Author/s:
Li, JS-H;Chan, W-S;Zhou, R

Title:
Semicoherent Multipopulation Mortality Modeling: The Impact on Longevity Risk Securitization

Date:
2017-09

Citation:
Li, J. S. -H., Chan, W. -S. & Zhou, R. (2017). Semicoherent Multipopulation Mortality Modeling: The Impact on Longevity Risk Securitization. Journal of Risk and Insurance, 84 (3), pp.1025-1065. https://doi.org/10.1111/jori.12135.

Persistent Link:
https://hdl.handle.net/11343/291687

Semi-Coherent Multi-Population Mortality Modeling: The Impact on Longevity Risk Securitization

Abstract

Multi-population mortality models play an important role in longevity risk transfers involving more than one population. Most of the existing multi-population mortality models are built on the hypothesis of coherence, which assumes that there always exists a force that brings the mortality differential between any two populations back to a constant long-term equilibrium level. This hypothesis prevents diverging long-term forecasts, which do not seem to be biologically reasonable. However, the coherence assumption may be perceived by market participants as too strong and is in fact not always supported by empirical observations. In this paper, we introduce a new concept called 'semi-coherence', which is less stringent in the sense that it permits the mortality trajectories of two related populations to diverge, as long as the divergence does not exceed a specific tolerance corridor, beyond which mean-reversion will come into effect. We further propose to produce semi-coherent mortality forecasts by using a vector threshold autoregression. The proposed modeling approach is illustrated with mortality data from US and English and Welsh male populations, and is applied to several pricing and hedging scenarios.

Keywords: Longevity bonds; q-forwards; S-forwards; Population basis risk; Vector threshold autoregression

1 Introduction

Longevity risk has become a high profile risk in recent years, due partly to the changes in insurance regulations after the global financial crisis. For instance, Solvency II, which is presently scheduled to be implemented in January 2016, requires insurers operating in the European Union to hold longevity risk solvency capital to cushion against the adverse financial consequences arising from unexpected changes in mortality rates. Another reason is the deep interest rate cuts followed by quantitative easing. At the current low interest rates, life-contingent cash flows payable in the distant future, which are subject to the greatest amount of longevity risk, are not discounted that heavily.

Securitization is increasingly seen as a solution to the problem of longevity risk. Through securitization, a financial institution can transfer its longevity risk exposure to one or more coun-

terparties who are interested in accepting the risk exposure for a premium. The potential of the longevity risk transfer market is huge. It is estimated that in the United Kingdom alone, there are approximately £1 trillion of defined-benefit pension scheme liabilities managed by corporate pension schemes and around £200 billion of insured annuity-in-payment liabilities managed by life insurance companies (Bugg et al., 2010). The market has recently seen several large transactions, exemplified by the £3 billion longevity swap that Rolls Royce transacted with Deutsche Bank in 2011 to cover the longevity risk of the 37,000 pensioners in its pension plan (Coughlan et al., 2013).

However, relative to its potential size, the longevity risk transfer market has developed rather slowly (Coughlan et al., 2013). The market has to overcome many challenges before it can become liquid, with substantial volumes on both the supply and demand sides. As Cairns (2013a) pointed out, the key challenge on the modeling front is to develop multi-population stochastic mortality models, which are crucially important for the following reasons.

Natural buyers of longevity risk are limited only to life (re-)insurance companies with a concentration on policies paying death benefits and perhaps companies in the health care, long-term care and pharmaceutical industries that would benefit from increased longevity. To further expand in size, the longevity risk transfer market has to reach out to other potential investors, such as hedge funds and asset managers, who may be attracted to this new asset class due to its low correlation with other market risk factors (Ribeiro and di Pietro, 2009). These investors prefer standardized index-based securities to customized deals, because the former are more transparent and liquid. By contrast, hedgers may prefer customized hedging solutions, due to their concerns about the population basis risk that arises from the difference in mortality experience between the populations associated with their portfolios and the populations to which the standardized hedging instruments are linked. Multi-population stochastic mortality models permit hedgers to better measure the potential population basis risk, giving them more confidence in hedging with standardized securities.

