Lee (2005), Cairns *et al*. (2011), Dowd *et al*. (2011), and Li *et al*. (2015*a*). Non-divergence has been supported by certain empirical observations. For instance, Wilson (2001) found a global convergence in mortality levels by comparing the distributions of the global population by life expectancy over three different time periods: 1950–1955, 1975–1980, and 2000. However, some researchers believe that the non-divergence constraints are too strong. Oeppen & Vaupel (2002) studied the trend of highest female life expectancy across countries and observed that US females have moved away from this trend in recent years. Li *et al*. (2015*b*) introduced the semi-coherent concept to allow temporary divergence, and demonstrated that the assumption of divergence/non-divergence has significant impact on longevity risk pricing and hedging. Inspired by the literature, we compare the value of deep-deferred annuities under divergence and non-divergence assumptions, and evaluate the model risk stemming from this assumption.

As mentioned above, parameter risk is also of great concern because of the scarcity and low quality of mortality data at high ages. To quantify the parameter risk in valuing deep-deferred annuities, we use the bootstrap method to determine the variability of estimated parameters and its impact on annuity values.

The remainder of this paper is organised as follows: section 2 introduces mortality tail curves; section 3 describes the single-population and two-population Lee–Carter models; sections 4 and 5 examine model risk and parameter risk in valuing deep-deferred annuities, respectively; and section 6 concludes the paper.

2. Mortality Tail Curves

2.1. Three shapes of mortality curves

Since a deep-deferred annuity does not start to pay benefits until the annuitant reaches a very advanced age, its cash flow may be greatly affected by the tail of a mortality curve. To examine the impact of mortality tail assumption on the annuity value, we consider the following three models:

(1) Gompertz law (Gompertz, 1825):

$$\mu_x = e^{ax+b}$$

where μ_x is the force of mortality for a life at age x, and a and b are constants. Under Gompertz law, the force of mortality increases exponentially with age. The mortality curve is concave upward without bound. Gompertz law has been used widely because of its simplicity and tractability.

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

(2) Cubic model with a flat tail (Panjer & Russo, 1992; Panjer & Tan, 1995):

$$\mu_x = \min\left(ax^3 + bx^2 + cx + d,\ \ln 2\right)$$

where a, b, c, and d are constants. The force of mortality increases as a cubic function of age until it reaches the cap, ln2, and then remains at ln2 for higher ages. This mortality curve models a late-life mortality plateau. The recently released RP-2014 Mortality Tables[1] by the Society of Actuaries (SoA) uses a similar mortality curve, and sets the death probability to 0.5 for males over 110-year-old and females over 112-year-old.

(3) Logistic model (Perks, 1932)

$$\mu_x = \frac{e^d + e^{ax+b}}{1 + e^{ax+c}}$$

where a, b, c, and d are constants. The force of mortality accelerates first, and then decelerates at higher ages, gradually approaching an asymptotic maximum. There are two variants of this model: Beard (1963), where $e^d = 0$; and Kannisto (1992), where $e^d = 0$ and $b = c$.

2.2. Data and fitting

We use Japanese mortality data in year 2012 with the age range of 75–105[2] to fit the three mortality curves described above.

We recognise that the market for deep-deferred annuities mainly lies in the United States. We choose to use Japanese mortality data because there is no US annuitant mortality data publicly available. Our paper focuses on how the assumptions of mortality tail curve and mortality improvement at high ages affect deep-deferred annuity valuation. High age mortality data are crucial for our study. One alternative is to use the US national population data, but it is recognised that annuitant mortality is much lighter than national population mortality due to adverse selection (Mitchell & McCarthy, 2002). Using the national population data may result in underestimation of demographic risk in deep-deferred annuities.

Japan is one of the countries that have the world's lowest mortality rates (Robine *et al.*, 2010), and it has a relatively large population. Japanese mortality data are of high quality since Japan implemented a national family registration system in 1872. Researchers (Robine & Saito, 2003) have found that Japanese mortality data offer some unique information on the demographic trends for extremely old population. Japanese data are also reliable for investigating the determinants and limits of human longevity and the important plasticity in longevity with regard to environmental conditions (Arai *et al.*, 2014). These features of Japanese mortality data prompt us to use them in studying demographic risk in deep-deferred annuity valuation.

The number of deaths at each age is assumed to follow a Poisson distribution with mean equal to the expected number of deaths under a mortality curve model:

$$D_{x,2012} \sim \text{Poisson}(E_{x,2012}\mu_{x,2012})$$

where $D_{x,t}$, $E_{x,t}$, and $\mu_{x,t}$ are the number of deaths, the number of exposures-to-risk, and the force of mortality of lives at age x in year t, respectively. Maximum likelihood estimation is used to

[1] The abbreviation "RP" stands for Retirement Plans.
[2] Data were downloaded from Human Mortality Database, http://www.mortality.org

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

Table 1. Parameter estimates of the three mortality curves.

	Males			Females		
Parameters	Gompertz	Cubic	Perks'	Gompertz	Cubic	Perks'
a	0.1115	1.1103×10^{-5}	0.1149	0.1320	1.0880×10^{-5}	0.1658
b	−11.8143	−0.0023	−14.5120	−14.1673	−0.0023	−16.9970
c		0.1548	−14.1351		0.1553	−16.5817
d		−3.5622	−5.2205		−3.5479	−5.7801
$\ln(L)$	−586.36	−198.73	−201.98	−1271.52	−331.85	−244.56
BIC	1179.59	411.20	417.70	2549.91	677.44	502.86

Note: BIC, Bayesian information criteria.

estimate the parameters in the mortality curve models. The log-likelihood function can be expressed as

$$\sum_x (D_{x,2012} \ln \mu_{x,2012} - \mu_{x,2012} E_{x,2012}) + \text{constant}$$

To maximise the log-likelihood, we can use standard Newton's method or optimisation functions in computational softwares. We obtain our estimates using the optimisation function in Matlab. Table 1 summarises the parameter estimates.

We use the Bayesian information criteria (BIC)[3] to compare the goodness of fit of the three models. Perks' logistic model has the lowest BIC for females and is thus the best fit for female mortality data. While Cubic model is the best fit for male mortality data, Perks' model fits male mortality data slightly worse. We have also fitted Beard's and Kannisto's logistic models, and found that they do not yield better goodness of fit than Perks' model. Therefore, we will only consider Perks' logistic model hereinafter.

The upper panel of Figure 1 plots raw mortality rates and fitted curves for males and females. The three curves have similar fits for ages under 95; however, they show very different patterns for ages over 95. The fitted male mortality curve using Perks' model is concave upward before age 97 and concave downward after age 97. It gradually approaches an asymptote of 0.686. The fitted female mortality curve using Perks' model has a reflection point at age 100 and an asymptote of 0.660. The fitted curve using cubic model is concave upward for ages over 75. It reaches the imposed cap of ln2 at age 106 for males and age 109 for females, respectively.

The goodness of fit of Gompertz law is much less accurate than that of the other two models. The fitted curve deviates significantly from the raw mortality rates beyond age 95 for males and age 97 for females. However, Gavrilov & Gavrilova (2011) found that Gompertz law fits the US Social Security Administration (SSA) Death Master File mortality data[4] of ages 88–106 better than logistic

[3] The BIC is defined as $BIC = -2 \ln(L) + k \ln(n)$, where $\ln(L)$ is the log-likelihood, k the number of parameters, and n the number of data points. The smaller the BIC is, the better fit the model has.

[4] There is no publicly available US annuity portfolio data that can be used for our purpose. As pointed out in Gavrilov & Gavrilova (2011), mortality rates of ages 85–88 before 1970 cannot be estimated using the SSA Death Master File mortality data due to incomplete death registration before 1970. The SSA life tables cannot be used either, because mortality rates for age 95 and above are extrapolated. Our results will be significantly affected by the extrapolation method employed in the SSA life tables.

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

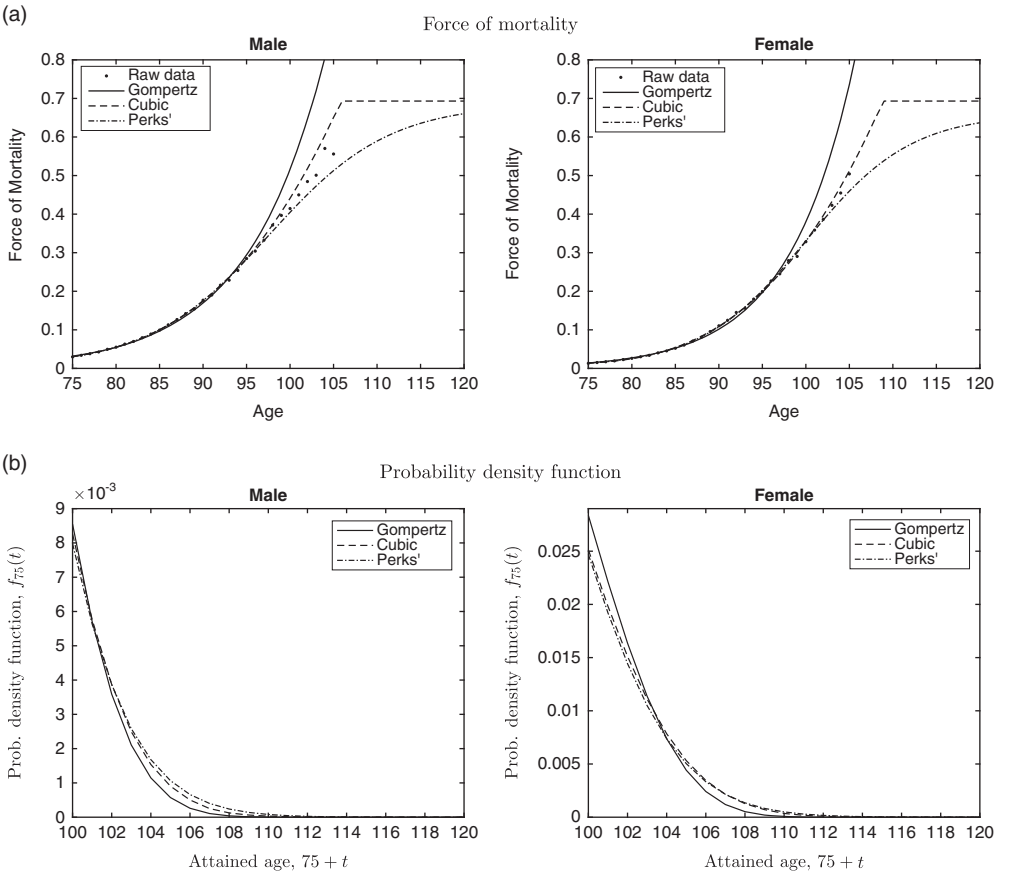

Figure 1. (a) Raw mortality rates and the three fitted mortality curves for a male and female currently at age 75. (b) Probability density function of the future lifetime of a person aged 75 in 2012 using the three mortality curves.

models. They concluded that improvement of data quality at advanced ages results in a mortality trajectory closer to that of the Gompertz law. Therefore, determining which model fits best appears to be very much data dependent. Practitioners should carefully choose the model for mortality tail curve based on the mortality experience of their own annuity portfolios. The large difference between the three fitted mortality curves suggests that mortality tail curve may be an important source of model risk for deep-deferred annuity valuation.

Denote $f_{75}(t)$ as the probability density function of the future lifetime of a person aged 75 in 2012, and $f_{75}(t) = {}_tp_{75}\mu_{75+t,2012}$, where ${}_tp_{75}$ is the probability that this person survives t years. In the lower panel of Figure 1, we plot $f_{75}(t)$, $t = 25, 26, \ldots, 45$ for both males and females using the three mortality tail curves. Gompertz law has the lightest tail because it yields the lowest probability of surviving to very high ages among the three mortality tail curves. We also notice that the value of $f_{75}(t)$ for males is much smaller than that for females, because males have much lower probability of surviving to very high ages. In order to display the difference between the density functions using various mortality tail curves clearly, we use different scales for males and females in these two plots.