The European Union's rules on gender-neutral pricing have recently become effective. The rules require insurers in Europe to charge the same prices to women and men for the same insurance products without distinction on the grounds of sex. Life insurers in Europe therefore have a pressing need for multi-population stochastic mortality models, which they can use to simultaneously forecast the mortality of both genders. Multi-population models can also assist them with the calculation of solvency risk capital and the amount of risk exposure that should be transferred through reinsurance or capital markets.

In the market there exist mortality-linked securities that are associated with more than one population. For instance, in 2010, Swiss Re issued a series of eight-year longevity bonds (known as the Kortis deal) valued at $50 million to hedge its life insurance exposure in the United States and pension exposure in the United Kingdom (Blake et al., 2014). The principal repayment of this bond depends on the divergence in realized mortality improvements between males aged 55-65 in

the United States and males aged 75-85 in England and Wales. To value securities similar to the Kortis deal, a multi-population stochastic mortality model is indispensable.

In the past few years, there has been a significant volume of research work on the problem of multi-population mortality modeling. Several multi-population stochastic mortality models have been proposed by researchers, including Ahmadi and Li (2014), Cairns et al. (2011), Chen et al. (2013), Dowd et al. (2011), Hatzopoulos and Haberman (2013), Hunt and Blake (2013), Jarner and Kryger (2011), Li and Hardy (2011), Li and Lee (2005), Lin et al. (2013), Yang and Wang (2013) and Zhou et al. (2013, 2014). The majority of the recent developments in multi-population mortality modeling have been built on an important concept called coherence, originally proposed by Li and Lee (2005).

In coherent mortality forecasting, the future mortality rates of two or more related populations are modeled jointly in such a way that there is always a force that brings the spread between the mortality rates of any two populations being modeled back to a constant level. Hence, coherent mortality forecasts do not diverge indefinitely over time. For example, in Li and Lee's (2005) augmented common factor model, coherence is achieved by modeling each population-specific time trend by a first order autoregressive process. The coherence hypothesis is supported in part by the argument that differences in mortality between related populations should not increase over time indefinitely if their similar socioeconomic conditions and close connections continue. It is also 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, 1950-55, 1975-80 and 2000. The same conclusion was drawn by White (2002) who considered mortality data from 21 high-income countries over the period of 1955 to 1996.

Nevertheless, whether a mortality convergence can be observed depends very much on the observation window and the populations involved. Oeppen and Vaupel (2002), who introduced the notion of best-practice life expectancy, noted that although a number of countries including Japan and Chile are converging to the best-practice level, some countries such as the United States have been moving away from it in recent decades. Hatzopoulos and Haberman (2013) computed the probabilities of non-coherence (non-convergence) for the residual age-period model structures for a group of 35 national populations. Their estimation results indicated that for some populations, for example, England and Wales, the probabilities of non-coherence were large. The conclusions of several other studies have contradicted the hypothesis of coherence. Based on data from the U.S. National Longitudinal Mortality Study, Waldron (2007) found that for individuals who died in 1972-2001 at ages 60-89, the top half of the average relative earnings distribution experienced faster mortality improvement than the bottom half. This pattern, according to Mackenbach et al. (2003), also exists in some other developed countries, including Finland, Sweden, Norway, Denmark, England and Wales, and Italy. Therefore, the hypothesis of coherence may be too strong and may not always be appropriate. If it turns out that the coherence hypothesis is not true, then a coherent multi-population mortality model may understate the range of possible mortality

differentials between two populations. The model may then underestimate the population basis risk in index-based longevity hedges and mis-price longevity securities that are structured like the Kortis deal.

In this paper, we address the aforementioned problem by introducing a new concept called semi-coherence, which is less stringent than full-coherence in the sense that it permits the mortality trajectories of two related populations to diverge, as long as the divergence does not exceed a specific tolerance corridor, beyond which mean-reversion will come into effect. Compared with fully-coherent forecasts, semi-coherent forecasts exhibit a number of distinctive features. First, semi-coherence does not strictly require mean-reversion to begin immediately at the forecast origin. Instead, under a semi-coherent mortality model, mean-reversion begins at random future time points when the mortality divergence between two populations goes beyond the specified tolerance corridor. Second, in a semi-coherent forecast, the mortality trajectories of two related populations are permitted to diverge over certain periods of time, thereby covering a wider range of possible mortality differentials. However, in contrast to a non-coherent forecast, the expected mortality trajectories do not diverge indefinitely, as mean-reversion is enforced when the divergence becomes too large. The term semi-coherence is used to describe the concept we propose because it can be regarded as a compromise between full-coherence, which corresponds to the special case in which the upper and lower bounds of the tolerance corridor are equal, and non-coherence, which corresponds to the other extreme in which the upper and lower bounds are positive and negative infinities, respectively.

We further propose to produce semi-coherent mortality forecasts by a vector threshold autoregressive (VETAR) process, which was developed by Tsay (1998) as a generalization of its univariate counterpart, originally proposed by Tong and Lim (1980). A VETAR process contains multiple regimes, each of which is characterized by a different parameter set and possibly a different autoregressive order, so that the behaviour of the time series changes once it enters a different regime. The switch from one regime to another is determined by a threshold variable, which depends on the past values of the time series. To generate semi-coherent mortality forecasts, we model the evolution of the period effects in a two-population Lee-Carter structure by a three-regime VETAR process, of which the threshold variable is defined as the average of the differences between the period effects of the two populations over the past few years. The VETAR process can sojourn in the middle regime, in which the period effects of the two populations reduce at different expected speeds. When the threshold variable becomes too high or too low, the process switches to one of the other two regimes, in which the reduction speeds are adjusted to enforce mean-reversion.

The rest of this paper is organized as follows. Section 2 describes the mortality data we use for illustration purposes. Section 3 discusses the concept of semi-coherence in greater depth. Section 4 presents the VETAR process and demonstrates how it can be utilized to produce semi-coherent mortality forecasts. Section 5 evaluates the proposed modeling approach and compares it with existing fully-coherent and non-coherent modeling methods. Sections 6 and 7 investigate the

effects of different forms of coherence on longevity risk hedging and pricing. Finally, Section 8 concludes the paper with some directions for future research.

2 Data

The illustrations in this paper are based on mortality data from the male populations of the United States and England and Wales. We consider a sample period of 1901 to 2010 and a sample age range of 55 to 85.

For United States males, the death counts and central death rates from 1901 to 1932 are provided by the Centers for Diseases Control and Prevention (CDC).[1] The exposure counts needed in the estimation process are computed by dividing each death count by its corresponding central death rate. The values provided by the CDC were originally arranged by decennial age groups 55-64, 65-74 and 75-84 and an open age group 85+. We convert them to values by individual ages with the data disaggregation method proposed by Boot et al. (1967). The death and exposure counts by individual ages beyond 1932 are obtained from the Human Mortality Database (2014).

For English and Welsh males, all of the required data (historical death and exposure counts by individual ages) are obtained from the Human Mortality Database (2014).

Throughout this paper, population $i = 1$ refers to English and Welsh males and population $i = 2$ refers to United States males.

3 The Concept of Semi-Coherence

To keep the exposition simple, we only consider models that belong to the Lee-Carter family and contain no cohort effects. However, the concepts discussed in this section can be applied readily to other models, including those that incorporate cohort effects.

Let us begin this discussion with a brief review of what is meant by non-coherence and full-coherence. For simplicity, we focus on modeling only two populations simultaneously.

3.1 Non-Coherence

Mortality forecasts generated by an arbitrary combination of single-population models are generally non-coherent. Consider a combination of two unrelated single-population Lee-Carter models:

$$\ln(m(x,t,i)) = \alpha_x^{(i)} + \beta_x^{(i)}\kappa_t^{(i)}, \qquad i = 1, 2, \tag{3.1}$$

where $m(x,t,i)$ is population i's central rate of death at age x and in year t, $\alpha_x^{(i)}$ is a parameter indicating population i's average mortality level at age x, $\kappa_t^{(i)}$ is the period effect index representing