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

We construct the complete mortality curve with raw mortality rates for ages below 75 and fitted mortality rates for ages above 75 using the models described in section 2.1 because the raw mortality rates for younger ages are believed to be reliable. We use raw mortality rates for ages below 75 in the mortality curve to avoid fitting errors. In addition, using the same mortality rates for younger ages ensures that the difference of annuity values comes solely from the mortality tail curve assumption.

3. Mortality Improvement Rate

3.1. Single-population Lee–Carter model

In the main text, we use the Lee–Carter model to obtain mortality improvement rates, because it is often regarded as the benchmark for mortality modelling. In Appendix A, we introduce the Mitchell *et al.* (2013) model and show that these two models result in similar findings.

Mathematically, the Lee–Carter model can be expressed as follows:

$$\ln m_{x,t} = \alpha_x + \beta_x \kappa_t$$

where $m_{x,t}$ is the central death rate for a life aged x in year t, α_x the age-specific mortality level, κ_t the time-varying mortality index, and β_x the age-specific sensitivity to the mortality index. Following Brouhns *et al.* (2002), we assume that $D_{x,t}$ follows a Poisson distribution with mean equal to $E_{x,t}m_{x,t}$. Note that Poisson assumption is commonly violated by overdispersion, which is often evident in death count data due to population heterogeneity. Li *et al.* (2009) addressed this problem by adding a gamma distribution for the unobserved heterogeneity. For simplicity, we continue to use the classic Poisson distribution and maximum likelihood estimation to obtain parameter estimates. To ensure parameter uniqueness, two identification constraints are imposed: $\sum_x \beta_x = 1$ and $\sum_t \kappa_t = 0$. κ_t is further modelled by an autoregressive integrated moving average (ARIMA) model.

Mortality improvement rate at age x and time t can be written as follows:

$$r_{x,t} = \ln \frac{m_{x,t}}{m_{x,t-1}} = \ln \frac{e^{\alpha_x + \beta_x \kappa_t}}{e^{\alpha_x + \beta_x \kappa_{t-1}}} = \beta_x (\kappa_t - \kappa_{t-1}) \tag{3.1}$$

Equation (3.1) indicates that mortality improvement rate for age x is determined by β_x and the change in κ_t. It is important to note that the cohort effect is another important factor driving mortality improvement due to the phenomenon that people born in certain years have experienced more rapid mortality improvement than others. Cohort effect has been observed in both United Kingdom and Japanese populations by Willets (2004). Several researchers (Renshaw & Haberman, 2003; Cairns *et al.*, 2009) have incorporated cohort effect in mortality modelling, and it is considered an important factor in annuity pricing and valuation. Although the Lee–Carter model cannot capture cohort effect, it has been chosen because the focus of our study is on the uncertainty stemming from the assumption of mortality tail curve and the dependence structure of male and female mortality. Future research may consider a collection of mortality models with cohort effect in the analysis.

The forecast of mortality rates is the multiplication of the base year mortality curve and the projected mortality improvement rate. This two-step approach allows us to separately examine the impact of mortality improvement model and the base mortality curve on deep-deferred annuity valuation.

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

In this paper, we use 2012 as the base year. For a life aged x at the beginning of 2013, his/her death rate in year $2013+s$, where $s = 0, 1, 2, \ldots$, can be expressed as follows:

$$\begin{aligned} m_{x+s,2013+s} &= \mu_{x+s,2012} \prod_{j=0}^{s} e^{\gamma_{x+s,2013+j}} \\ &= \mu_{x+s,2012} \prod_{j=0}^{s} e^{\beta_{x+s}\left(\kappa_{2013+j} - \kappa_{2013+j-1}\right)} \\ &= \mu_{x+s,2012} e^{\beta_{x+s}\left(\kappa_{2013+s} - \kappa_{2012}\right)} \end{aligned}$$

where $\mu_{x+s,2012}$ is obtained from the base mortality curve in section 2. This approach is analogous to the mortality improvement scale approach used in the SoA's Mortality Improvement Scale MP – 2015. In SoA's approach, future mortality rates in year $t+s$ are calculated by

$$m_{x,t+s} = m_{x,t} \prod_{j=1}^{s} (1 - IS(x, t+j))$$

where IS (x; s) is the improvement scale, retrieved from a two-dimensional table with one dimension for age and the other for calendar year. The base year mortality curve is the RP-2014 Mortality Table. However, this tabulated approach can only provide point forecasts of future mortality rates, but not the uncertainty of the forecasts. In contrast, our approach is stochastic and can capture the uncertainty in mortality improvement rates.

If we use the Lee–Carter model directly, the forecasted value of $m_{x+s,2013+s}$ is

$$\begin{aligned} m_{x+s,2013+s} &= e^{\alpha_{x+s} + \beta_{x+s}\kappa_{2013+s}} \\ &= e^{\alpha_{x+s} + \beta_{x+s}\kappa_{2012} + \beta_{x+s}\left(\kappa_{2013+s} - \kappa_{2012}\right)} \\ &= \hat{m}_{x+s,2012} e^{\beta_{x+s}\left(\kappa_{2013+s} - \kappa_{2012}\right)} \end{aligned}$$

where $\hat{m}_{x+s,2012} = e^{\alpha_{x+s} + \beta_{x+s}\kappa_{2012}}$ and it is the fitted value of $m_{x+s,2012}$. Therefore, using the Lee–Carter model to directly forecast mortality rates is equivalent to our two-step approach with the base mortality curve set to the fitted Lee–Carter mortality rates in the base year.

We fit the Lee–Carter model to the Japanese male and female mortality data from the sample period of 1965–2012. Due to the concern of data reliability, we use the sample age range of 60–95. Data before year 1965 were dropped due to greater volatility compared to subsequent years. Figure 2 demonstrates the estimated values of α_x, β_x, and κ_t for males and females, respectively. α_x increases almost linearly with age, indicating that mortality rates increase with age. As shown in equation (3.1), mortality improvement rate is determined by β_x and $\kappa_t - \kappa_{t-1}$. Since κ_t decreases with t and β_x is positive, mortality rates decrease over time for all ages. β_x attains its maximum value at approximately age 75, thus suggesting that people who were approximately age 75 experienced faster mortality improvement than other age groups. We also observe that female κ_t has a steeper slope than male κ_t, and females over age 77 have higher β_x than males. This observation implies that the mortality improvement of females over age 77 is faster than that of males. Note that other countries may have different experiences. For example, the mortality of UK males has improved more rapidly than females since 1970, according to the 2012-based National Population Projections report by the Office for National Statistics (2013).

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

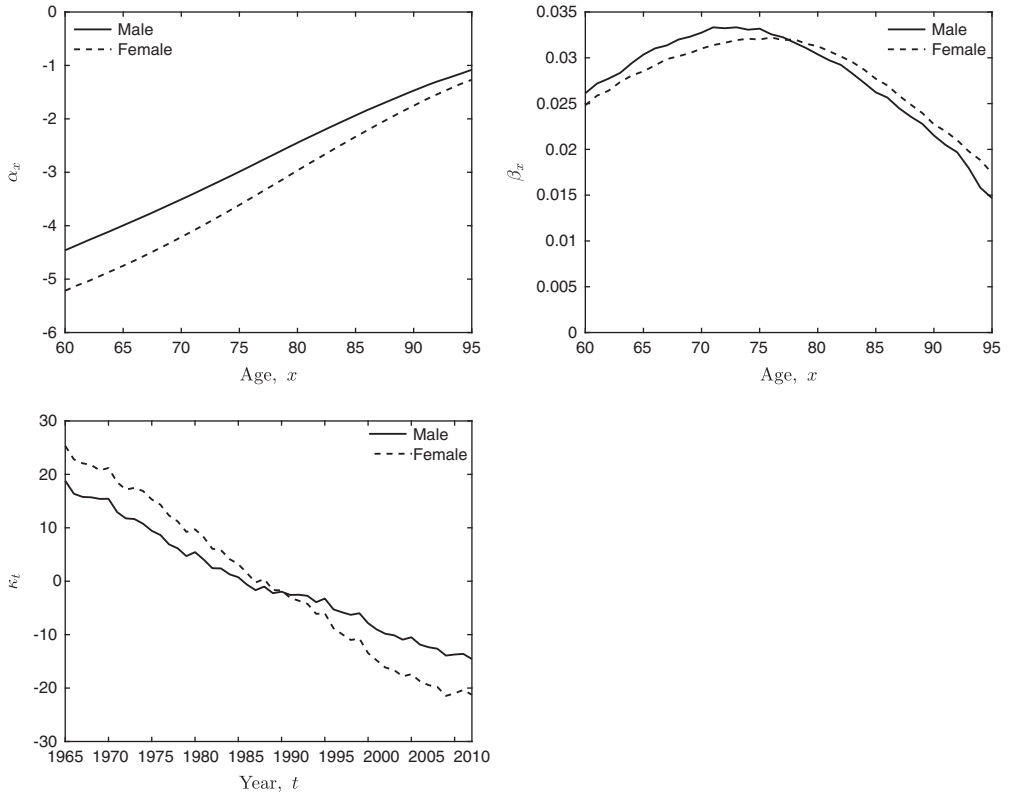

Figure 2. Estimates of α_x, β_x, and κ_t.

We selected ARIMA(0, 1, 1) model for both male and female κ_t based on BIC values. The ARIMA(0, 1, 1) model can be expressed as follows:

$$\Delta\kappa_t = a_0 + z_t + a_1 z_{t-1}$$

where $\Delta\kappa_t = \kappa_t - \kappa_{t-1}$ and $\{z_t\}$ are independently and identically distributed normal random variables with mean 0 and variance σ^2. The parameter estimates for the ARIMA model and their standard errors (in parentheses) are shown in Table 2. Depending on the choice of identifiability constraints, the estimates of κ_t may change, thereby changing the estimates of ARIMA parameters as well. However, the structure of the fitted ARIMA model and future morality forecast will remain the same.

Since the sample age range of the data is 60–95, we can only obtain the estimates of α_x and β_x up to age 95. For ages above 95, we extrapolate the patterns of α_x and β_x using deterministic functions. It is reasonable to assume that β_x is non-negative, because mortality is expected to continue to improve, or at least not to deteriorate. With this constraint, we consider two models for β_x: a Gaussian function with an asymptote of 0, and a quadratic function with a minimum value of 0. Using Gaussian function, β_x decreases slowly and approaches 0 as x goes to infinity. Using a quadratic function, β_x decreases quickly and reaches the floor of 0 when $x = 105$. Both cases are displayed in Figure 3. Since α_x does not determine mortality improvement rate, as seen in equation (3.1), the choice of extrapolation function will not affect our conclusion. We can use either a Gaussian function or linear function with a cap of 0 to extrapolate α_x.

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

Table 2. Parameter estimates of ARIMA(0, 1, 1) for κ_t in the single-population Lee–Carter model.

Gender	a_0	a_1	σ
Male	−0.6794 (0.0662)	−0.3784 (0.1350)	0.5606 (0.1649)
Female	−0.9521 (0.0835)	−0.3728 (0.1397)	0.8150 (0.2238)

Note: ARIMA, autoregressive integrated moving average.

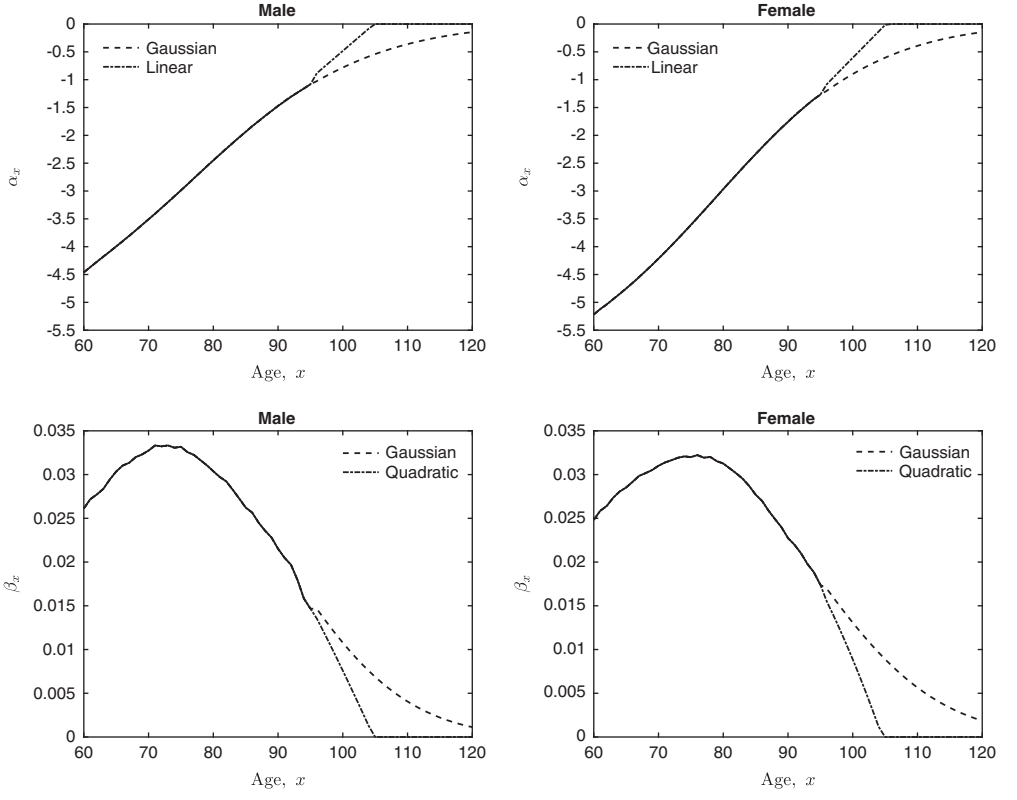

Figure 3. Extrapolation of α_x and β_x at high ages.

3.2. Two-population Lee–Carter model

To build a two-population model, we use one Lee–Carter model for each population:

$$\ln m_{x,t}^{(i)} = \alpha_x^{(i)} + \beta_x^{(i)} \kappa_t^{(i)}$$

where $i = 1$ for males and 2 for females. Following Zhou *et al.* (2014), we further model $\left[\Delta\kappa_t^{(1)}, \Delta\kappa_t^{(2)}\right]'$ by a pth order vector autoregressive (VAR) process:

$$\begin{pmatrix} \Delta\kappa_t^{(1)} \\ \Delta\kappa_t^{(2)} \end{pmatrix} = A_0 + A_1 \begin{pmatrix} \Delta\kappa_{t-1}^{(1)} \\ \Delta\kappa_{t-1}^{(2)} \end{pmatrix} + A_2 \begin{pmatrix} \Delta\kappa_{t-2}^{(1)} \\ \Delta\kappa_{t-2}^{(2)} \end{pmatrix} + \ \ldots \ + A_p \begin{pmatrix} \Delta\kappa_{t-p}^{(1)} \\ \Delta\kappa_{t-p}^{(2)} \end{pmatrix} + \begin{pmatrix} z_t^{(1)} \\ z_t^{(2)} \end{pmatrix}$$

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

Table 3. *p*-values of likelihood ratio tests for determining vector autoregressive (VAR) lag order.

	Alternative		
Null	VAR(2)	VAR(3)	VAR(4)
VAR(1)	0.0309	0.0687	0.1347
VAR(2)		0.4179	0.5167
VAR(3)			0.5822

where A_0 is a 2×1 vector, A_j, for $j = 1, 2, \ldots, p$, is a 2×2 matrix, and $[z_t^{(1)}, z_t^{(2)}]'$ follows a bivariate normal distribution with 0 mean and variance–covariance matrix of Σ.

We use likelihood ratio tests to determine the lag order p. Using a significance level of 5%, a p-value smaller than 5% implies strong evidence against the null model. Table 3 summarises the test results and shows that the tests favour VAR(1), VAR(2), and VAR(3) over VAR(4), indicating that lag order of <4 should be used. Similarly, the tests favour VAR(1) and VAR(2) over VAR(3), indicating that lag order of <3 should be used. Lastly, the test of VAR(1) versus VAR(2) rejects VAR(1) at the p-value of 0.0309. Therefore, we select VAR(2) to model $\left[\Delta\kappa_t^{(1)}, \Delta\kappa_t^{(2)}\right]'$.

The parameter estimates and their standard errors (in parentheses) are displayed below:

$$A_0 = \begin{bmatrix} -1.1174 \\ (0.2118) \\ -1.2597 \\ (0.2639) \end{bmatrix}, A_1 = \begin{bmatrix} 1.0202 & -1.1513 \\ (0.4252) & (0.3313) \\ 0.9640 & -1.0658 \\ (0.5299) & (0.4129) \end{bmatrix}, A_2 = \begin{bmatrix} -0.6644 & 0.4750 \\ (0.4292) & (0.3846) \\ -1.2924 & 1.0120 \\ (0.5349) & (0.4793) \end{bmatrix}, \text{ and } \Sigma = \begin{bmatrix} 0.4307 & 0.5075 \\ 0.5075 & 0.6689 \end{bmatrix}$$

This two-population Lee–Carter model can capture the correlation between the mortality rates of the two populations. However, it does not impose any non-divergence conditions. Researchers (Li & Lee, 2005; Cairns *et al.*, 2011; Dowd *et al.*, 2011; Li *et al.*, 2015*b*) have argued that mortality rates of different populations should not diverge indefinitely due to biological reasonableness. To incorporate the non-divergence assumption and study how it affects deep-deferred annuity valuation, we consider the following constraints:

1. $\beta_x^{(1)} = \beta_x^{(2)}$;
2. $\Delta\kappa_x^{(1)}$ and $\Delta\kappa_x^{(2)}$ share the same unconditional mean.

Under these two constraints, the mortality improvement rates of the two populations have the same unconditional mean. The female and male mortality rates will eventually run in parallel. We follow Zhou *et al.* (2014) and estimate this model by maximum likelihood estimation.

Figure 4 plots the estimates of $\alpha_x^{(i)}$, $\beta_x^{(i)}$, and $\kappa_t^{(i)}$ when the non-divergence constraints are applied. We observe that $\alpha_x^{(i)}$ and the common $\beta_x^{(i)}$ follow similar patterns as those in Figure 2. Therefore, we continue to extrapolate $\alpha_x^{(i)}$ by linear function and $\beta_x^{(i)}$ by Gaussian or quadratic function. The new parameter estimates of the VAR(2) model for $[\Delta\kappa_t^{(1)}, \Delta\kappa_t^{(2)}]'$ are $A_0 = \begin{bmatrix} -0.9872 \\ -1.0340 \end{bmatrix}$, $A_1 = \begin{bmatrix} 1.1344 & -1.2043 \\ 0.9852 & -1.0241 \end{bmatrix}$, $A_2 = \begin{bmatrix} -0.7530 & 0.6210 \\ -1.4681 & 1.2480 \end{bmatrix}$, and $\Sigma = \begin{bmatrix} 0.4674 & 0.5397 \\ 0.5397 & 0.6963 \end{bmatrix}$.

Figure 5 shows the mean mortality forecasts of an 80-year-old Japanese male and an 80-year-old Japanese female over 40 years using the three dependence structures: independent (left), correlated

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

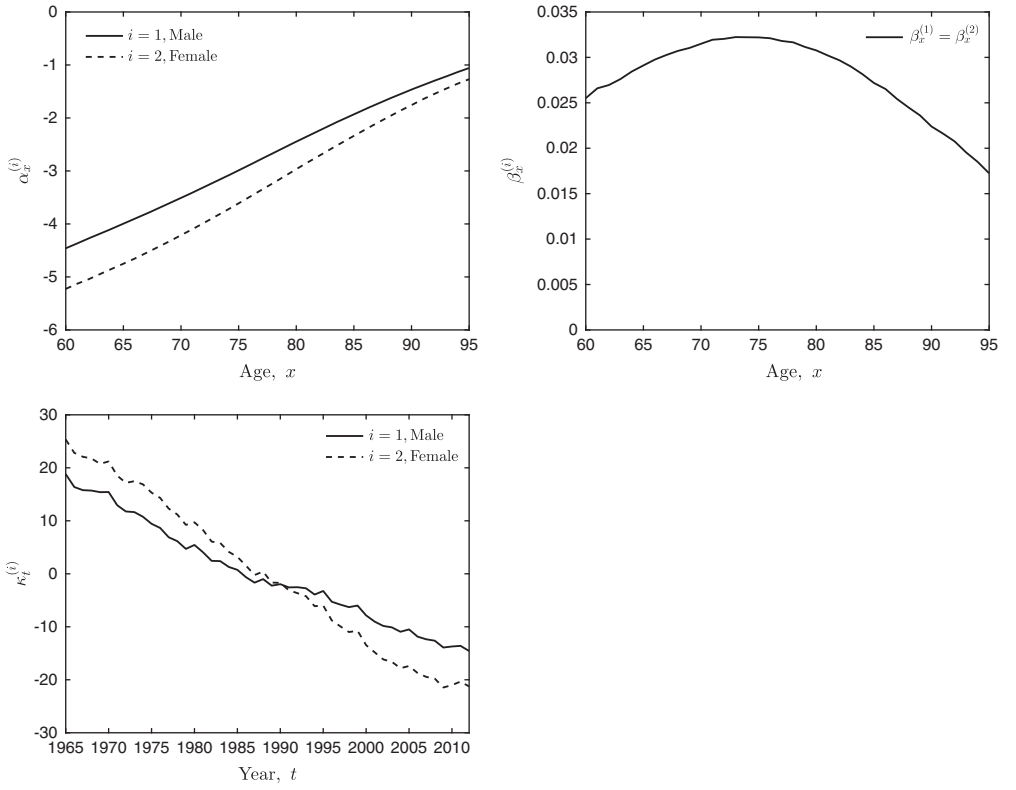

Figure 4. Estimates of $\alpha_x^{(i)}$, $\beta_x^{(i)}$, and $\kappa_t^{(i)}$ of the two-population Lee–Carter model with non-divergence constraints.

but allowing divergence (centre), and correlated with non-divergence constraints (right). The upper panel uses Gaussian extrapolation for β_x and the lower panel uses quadratic function with a floor of 0. Perks' model is used in the base mortality curve for illustration. When using Gaussian extrapolation for β_x, there is no obvious difference between the left and centre plots. Incorporating correlation without imposing non-divergence constraints leads to marginal changes in the mean forecasts of both male and female mortality. However, comparing the centre and right plots of the upper panel, we find that the non-divergence constraints shift male mortality significantly downward and female mortality slightly upward. This may have a major impact on the annuity values.

To understand the reasons behind this difference, we calculate the unconditional mean of $\Delta\kappa_t^{(1)}$ and $\Delta\kappa_t^{(2)}$ under all three structures, as presented in Table 4. When no constraints are imposed, $E(\Delta\kappa_t^{(1)})$ is significantly higher than $E(\Delta\kappa_t^{(2)})$, indicating slower mortality improvements for males. After imposing the non-divergence constraints, $E(\Delta\kappa_t^{(1)})$ decreases and $E(\Delta\kappa_t^{(2)})$ increases, such that they agree on the same value. With faster mortality improvement, male mortality forecasts become lower. The opposite occurs for female mortality forecasts.