[1] www.cdc.gov

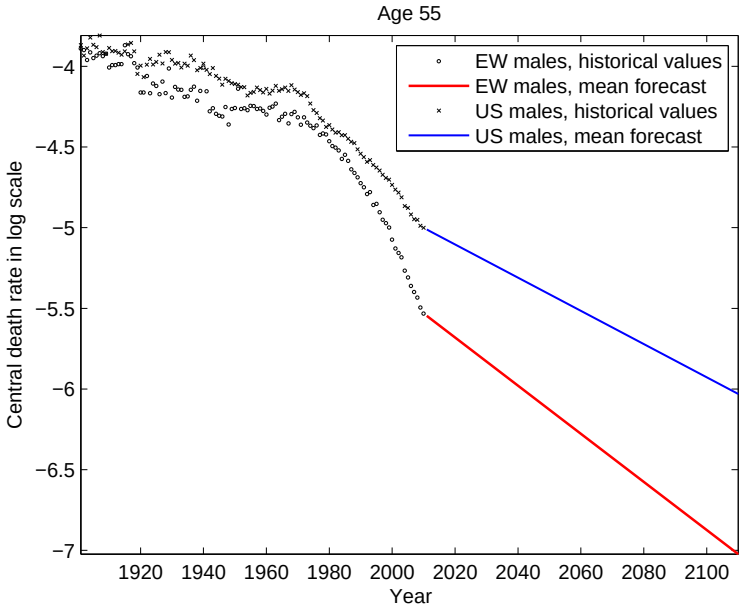

Figure 1: The mean forecast of the central death rates for age 55 derived from two independent Lee-Carter models, English and Welsh (EW) males and United States (US) males.

population i's overall level of mortality at time t, and $\beta_x^{(i)}$ is a parameter reflecting the sensitivity of the log central death rate to changes in $\kappa_t^{(i)}$. We obtain estimates of $\alpha_x^{(i)}$, $\beta_x^{(i)}$ and $\kappa_t^{(i)}$ by using the method of maximum likelihood under the assumption of Poisson death counts (Wilmoth, 1993).

To capture the evolution of $\ln(m(x,t,i))$ over time, each series of $\kappa_t^{(i)}$ is further modeled by a random walk with drift; that is,

$$\kappa_t^{(i)} = c^{(i)} + \kappa_{t-1}^{(i)} + \epsilon_t^{(i)}, \qquad i = 1, 2, \tag{3.2}$$

where $c^{(i)}$ is a constant, and $\{\epsilon_t^{(i)}\}$ is a sequence of i.i.d. normal random variables such that $\epsilon_t^{(1)}$ and $\epsilon_t^{(2)}$ are independent. By extrapolating from the two stochastic processes, mortality forecasts can be obtained readily.

The independent Lee-Carter models are fitted to data from the male populations of England and Wales and the United States. The resulting mean forecast of the central death rates for age 55 is depicted in Figure 1. We call this mortality forecast non-coherent, because the projected difference in the mortality rates of the two populations diverges indefinitely over time. We may also describe this mortality forecast as biologically unreasonable, because the indefinite divergence is difficult, if not impossible, to justify on the grounds of biology, medicine or environment (Cairns et al., 2006; Dowd et al., 2010).

3.2 Full-Coherence

The problem of indefinitely diverging mortality forecasts was acknowledged and addressed by Li and Lee (2005), who proposed the concept of coherent mortality forecasting. A multi-population mortality model is regarded as (fully) coherent if there always exists a force that brings the mortality differential between any two populations being modeled back to a constant long-term equilibrium level. Hence, in expectation terms, the mortality differentials projected from a fully-coherent model are level rather than diverging over the long run.

Fully coherent Lee-Carter mortality forecasts can be derived in different manners. For example, Cairns et al. (2011) considered the following two-population extension of the Lee-Carter model:

$$\ln(m(x,t,i)) = \alpha_x^{(i)} + \beta_x \kappa_t^{(i)}, \qquad i = 1, 2, \tag{3.3}$$

where $\alpha_x^{(i)}$ represents population i's average mortality level at age x, $\kappa_t^{(i)}$ is the time-varying parameter for population i, and β_x measures the sensitivity of $\ln(m_(x,t,i))$ to changes in $\kappa_t^{(i)}$.[2] To ensure that the resulting forecast is fully coherent, this model structure requires both populations to share the same set of sensitivity parameters, β_x's. As before, the parameters in equation (3.3) are estimated by maximum likelihood.