When using quadratic extrapolation for β_x, the three plots in the lower panel of Figure 5 show no difference for ages above 105. In Figure 3, we have seen that β_x decreases quickly and reaches the floor of 0 at $x = 105$. Male and female mortality will not improve further after age 105. Therefore, the choice of correlation structure has no influence on the mortality forecasts for ages above 105.

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

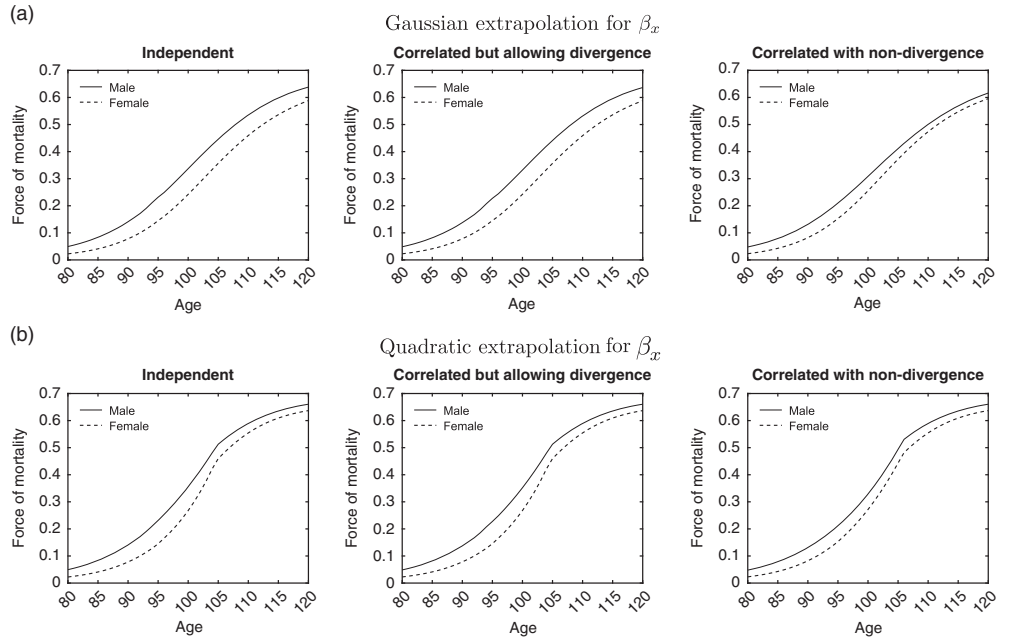

Figure 5. Force of mortality forecasts of an 80-year-old over the next 40 years using the three dependence structures, assuming Gaussian extrapolation (a) and quadratic extrapolation (b) for β_x, and the Perks' mortality curve.

Table 4. Unconditional mean of $\Delta\kappa_t^{(1)}$ and $\Delta\kappa_t^{(2)}$ under various correlation structures.

Correlation structures	$E(\Delta\kappa_t^{(1)})$	$E(\Delta\kappa_t^{(2)})$
Independent	−0.6794	−0.9521
Correlated but allowing divergence	−0.7128	−0.9731
Correlated with non-divergence	−0.8213	−0.8213

4. Model Risk

4.1. Measuring model risk

According to Draper (1995), a model formalises assumptions about how unknown quantities are related to known quantities. A model typically has two parts: structural assumption, such as the form of the link function and the choice of error distribution in a generalised linear model; and parameters, whose meaning is specific to a given choice of structural assumptions. The uncertainty of a model arises from two sources:

1. Uncertainty in the values of parameters given a specific model structure, i.e. parameter risk; and
2. uncertainty in the model structure itself, i.e. model risk.

In this section, we aim to quantify the exposure of annuity value to the second source of uncertainty. To simplify the problem and maintain our focus on model risk, we assume that annuity value is the

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

actuarial present value of all future benefit payments. Consider an n-year deferred annuity issued to a life aged x at the beginning of 2013. This annuity pays \$10,000 at the beginning of each year contingent on the survival of the annuitant. Its value at issue can be expressed as

$$ {}_{n|}\ddot{a}_x = \sum_{k=n}^{\omega-1} 10,000(1+i)^{-k}\, {}_kp_x $$

where ω is the limiting age, i the annual effective interest rate, and ${}_kp_x$ the probability that this life survives k years. When $n = 0$, this annuity becomes an immediate annuity. Assuming constant force of mortality between integral ages, ${}_kp_x$ can be approximated by

$$ {}_kp_x = \prod_{s=0}^{k-1} p_{x+s} = \prod_{s=0}^{k-1} e^{-m_{x+s,2012+(1+s)}} $$

Danielsson *et al.* (2015) proposed to use risk ratio to quantify model risk. The risk ratio measures the level of disagreement amongst candidate models by calculating the ratio of the maximum to the minimum risk forecasts. The difference between risk ratio and one indicates the extent to which the candidate models disagree. When risk forecasts come from similar models, then risk ratio should be close to one. In this paper, we use 95% value at risk (VaR) of annuity values as the risk forecasts. If we have N candidate models, then the risk ratio can be expressed as

$$ \text{Risk ratio} = \frac{\max_i \text{VaR}^i_{95\%}({}_{n|}\ddot{a}_x)}{\min_i \text{VaR}^i_{95\%}({}_{n|}\ddot{a}_x)} \tag{4.1} $$

where i denotes the ith model and $i = 1, 2, \ldots, N$. The mortality model we use for annuity valuation is comprised of two parts: mortality tail curve and mortality improvement rate model. We examine the model risk in these two parts separately. Since we do not know the exact distribution of annuity value, we use the 95th percentile of simulated annuity values as the 95% VaR.

4.2. Model risk in mortality tail curves

To start with, we assume there is no mortality improvement and examine how the annuity value changes with the choice of mortality tail curves. With zero mortality improvement rate, we have ${}_kp_x = \prod_{s=0}^{k-1} e^{-m_{x+s,2012}}$. Once the mortality curve in 2012 is determined, there is no uncertainty in ${}_kp_x$. Therefore, annuity value is a constant and the 95% VaR of annuity value is equal to this constant.

We use immediate annuities and deep-deferred annuities with various deferral periods for illustration. Assume that these annuities are purchased at age 75 and annual effective interest rate is 2%. Table 5 lists risk ratios and average annuity values using various mortality tail curves. Although we observed a large discrepancy of mortality rates at high ages in Figure 1, Table 5 shows that there is insignificant difference in immediate annuity values from using different mortality curves because the three mortality curves only start to diverge at approximately age 95. Immediate annuity values are heavily weighted by benefits paid in earlier years due to less discounting and higher survival probability. The actuarial present value of benefits paid after age 95 is relatively small. Therefore, the choice of mortality tail curve does not have a significant impact on the values of immediate annuities.

As expected, deep-deferred annuities are much less costly than immediate annuities. The value of a 15-year deferred annuity issued to a 75-year-old male is ~9.7% of the corresponding immediate annuity value because a 75-year-old male has only an ~54% chance of surviving to age 90, and the annuity payments at high ages are deeply discounted.

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

Table 5. Annuity values under various mortality tail curves and zero mortality improvement, $i = 2\%$.

	Annuity value			
Annuity	Gompertz	Cubic	Perks	Risk ratio
Males				
$\ddot{a}_{75}$	105,437	105,484	105,501	1.001
${}_{10\|}\ddot{a}_{75}$	28,870	28,597	28,614	1.010
${}_{15\|}\ddot{a}_{75}$	10,262	10,093	10,098	1.017
${}_{20\|}\ddot{a}_{75}$	2,260	2,310	2,349	1.040
Females				
$\ddot{a}_{75}$	133,481	133,598	133,637	1.001
${}_{10\|}\ddot{a}_{75}$	49,518	49,270	49,335	1.005
${}_{15\|}\ddot{a}_{75}$	22,732	22,339	22,407	1.018
${}_{20\|}\ddot{a}_{75}$	7,220	7,179	7,161	1.008

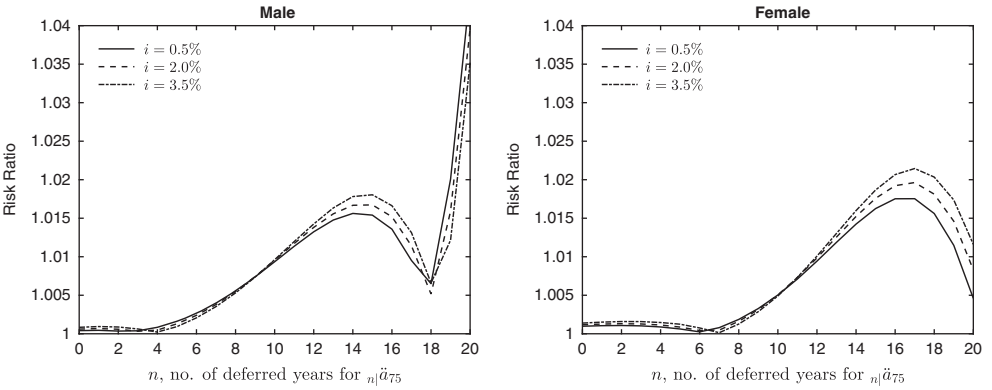

Figure 6. Risk ratios of ${}_{n|}\ddot{a}_{75}$ amongst the three mortality tail curves and zero mortality improvement rate.

Although deep-deferred annuities have a low-cost advantage, they have a higher model risk than immediate annuities. Figure 6 plots the risk ratios of ${}_{n|}\ddot{a}_{75}$ for $n = 0, 1, 2, \ldots, 20$ under various interest rate assumptions. While the focus of this discussion is on males, the female plot can be analysed in a similar fashion.

The pattern of risk ratio is determined largely by the relative impacts of mortality rates in two age ranges: 82–94 and above 95. As the deferral period increases, the impact of age range 82–94 decreases while that of ages above 95 increases due to increasing weights of payments in later life. Recall that in Figure 1, Gompertz law has the lowest force of mortality for males in the age range of 82–92 and the highest force of mortality for males who are age 95 and above among the three mortality curves. In contrast, Perks' model has the highest or close to the highest force of mortality for males in the age range of 82–92 and the lowest force of mortality for males who are age 95 and above.

When $n \leq 18$, the impact of age range 82–94 outweighs that of ages over 95 and causes Gompertz law to yield the highest annuity value. Because the gap between Gompertz and Perks' mortality

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

Table 6. Average annuity values under various mortality tail curves, assuming single-population Lee–Carter model for mortality improvement, Gaussian extrapolation for β_x and $i = 2\%$.

	Average annuity value			
Annuity	Gompertz	Cubic	Perks	Risk ratio
Males				
$\ddot{a}_{75}$	111,439	111,581	111,645	1.002
${}_{10\|}\ddot{a}_{75}$	34,086	33,912	33,977	1.004
${}_{15\|}\ddot{a}_{75}$	13,769	13,676	13,730	1.005
${}_{20\|}\ddot{a}_{75}$	3,730	3,856	3,945	1.061
Females				
$\ddot{a}_{75}$	143,867	144,304	144,385	1.004
${}_{10\|}\ddot{a}_{75}$	59,280	59,386	59,481	1.005
${}_{15\|}\ddot{a}_{75}$	30,668	30,621	30,720	1.005
${}_{20\|}\ddot{a}_{75}$	12,000	12,241	12,267	1.028

curves increases from age 82 to age 88, the risk ratio increases for $n < 15$. When $15 \le n \le 18$, the risk ratio decreases. The reasons are twofold: (1) the gap between Gompertz mortality curve and Perks' mortality curve shrinks from age 89 to age 92; (2) the impact of mortality at age 95 and above weighs more in annuity values as n increases. With n continuing to increase, the impact of mortality at age 95 and above eventually outweighs that of mortality in the age range of 82–94. This results in Perks' model having the highest annuity value for $n > 18$, and hence a kink point at $n = 18$ in the left plot of Figure 6. As the gap between Perks' and Gompertz mortality curves enlarges quickly for ages above 95, the risk ratio also increases steeply for $n > 18$.

The impact of interest rate also depends on the length of the deferral period. We observe that higher interest rate leads to higher risk ratio for $10 \le n \le 18$, but lower risk ratio for $n > 18$. When interest rate increases, the present value of benefit payment becomes lower. However, the present value of payments made in the age range of 82–94 is less affected than the value of payments made at age 95 and above due to discounting. Therefore, the impact of mortality in the age range of 82–94 on annuity value is amplified for deferred annuities with $10 \le n \le 18$, thus resulting in higher risk ratio. When $n > 18$, annuity value is mostly determined by the mortality at age 95 and above. With a lower interest rate, payments made at higher ages when a larger discrepancy of mortality rate is observed between Gompertz and Perks' mortality curve gain more weight in annuity value. The risk ratio increases as a result.

Next we incorporate mortality improvement rates and further study the model risk from mortality tail curves. We first use the single-population Lee–Carter model to obtain 1,000 trajectories of future mortality improvement rates. We then multiply the improvement rates by the base year mortality curve to calculate simulated trajectories of mortality rates. Again, we consider the three types of mortality tail curves described in section 2. Simulated annuity values are determined for each combination of mortality improvement trajectory and base mortality curve. Table 6 lists the average simulated annuity value using various mortality tail curves and the risk ratios assuming Gaussian extrapolation for β_x and $i = 2\%$.

After incorporating mortality improvement trend, the probability of a 75 years old surviving to advanced ages increases. As a result, annuity prices increase significantly compared to those in

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

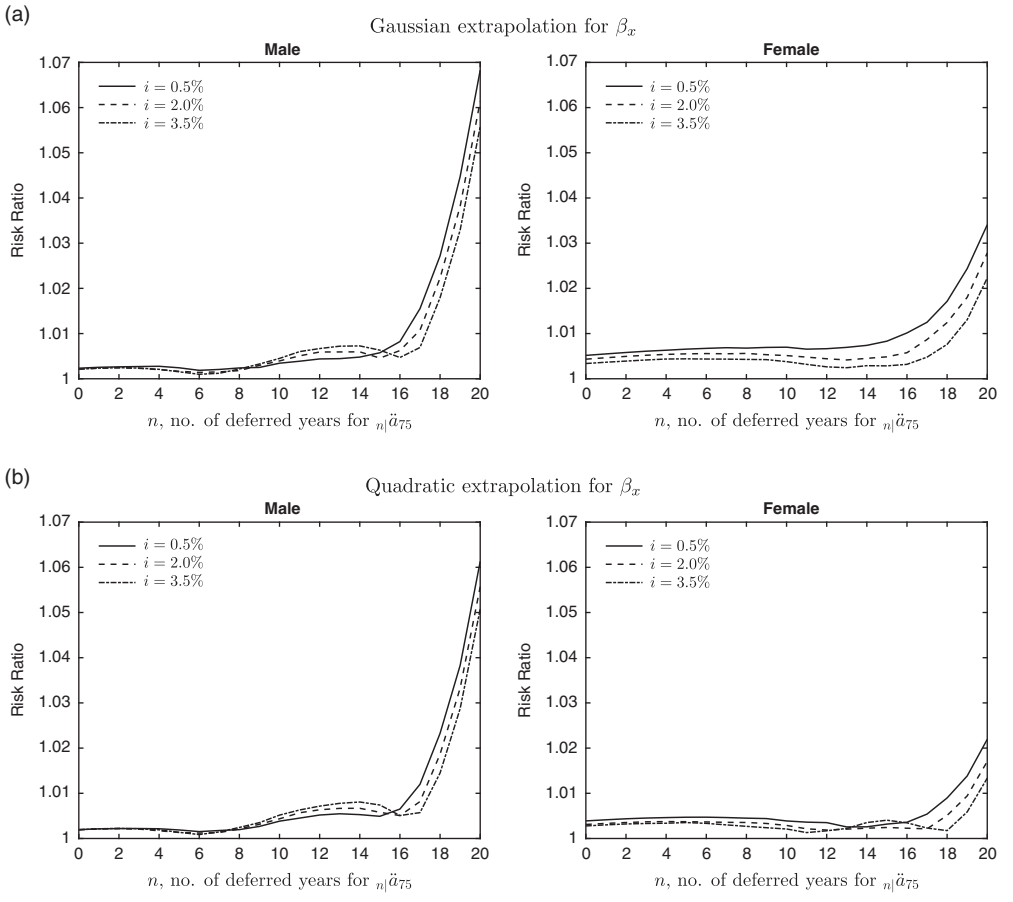

Figure 7. Risk ratios of ${}_{n|}\ddot{a}_{75}$ amongst the three mortality tail curves, assuming single-population Lee–Carter model for mortality improvement, Gaussian extrapolation for β_x (a) and quadratic extrapolation for β_x (b).

Table 5, especially for annuities with a longer deferral period. The risk ratio for 20-year deferred annuity becomes higher. The value of a 20-year deferred annuity is largely driven by mortalities at ages above 95. With higher surviving probabilities, the actuarial present value of payments made at high ages gains more weight in annuity value. The large discrepancy between different mortality curves at age 95 and above exerts more impacts, thereby resulting in higher risk ratios. In contrast, the values of 10-year and 15-year deferred annuities are mostly determined by mortalities at ages below 95, where the discrepancy between mortality curves is small. Hence, their risk ratios are also small.

In Figure 7, we plot the relation between risk ratios amongst mortality tail curves and length of deferral period assuming various interest rates and extrapolation methods for β_x. Due to the reasons explained above, risk ratios increase sharply when the deferral period is >17. We also observe that risk ratios using Gaussian extrapolation for β_x are slightly higher than those using quadratic extrapolation. Gaussian extrapolation admits more mortality improvement for ages above 95 and thus increases the weights of payments at these ages in the annuity value. Since the discrepancy

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

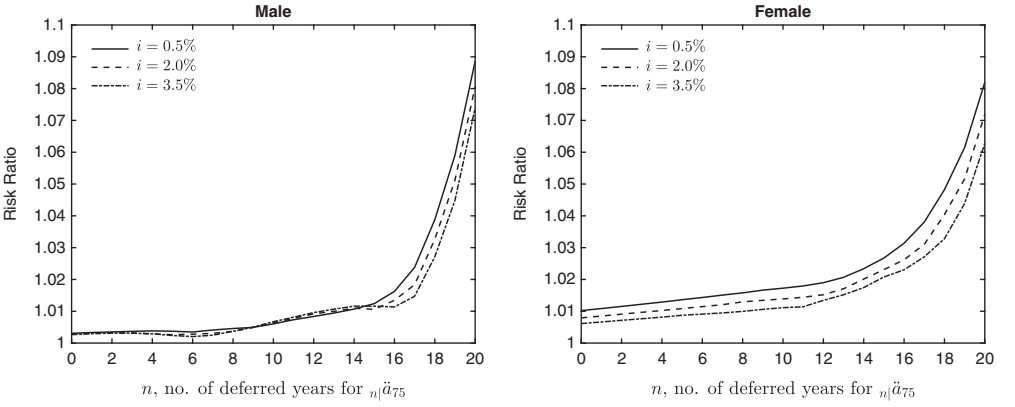

Figure 8. Risk ratios of ${}_{n|}\ddot{a}_{75}$ amongst six combinations of mortality tail curve and extrapolation method for β_x, assuming single-population Lee–Carter model for mortality improvement.

between mortality curves for these ages is particularly large, higher weights of payments at these ages lead to higher risk ratios.

Overall, the model risk from mortality tail curve appears to be low for annuities issued to a 75-year-old male with a deferral period of less than 17 years. When the deferral period is longer than 17 years, the model risk increases sharply. In practice, deferred annuities often have payments starting no later than age 85. Therefore, model risk from mortality tail curve should not be a major concern.

Figure 8 examines the total model risk from the two sources: mortality tail curve and extrapolation method for β_x. The risk ratios are based on six combinations of mortality tail curve and extrapolation method for β_x, since we have three mortality tail curves and two extrapolation methods. Since extrapolation method for β_x is part of the mortality improvement model, the effects of adding mortality improvement are twofold: (1) amplifying the model risk from mortality tail curves as shown in Table 6; and (2) introducing model risk from extrapolation method for β_x. A comparison of Figures 7 and 8 indicates that male annuity values are subject to more model risk from mortality tail curve while female annuity values are subject to more model risk from extrapolation method for β_x.

4.3. Model risk in two-population Lee–Carter model

In this section, we study model risk in the two-population Lee–Carter model. We consider three different assumptions about the dependence structure between male and female mortality improvement:

- independent;
- correlated but allowed to diverge over the long run;
- correlated and non-divergent over the long run.

Table 7 lists risk ratios and average annuity values using various dependence structures assuming Gaussian extrapolation for β_x, Perks' mortality curve, and 2% annual effective interest rate. The difference in average annuity values under these three structures can be explained by the mortality forecasts in Figure 5. When imposing the non-divergence constraints with Gaussian extrapolation for β_x, male mortality shifts down significantly and female mortality shifts up slightly. This leads to

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

Table 7. Average annuity values using various dependence structures assuming Gaussian extrapolation method for β_x, Perks' mortality curve, and $i = 2\%$.

	Average annuity value			
Annuity	Independent	Allowing divergence	Non-diverging	Risk ratio
Males				
$\ddot{a}_{75}$	111,645	112,512	113,852	1.028
${}_{10\|}\ddot{a}_{75}$	33,977	34,697	35,938	1.081
${}_{15\|}\ddot{a}_{75}$	13,730	14,200	15,163	1.143
${}_{20\|}\ddot{a}_{75}$	3,945	4,151	4,710	1.268
Females				
$\ddot{a}_{75}$	144,385	144,191	142,546	1.003
${}_{10\|}\ddot{a}_{75}$	59,481	59,305	57,718	1.006
${}_{15\|}\ddot{a}_{75}$	30,720	30,594	29,241	1.015
${}_{20\|}\ddot{a}_{75}$	12,267	12,203	11,313	1.029

higher male annuity values and lower female annuity values. A comparison of Tables 5–7 indicates that model risk caused by two-population Lee–Carter model is much more significant than that caused by mortality for male annuities. For female annuities, the model risk from all three sources is low.

Figure 9 plots the risk ratios of ${}_{n|}\ddot{a}_{75}$, for $n = 0, 1, \ldots, 20$, among the three dependence structures assuming Gaussian extrapolation for β_x and Perks' mortality curve. Annuities with a longer deferral period have cash flows concentrated at higher ages. Since the difference of mortality forecasts from using the three structures increases with the length of the forecasting period, a longer deferral period results in higher risk ratios. Gaussian extrapolation for β_x allows for more mortality improvement and heavier cash flows at advanced ages than quadratic extrapolation. It amplifies the impact of dependence structures on annuity values, thereby resulting in higher risk ratios than quadratic extrapolation.

Figure 10 studies how risk ratios change when we gradually add the three sources of model risk. Annual effective interest rate is assumed to be 2%. The left panel shows that the model risk of male annuity values is high, and it is primarily attributable to dependence structure. The model risk from extrapolation method for β_x and mortality curve is almost negligible for annuities with deferral period of less than 10 years; however, this risk becomes more significant when the deferral period increases. The right panel shows that the model risk of female annuity values mainly comes from dependence structure and extrapolation method for β_x, and this risk is much lower than the model risk of male annuity values. Overall, the model risk from the two-population mortality improvement model is much more significant than that from mortality tail curve.

5. Parameter Risk

5.1. Incorporating parameter risk

According to Draper (1995), a major source of uncertainty is parameter risk. The values of parameters given a specific model structure have uncertainty because only a finite set of data are used to estimate these parameters and the estimates cannot be exact. Due to the sparseness of high age mortality data, parameter risk may be more severe for deep-deferred annuities.