Another condition for full-coherence is that $\kappa_t^{(1)} - \kappa_t^{(2)}$, the difference between the period effects of the two populations, is mean-reverting. Accordingly, Cairns et al. (2011) modeled the period effects $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ by the following stochastic processes:

$$\kappa_{t+1}^{(1)} = \kappa_t^{(1)} + \mu_\kappa + N_\kappa(t+1), \tag{3.4}$$

$$\Delta_\kappa(t) = \kappa_t^{(1)} - \kappa_t^{(2)}, \tag{3.5}$$

$$\Delta_\kappa(t+1) = \mu_{\Delta_\kappa} + \phi_{\Delta_\kappa} \Delta_\kappa(t) + N_{\Delta_\kappa}(t+1), \tag{3.6}$$

where μ_κ is a constant, and ϕ_{Δ_κ} is another constant with an absolute value that is strictly less than 1. These stochastic processes ensure non-divergence between $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ and hence between $m(x,t,1)$ and $m(x,t,2)$, because $\Delta_\kappa(t)$ will revert to a constant $\mu_{\Delta_\kappa}/(1-\phi_{\Delta_\kappa})$ over the long run. In this set up, population 1 is regarded as the dominant population, since it drives the mortality dynamics of both populations under consideration.

Additionally, Cairns et al. (2011) observed empirically that significant positive correlations between the random effects of different populations exist, so they treated the innovations $N_\kappa(t)$ and $N_{\Delta_\kappa}(t)$ jointly as a sequence of i.i.d. bivariate normal random vectors with zero mean. This distributional assumption allows us to derive the processes' log-likelihood, which is then maximized to obtain parameter estimates.

Equations (3.3) and (3.6) are fitted to the mortality data from English and Welsh and United States male populations. The resulting mean forecast for age 55 is displayed in Figure 2. Equation

[2] This model structure appears in an earlier (and fuller) version of Cairns et al. (2011), which is available at http://www.ma.hw.ac.uk/_andrewc/papers/ajgc54.pdf.

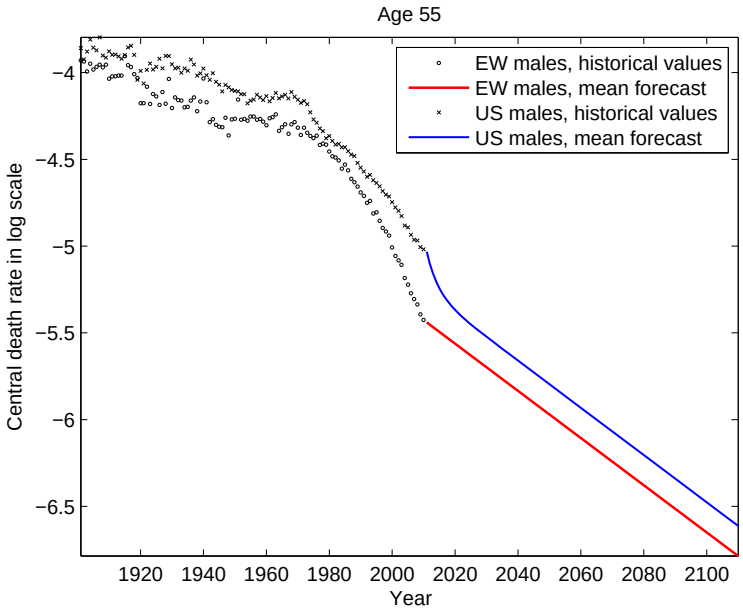

Figure 2: The mean forecast of the central death rates for age 55 derived from the fully-coherent two-population Lee-Carter model considered by Cairns et al. (2011), English and Welsh (EW) males and United States (US) males.

(3.4) implies that the expected trajectory of $\ln(m(x,t,1))$ is a straight line with a slope of $\beta_x\mu_\kappa$. On the other hand, equations (3.5) and (3.6) ensure that the trajectory of $\ln(m(x,t,2))$ starts reverting to its long-term equilibrium path, which has the same gradient as that of the dominant population's trajectory, immediately at the forecast origin. It is observed that the reversion to the long-term equilibrium is quick, taking less than 10 years. Due to the autoregressive parameter ϕ_{Δ_κ} in equation (3.6), any future deviation from the long-term equilibrium will be pulled back to zero.