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

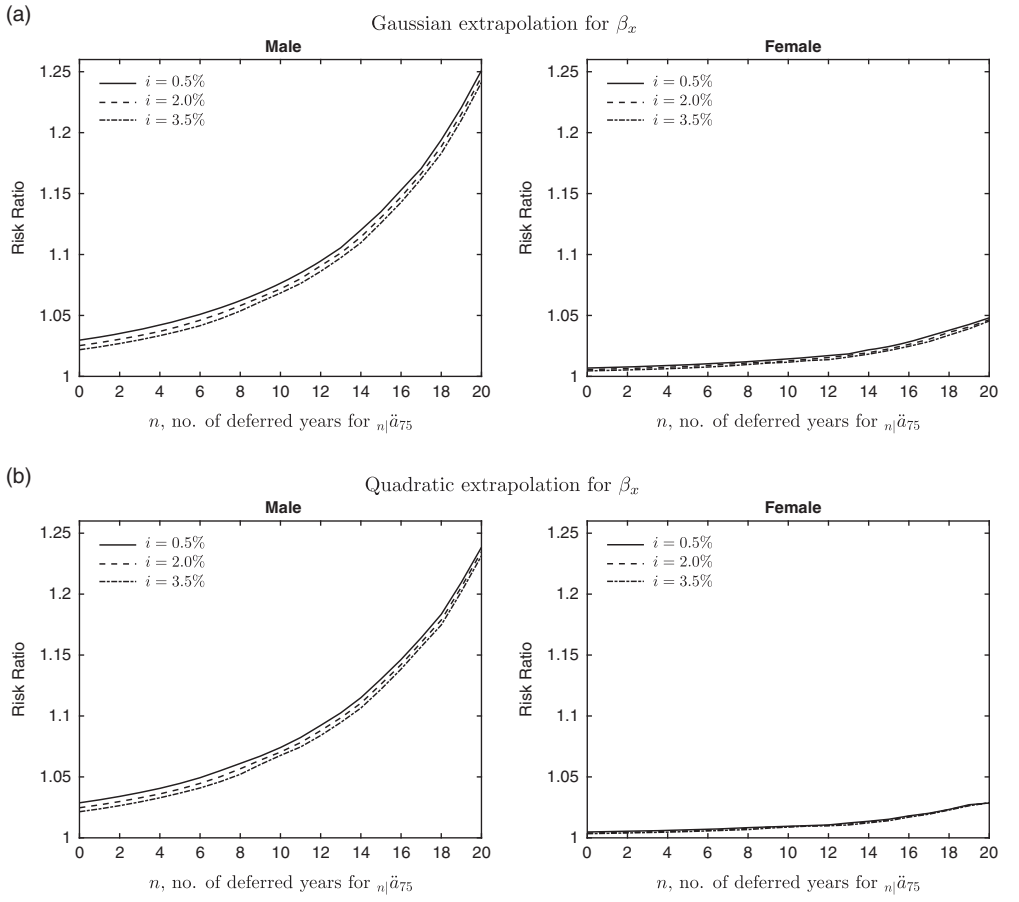

Figure 9. Risk ratios of ${}_{n|}\ddot{a}_{75}$ amongst various correlation structures with Perks' mortality curve, Gaussian extrapolation for β_x (a) and quadratic extrapolation for β_x (b).

To quantify parameter uncertainty in mortality tail curve and mortality improvement model, we adopt the parametric bootstrap method. To apply bootstrap for mortality tail curve, we follow Brouhns *et al.* (2005) and bootstrap death counts by applying a Poisson noise to the observed numbers of deaths as follows:

1. Generate one bootstrap sample of death counts in year 2012, assuming that the bootstrapped death counts are from Poisson distributions with mean equal to the observed number of deaths.
2. Estimate the three mortality tail curves by using the bootstrap sample.
3. Construct base mortality curves using each estimated mortality tail curve.

Since we use a two-stage method to fit the Lee–Carter model, parameter risk comes from both stages and bootstrapping should be performed for both stages as well. In the first stage, we estimate the Lee–Carter model and obtain estimates of $\alpha_x^{(i)}$, $\beta_x^{(i)}$, and $\kappa_t^{(i)}$. To capture the parameter risk in this stage, we again follow Brouhns *et al.* (2005) and bootstrap death counts by applying a Poisson noise to the observed numbers of deaths. In the second stage, we estimate the extrapolation function using

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

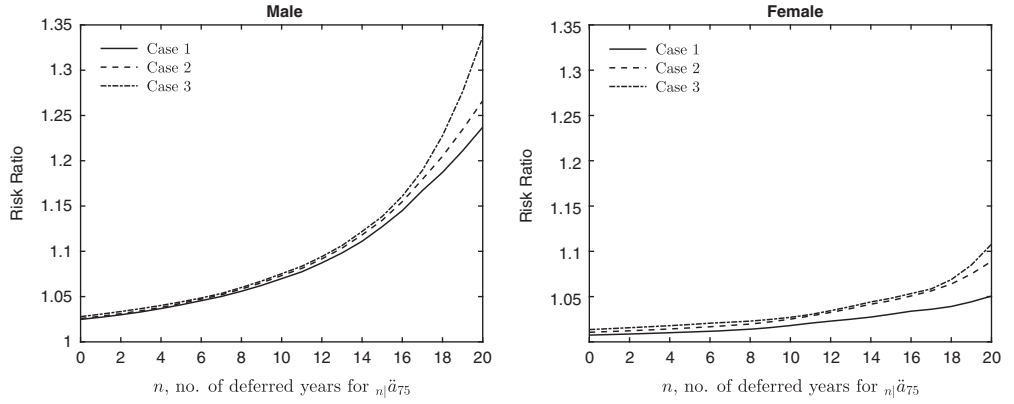

Figure 10. Comparison of risk ratios of $_{n|}\ddot{a}_{75}$ in three cases: Case 1: risk ratios amongst dependence structures, assuming Gaussian extrapolation for β_x and Perks' mortality curve; Case 2: risk ratios amongst six combinations of dependence structure and extrapolation method for β_x, assuming Perks' mortality curve; and Case 3: risk ratios amongst 18 combinations of dependence structure, extrapolation method for β_x, and mortality tail curve. $i = 2\%$.

the estimates of $\alpha_x^{(i)}$ and $\beta_x^{(i)}$, and estimate the time series model using the estimates of $\kappa_t^{(i)}$ from the first stage. To take into account the parameter risk in this stage, we apply residual bootstrap.

The detailed procedures to apply bootstrap in the Lee–Carter model are as follows:

1. Generate one bootstrap sample of death counts for the sample period of 1965–2012, assuming that the bootstrapped death counts are from Poisson distributions with mean equal to the observed number of deaths.
2. Estimate the mortality improvement rate model using the bootstrap sample and obtain estimates of $\alpha_x^{(i)}$, $\beta_x^{(i)}$, and $\kappa_t^{(i)}$.
3. Estimate the extrapolation functions for $\alpha_x^{(i)}$ and $\beta_x^{(i)}$.
4. Generate one bootstrap sample of $\alpha_x^{(i)}$ and $\beta_x^{(i)}$ using residual bootstrap[5] and re-estimate the extrapolation functions by using this sample.
5. Extrapolate $\alpha_x^{(i)}$ and $\beta_x^{(i)}$ to desired age using the re-estimated function.
6. Estimate the time series model for $\kappa_t^{(i)}$.
7. Generate one bootstrap sample of $\kappa_t^{(i)}$ using residual bootstrap[6] and re-estimate the time series model by using this sample.
8. Simulate one path of $\kappa_t^{(i)}$, for $t = 2013, 2014, \ldots$, using the re-estimated time series model.
9. Calculate the path of future mortality improvement rates based on extrapolated $\alpha_x^{(i)}$ and $\beta_x^{(i)}$, and simulated path of $\kappa_t^{(i)}$.

[5] We re-sample with replacement the residuals from the initial estimation in step (3); and then generate the bootstrap sample by adding the re-sampled residuals to the fitted values of $\alpha_x^{(i)}$ and $\beta_x^{(i)}$ in step (3).

[6] We re-sample with replacement the residuals from the initial estimation in step 6); and then generate the bootstrap sample by adding the re-sampled residuals to the fitted values of $\kappa_t^{(i)}$ in step (6).

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

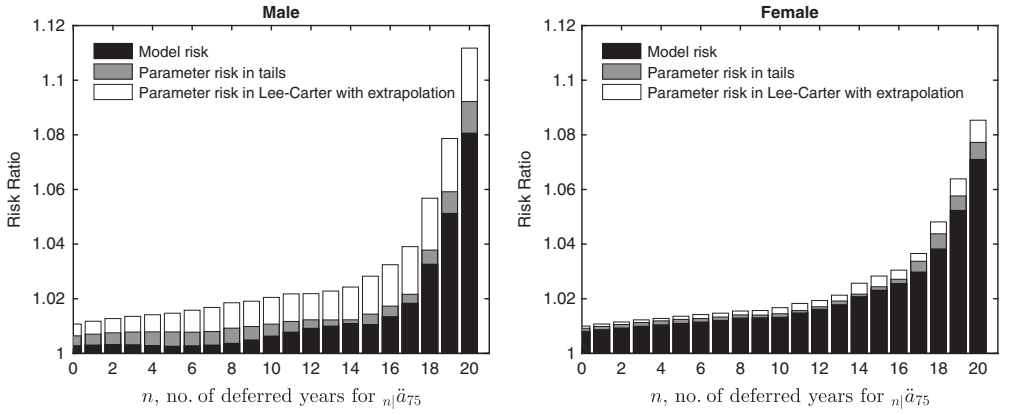

Figure 11. Risk ratios of $_{n|}\ddot{a}_{75}$ amongst six combinations of mortality tail curve and extrapolation method for β_x, with parameter uncertainty incorporated. $i = 2\%$.

After obtaining the base mortality curve and future mortality improvement rates, we determine the path of future mortality rates and calculate a simulated annuity value. These procedures are repeated 1,000 times to obtain 1,000 simulated annuity values. The new risk ratio is then calculated as follows:

$$\text{New risk ratio} = \frac{\max_i \text{VaR}^i_{95\%}\left({}_{n|}\ddot{a}_x\right) \text{ with bootstrapping}}{\min_i \text{VaR}^i_{95\%}\left({}_{n|}\ddot{a}_x\right) \text{ without bootstrapping}} \tag{5.1}$$

where $i = 1, 2, \ldots, N$ and N is the number of models investigated. The extent of parameter risk is indicated by the difference between the new risk ratio and the risk ratio without considering parameter uncertainty as defined in equation (4.1).

5.2. Parameter risk in single-population mortality model

A full single-population mortality model includes a base mortality curve and a single-population mortality improvement model. To examine its parameter risk, we follow the bootstrap procedures described in section 5.1 and obtain simulated annuity values for each combination of mortality tail curve and extrapolation method for β_x. We then determine $\text{VaR}_{95\%}\left({}_{n|}\ddot{a}_x\right)$ for all six models considered and use the highest one as the numerator in equation (5.1) to calculate the new risk ratio.

Figure 11 plots the risk ratios of $_{n|}\ddot{a}_{75}$, for $n = 0, 1, 2, \ldots, 20$, amongst six combinations of mortality tail curve and extrapolation method for β_x. We consider three cases: no bootstrapping, partial bootstrapping, and full bootstrapping. The dark lower part of each bar displays the risk ratio without considering parameter uncertainty. It represents the model risk from mortality tail curve and extrapolation method for β_x used in the single-population Lee–Carter model.

In partial bootstrapping, we apply bootstrap only to the mortality tail curve, and not to the mortality improvement model. Partial bootstrapping increases risk ratios by adding parameter uncertainty from mortality tail curve. The grey part of each bar depicts the increased risk, i.e. parameter risk from mortality tail curve.

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

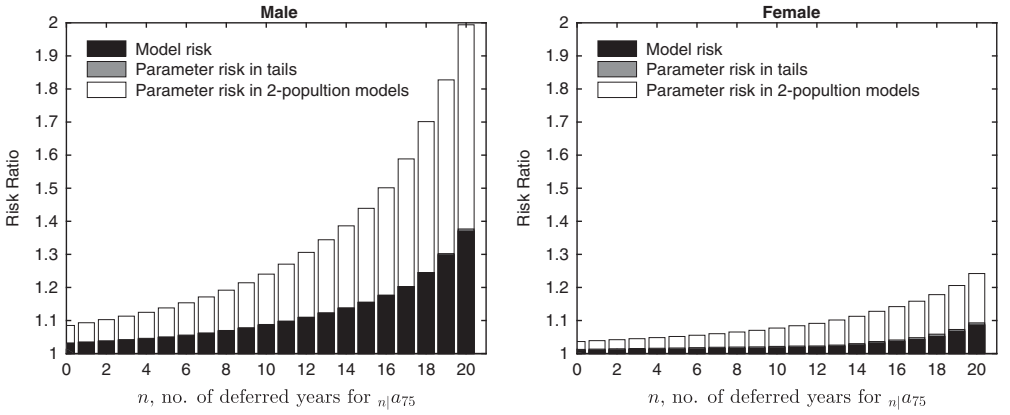

Figure 12. Risk ratios of $_{n|}\ddot{a}_{75}$ amongst 18 combinations of mortality tail curve, extrapolation method for β_x, and dependence structure, with parameter uncertainty incorporated. $i = 2\%$.

In full bootstrapping, we apply bootstrap to both the mortality tail curve and the mortality improvement model. Full bootstrapping further increases risk ratios by adding parameter uncertainty from the mortality improvement rate model. The upper white part of each bar represents the increased risk, i.e. parameter risk from the mortality improvement model.

For male annuity values, both model risk and parameter risk play important roles. However, parameter risk from mortality tail curve is much smaller than that from mortality improvement rate model. For female annuity values, model risk is of more concern, while parameter risk appears to be low.

5.3. Parameter risk in two-population mortality models

In section 4.3, we examined model risk from the dependence structure in the two-population Lee–Carter model. We found that the model risk is significant, especially for males, and this risk increases steadily with the length of deferral period. In this section, we study the parameter risk in the two-population mortality model and compare it with the model risk.

We consider 18 combinations of mortality tail curve, extrapolation method for β_x, and dependence structure. For each combination, we use the bootstrapping method described in section 5.1 to simulate 1,000 annuity values and calculate VaR $\left(_{n|}\ddot{a}_{75}\right)$. The highest VaR $\left(_{n|}\ddot{a}_{75}\right)$ is used to calculate the new risk ratio.

Figure 12 demonstrates the risk ratios of $_{n|}\ddot{a}_{75}$ amongst the 18 combinations of mortality tail curve, extrapolation method for β_x, and dependence structure. Three levels of bootstrapping are considered: no bootstrapping, partial bootstrapping, and full bootstrapping. In partial bootstrapping, we again apply bootstrap only to the mortality tail curve, and not to the mortality improvement model. The dark lower part of each bar depicts the risk ratios without parameter uncertainty. It measures the variation of VaR $\left(_{n|}\ddot{a}_{75}\right)$ stemming from model risk. The grey part is almost invisible indicating that parameter risk from mortality tail curve is negligible. The upper white part represents the increase of risk ratio due to parameter uncertainty in the two-population mortality improvement model, which turns out to be the most significant source of risk in both male and female annuity values. The overall risk in male annuity values is much higher than that in female annuity values.

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

6. Conclusion

As deep-deferred annuity is allowed in 401(K) plan, there will be more interest in purchasing and providing the annuity. In order to develop effective risk management solutions for this annuity, it is necessary to understand its risk profile. In this paper, we focus on the model risk and parameter risk in valuing deep-deferred annuities. We use risk ratio to measure the disagreement between different models. Higher risk ratio indicates higher model uncertainty in annuity valuation. This measure can be used for both model risk and parameter risk, and thus allows us to compare the magnitude of the two types of risks.

The mortality model we use for high ages is comprised of two parts: a base mortality curve and a mortality improvement model. This mortality model has the flexibility of adapting different assumptions of mortality tail and mortality improvement patterns, and thus allows us to decompose the demographic risk of deep-deferred annuity valuation into four components: model risk in mortality tail curve, model risk in mortality improvement model, parameter risk in mortality tail curve, and parameter risk in mortality improvement model.

In the single-population mortality model, model risk comes from the mortality tail curve and the extrapolation method for β_x. Our analysis shows that the model risk of male annuity values is mainly from the mortality tail curve, while the model risk of female annuity values is mainly from the extrapolation method for β_x. Bootstrap method is used to incorporate parameter risk. Both model risk and parameter risk play significant roles in male annuity values and most of the parameter risk comes from the Lee–Carter model. In contrast, model risk dominates in female annuity values.

When we factor in dependence between male and female mortality, model risk increases sharply for male annuity values but only marginally for female annuity values. Parameter risk becomes much more significant than model risk for both genders. Parameter risk is primarily attributed to the two-population Lee–Carter model and parameter risk from mortality tail curves is trivial. The overall risk in female annuity value is much smaller than that in male annuity values.

Immediate annuity has very low model risk and parameter risk compared to deep-deferred annuity. Since high age cash flows have low weights in immediate annuity values, the large discrepancy from using different mortality tail curves and mortality improvement assumptions has marginal impact on immediate annuity values. In contrast, deep-deferred annuities have cash flows concentrated at high ages and are thus more sensitive to these assumptions.

In Appendix A, we directly model the mortality improvement rates following Mitchell *et al.* (2013). The findings using the Mitchell *et al.* (2013) model are similar to those from using the Lee–Carter model. We should point out that we only consider a limited number of models in this paper and thus may underestimate model risk. It would therefore be warranted to examine more model structures in the future, such as M1–M8 models described in Cairns *et al.* (2009) and the two-population models in Li *et al.* (2015*b*), such that model risk is studied more comprehensively.

In the absence of an effective hedging instrument, the alarmingly high uncertainty in deep-deferred annuity valuation will lead to high reserve requirement and high cost of providing the annuity. In order to promote provisions of deep-deferred annuity, it is crucial to develop solutions that can provide effective longevity risk hedge. A possible venue of future work is to examine existing capital market solutions and study their hedging effectiveness in consideration of model risk and parameter risk.

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

Acknowledgements

The authors acknowledge the financial support from the Natural Science and Engineering Research Council of Canada. The authors would also like to thank participants at the Eleventh International Longevity Risk and Capital Markets Solutions Conference for their stimulating discussions in an earlier version of this paper.

References

Arai, Y., Inagaki, H., Takayama, M., Abe, Y., Saito, Y., Takebayashi, T. & Hirose, N. (2014). Physical independence and mortality at the extreme limit of life span: supercentenarians study in Japan. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, **69**(4), 486–494.

Beard, R.E. (1963). A theory of mortality based on actuarial, biological, and medical considerations. *Proceedings of International Population Conference, New York*, **1**, 611–625.

Beard, R.E. (1971). Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes. In W. Brass, Ed. *Biological aspects of demography* (pp. 57–68). Taylor & Francis, London.

Blake, D. & Turner, J.A. (2014). Longevity insurance annuities: lessons from the United Kingdom. *Benefits Quarterly*, **30**(1), 39–47.

Brouhns, N., Denuit, M. & Van Keilegom, I. (2005). Bootstrapping the Poisson log-bilinear model for mortality forecasting. *Scandinavian Actuarial Journal*, **2005**(3), 212–224.

Brouhns, N., Denuit, M. & Vermunt, J.K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, **31**(2002), 373–393.

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

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

Danielsson, J., James, K.R., Valenzuela, M. & Zer, I. (2015). Model risk of risk models. *Journal of Financial Stability*, forthcoming, https://doi.org/10.2139/ssrn.2425689.

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

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**(1), 45–97.

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

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

Gong, G. & Webb, A. (2010). Evaluating the advanced life deferred annuity – an annuity people might actually buy. *Insurance: Mathematics and Economics*, **46**(1), 210–221.

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

Haberman, S. & Renshaw, A. (2012). Parametric mortality improvement rate modelling and projecting. *Insurance: Mathematics and Economics*, **50**(3), 309–333.

Kannisto, V. (1992). *Presentation at a Workshop on Old-Age Mortality*. Odense University, Odense, Denmark.

Lee, R.D. & Carter, L.R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, **87**(419), 659–671.

Li, J.S.H. & Hardy, M.R. (2011). Measuring basis risk in longevity hedges. *North American Actuarial Journal*, **15**(2), 177–200.

Li, J.S.H., Chan, W.S. & Zhou, R. (2015*a*). Semi-coherent multi-population mortality modeling: the impact on longevity risk securitization. *Journal of Risk and Insurance* (in press).

Li, J.S.H., Hardy, M.R. & Tan, K.S. (2009). Uncertainty in mortality forecasting: an extension to the classical Lee-Carter approach. *Astin Bulletin*, **39**(1), 137–164.

Li, J.S.H., Zhou, R. & Hardy, M. (2015*b*). A step-by-step guide to building two-population stochastic mortality models. *Insurance: Mathematics and Economics*, **65**, 121–134.

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

Milevsky, M. (2005). Real longevity insurance with a deductible: an introduction to advanced-life delayed annuities (ALDA). *North American Actuarial Journal*, **9**(4), 109–122.

Mitchell, D., Brockett, P., Mendoza-Arriaga, R. & Muthuraman, K. (2013). Modeling and forecasting mortality rates. *Insurance: Mathematics and Economics*, **52**(2), 275–285.

Mitchell, O.S. & McCarthy, D. (2002). Estimating international adverse selection in annuities. *North American Actuarial Journal*, **6**(4), 38–54.

Nusbaum, T.J., Mueller, L.D. & Rose, M.R. (1996). Evolutionary patterns among measures of aging. *Experimental Gerontology*, **31**(4), 507–516.

Office for National Statistics (2013). National population projections, 2012-based statistical bulletin. Available online at the address http://www.ons.gov.uk/ons/dcp171778_334975.pdf. Accessed on February 1st, 2016.

Oeppen, J. & Vaupel, J.W. (2002). Broken limits to life expectancy. *Science*, **296**(5570), 1029–1031.

Panjer, H.H. & Russo, G. (1992). Parametric graduation of Canadian individual insurance mortality experience: 1982-1988. *Proceedings of the Canadian Institute of Actuaries*, **23**, 378–449.

Panjer, H.H. & Tan, K.S. (1995). *Graduation of Canadian Individual Insurance Mortality Experience: 1986-1992*. Canadian Institute of Actuaries, Retrieved from http://www.cia-ica.ca/docs/default-source/1995/9529e.pdf.

Perks, W. (1932). On some experiments on the graduation of tail curvetatistics. *Journal of the Institute of Actuaries*, **63**, 12–40.

Renshaw, A. & Haberman, S. (2003). Lee-Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections. *Applied Statistics*, **52**(1), 119–137.

Robine, J.M., Cheung, S.L.K., Saito, Y., Jeune, B., Parker, M.G. & Herrmann, F.R. (2010). Centenarians today: new insights on selection from the 5-COOP study. *Current Gerontology and Geriatrics Research*, 2010, Retrieved from http://dx.doi.org/10.1155/2010/120354.

Robine, J.M. & Saito, Y. (2003). Survival beyond age 100: the case of Japan. *Population and Development Review*, **29**, 208–228.