Although full-coherence has its merits, it is subject to several significant limitations. For instance, it does not seem straightforward to justify immediate, quick convergence to the long-term equilibrium. Other limitations of the fully-coherent mortality model can be seen from Figure 3, which shows the estimates of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ in equation (3.3). Since the 1990s, the gradients of the two trends have been quite different and there is no sign that the two trends are beginning to converge. Over the observation window, there are several time periods (e.g., 1940-1950, 1970-1980, 1990-2010) during which one trend reduces persistently and substantially faster than the other.

3.3 Semi-Coherence

Inspired by the limitations of non- and full-coherence, we propose a new concept called semi-coherence, which is less restrictive than full-coherence, but still includes some sort of mean-

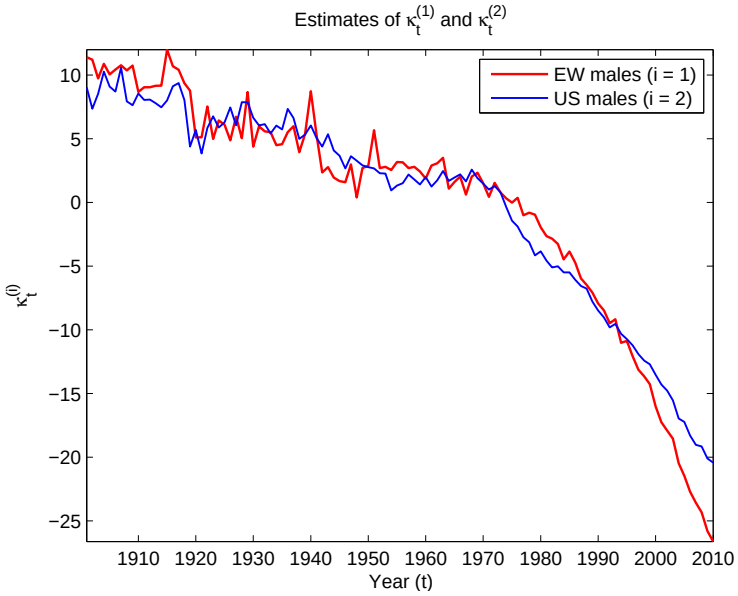

Figure 3: The estimates of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ in equation (3.3), English and Welsh (EW) males and United States (US) males.

reversion (with respect to the difference in mortality between two populations) to avoid an indefinite divergence.

The concept of semi-coherence lies in a tolerance corridor, which is imposed on the mortality differential between two populations. Provided that the differential lies within the tolerance corridor, the expected mortality trajectories of the two populations can diverge. However, when the tolerance corridor is exceeded, then the stochastic process driving the mortality dynamics of the two populations should switch to another state, in which the expected paces of mortality reduction for the two populations are changed in such a way that the differential (in absolute value) is expected to shrink over time. The concept of semi-coherence is illustrated graphically in the middle panel of Figure 4.

Without assuming a specific stochastic process, it can be deduced that under the semi-coherence hypothesis, mortality forecasts possess the following distinctive characteristics:

- Depending on the initial value of the mortality differential relative to the tolerance corridor, mean-reversion does not necessarily begin at the forecast origin.
- Generally speaking, mean-reversion is effective in only part of the projected mortality trajectories. Hence, the expected mortality trajectories of the two populations being modeled can diverge over certain periods of time, thereby covering a wider range of possible mortality differentials in comparison to a fully-coherent modeling approach.
- Mean-reversion comes into effect at random time points when the tolerance corridor is exceeded. The intermittent mean-reversions prevent the expected trajectories from diverging