Weitz, J.S. & Fraser, H.B. (2001). Explaining mortality rate plateaus. *Proceedings of the National Academy of Sciences*, **98**(26), 15383–15386.

Willets, R.C. (2004). The Cohort Effect: Insights and Explanations. *British Actuarial Journal*, **10**(4), 833–877.

Wilson, C. (2001). On the scale of global demographic convergence 1950-2000. *Population and Development Review*, **27**(1), 593–600.

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

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

Appendix A. Mitchell *et al.* (2013) model

Mitchell *et al.* (2013) directly models mortality change rates as follows:

$$r_{x,t} = \alpha_x + \beta_x \kappa_t$$

where $r_{x,t} = \ln\left(\frac{m_{x,t}}{m_{x,t-1}}\right)$, α_x is the average change in log mortality rate for age x, κ_t the mortality change index, and β_x the age-specific sensitivity to the mortality change index. Single value decomposition is used to estimate this model. To ensure parameter uniqueness, we let α_x to be the average change in log mortality rates for age x over the sample period, i.e. $\alpha_x = \frac{\sum_{t=1965}^{2012} r_{x,t}}{48}$, and also impose an identification constraint: $\sum_x \beta_x = 1$. κ_t is further modelled by an autoregressive moving average model. Figure A.1 demonstrates the estimated values of α_x, β_x, and κ_t for males and females, respectively. We choose AR(2) model for κ_t based on BIC for model selection. Since the constant term of the estimated model is not significant, we omit the constant term. The AR(2) model can be expressed as follows:

$$\kappa_t = a_1 \kappa_{t-1} + a_2 \kappa_{t-2} + z_t$$

where $\{z_t\}$ are independently and identically distributed normal random variables with mean 0 and variance σ^2. Parameter estimates of this model and their standard errors (in parentheses) are shown in Table A.1.

Since the sample age range of the data is 60–95, we can only obtain the estimates of α_x and β_x for $x = 60, 61, \ldots, 95$. For ages above 95, α_x and β_x cannot be directly obtained. We model α_x and β_x as deterministic functions of age x, and then extrapolate them to higher ages. Figure A.1 shows that α_x is slightly concave upward, and β_x appears to increase linearly. We simply use a linear function of x to model β_x. Since α_x represents the average log mortality improvement rate at age x and mortality is expected to improve over time, we assume that α_x is non-positive. With this constraint, we consider two models for α_x: a Gaussian function with an asymptote of 0, and a quadratic function with a cap of 0.

To build a two-population mortality improvement rate model, we use one Mitchell *et al.* (2013) model for each population. $[\kappa_t^{(1)}, \kappa_t^{(2)}]'$ is further modelled by a pth order VAR process. We select VAR(2) model based on BIC values. When estimating the model, we also find that the constant term is not significant. Therefore, we omit the constant term in the VAR(2) model.

To incorporate the non-divergence assumption and study how it affects deep-deferred annuity valuation, we consider the following constraints:

1. $\alpha_x^{(1)} = \alpha_x^{(2)}$;
2. $\beta_x^{(1)} = \beta_x^{(2)}$;
3. $\kappa_x^{(1)}$ and $\kappa_x^{(2)}$ share the same long-term unconditional mean.

Under these constraints, the mortality improvement rates of males and females have the same unconditional mean. Since the constant term of the VAR(2) model is set to 0, both $\kappa_x^{(1)}$ and $\kappa_x^{(2)}$ have zero unconditional mean and thus the third constraint is satisfied automatically.

Future mortality rates are then generated by multiplying the base mortality curve with the mortality improvement rates using the Mitchell *et al.* (2013) model. To study the model risk, we consider three mortality tail curves, two extrapolation methods for $\alpha_x^{(i)}$, and three dependence structures for males and females. Figure A.2 shows how the risk ratios change when we gradually add the three sources of model

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

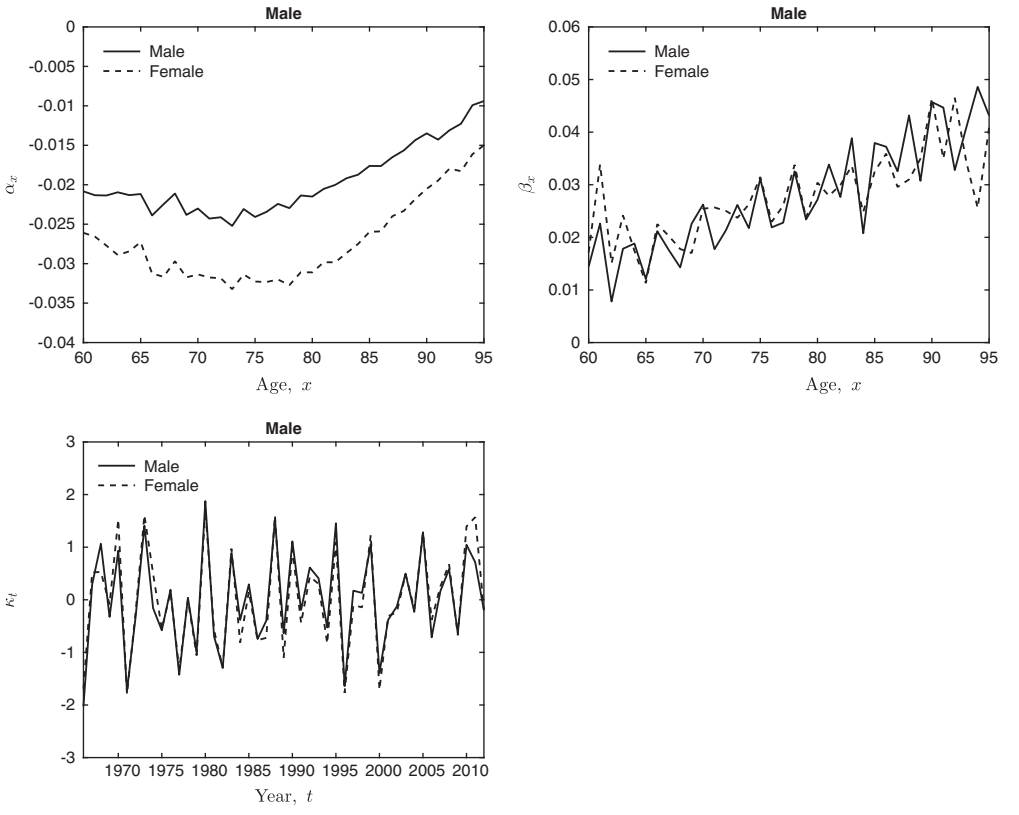

Figure A.1. Estimates of α_x, β_x, and κ_t for the Mitchell *et al.* (2013) model.

Table A1. Parameter estimates for κ_t in the single-population Mitchell *et al.* (2013) model.

Gender	a_1	a_2	σ
Male	−0.662853 (0.108711)	−0.630002 (0.0815341)	1.53245 (0.205336)
Female	−0.505672 (0.0989214)	−0.537079 (0.0930713)	1.59606 (0.26323)

risk. The model risk from the dependence structure in the two-population Mitchell model dominates in male annuity values. The model risk from dependence structure and extrapolation method for $\alpha_x^{(i)}$ is more significant than that from mortality tail curve for female annuity values. The overall model risk of female annuity values is much smaller than that of male annuity values. These findings are similar with what we observe in Figure 10 from using the Lee–Carter model for mortality improvement rates.

In Figure A.3, we plot the risk ratios of ${}_{n|}\ddot{a}_{75}$ amongst 18 combinations of morality tail curve, extrapolation method for $\alpha_x^{(i)}$, and dependence structure in the two-population Mitchell model. We consider three cases of bootstrapping: no bootstrapping, partial bootstrapping, and full bootstrapping. In partial bootstrapping, we apply bootstrap only to the mortality tail curve, and not to the two-population Mitchell model. The grey part is invisible indicating that parameter risk from mortality tail curve is negligible. The parameter risk in the two-population Mitchell model,

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press

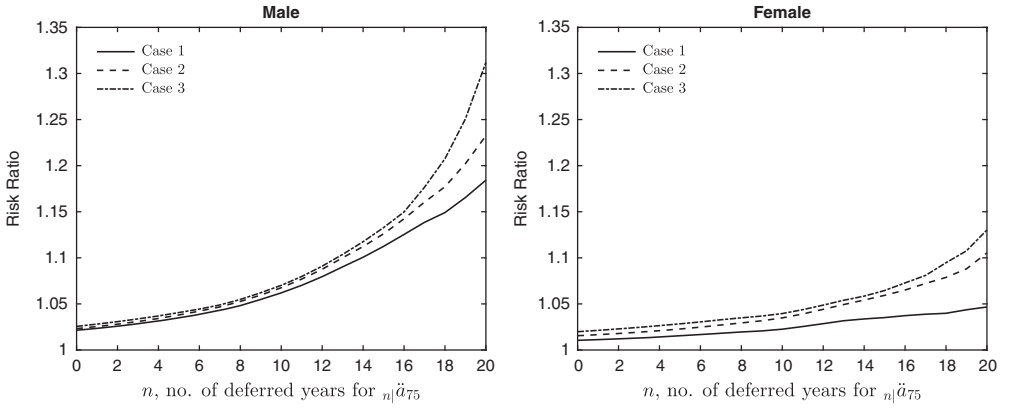

Figure A.2. Comparison of risk ratios of ${}_{n|}\ddot{a}_{75}$ in three cases, using the Mitchell *et al.* (2013) model for mortality improvement rates: Case 1: risk ratios amongst dependence structures, assuming Gaussian extrapolation for $\alpha_x^{(i)}$ and Perks' mortality curve; Case 2: risk ratios amongst six combinations of dependence structures and extrapolation methods for $\alpha_x^{(i)}$, assuming Perks' mortality curve; Case 3: risk ratios amongst 18 combinations of dependence structures, extrapolation methods for $\alpha_x^{(i)}$, and mortality tail curves. $i = 2\%$.

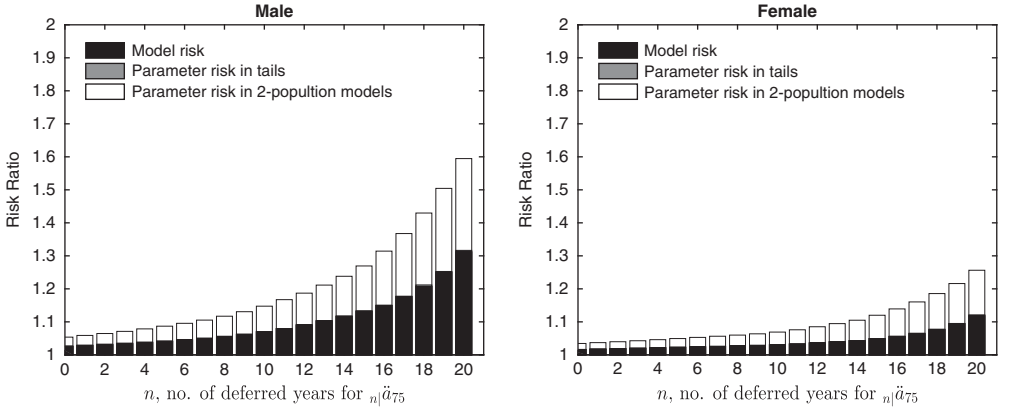

Figure A.3. Risk ratios of ${}_{n|}\ddot{a}_{75}$ amongst 18 combinations of mortality tail curve, extrapolation method for $\alpha_x^{(i)}$, and dependence structure in two-population Mitchell model, with parameter uncertainty incorporated. $i = 2\%$.

represented by the white part, is very significant for both male and female annuity values. The overall risk in male annuity values is much higher than that in female annuity values. These findings are again similar with what we observe in Figure 12 from using the Lee–Carter model for mortality improvement rates.

https://doi.org/10.1017/S174849951700001X Published online by Cambridge University Press