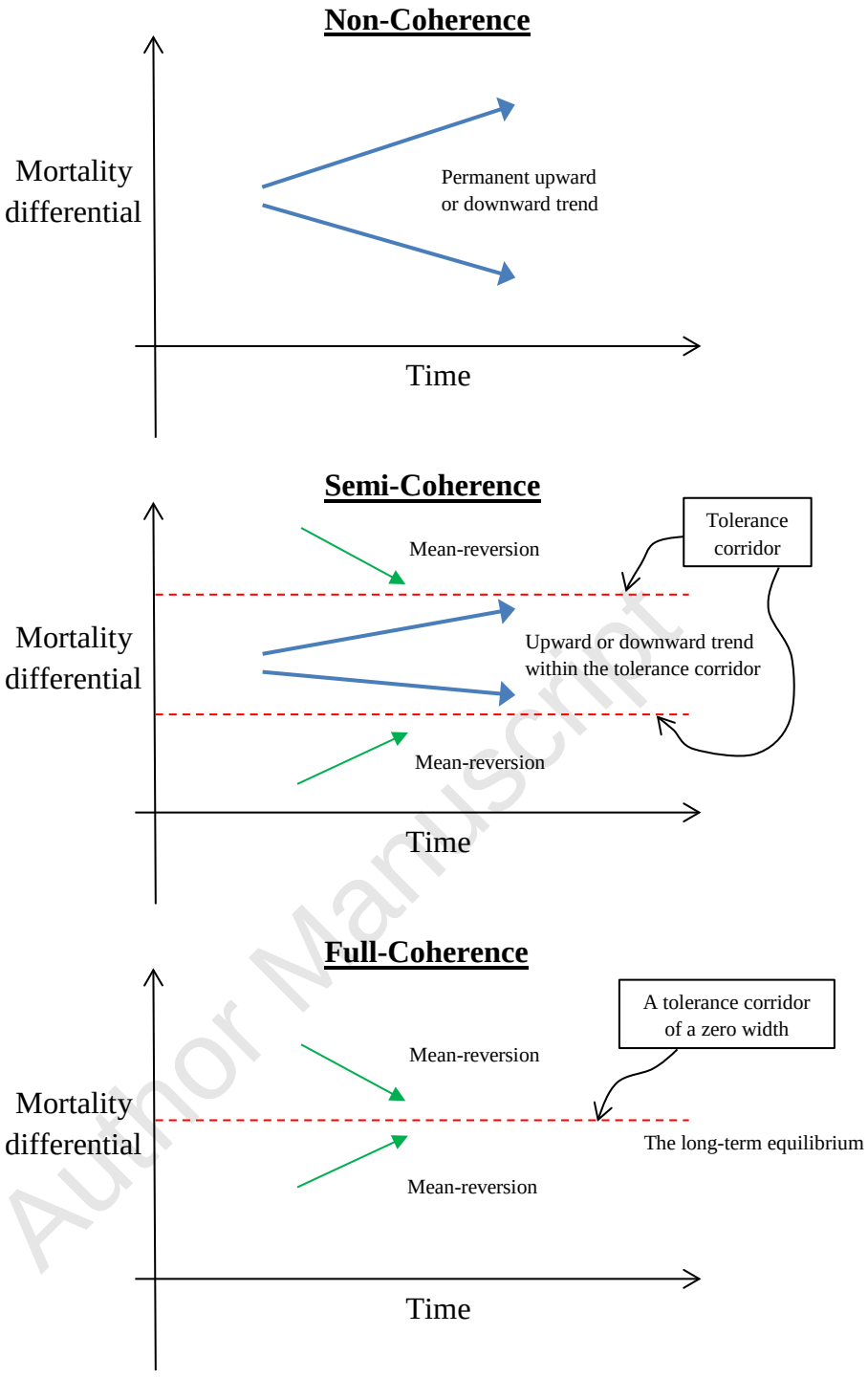

Figure 4: A graphical illustration of the three forms of coherence.

indefinitely.

We describe the concept we propose as semi-coherence, because in terms of restrictiveness it is in between full- and non-coherence. More specifically, it is less stringent than full-coherence, which can be regarded as the special case in which the width of the tolerance corridor is zero, but is purposely more restrictive than non-coherence, which corresponds to the other extreme in which the tolerance corridor has an infinite width. The distinctions between the three forms of coherence are illustrated diagrammatically in Figure 4.

In terms of modeling techniques, semi-coherence requires the underlying stochastic process to encompass multiple regimes. The switch between regimes should be driven by the mortality differentials between the two populations being modeled over a certain lookback period. This configuration is significantly different from typical regime-switching models, for example, the model of Hardy (2001), in which the change in regime is controlled by a Markovian random process.

4 The Vector Threshold Autoregression

4.1 The VETAR Process in its General Form

The general VETAR process contains $k \geq 2$ regimes, each of which is characterized by a different parameter set and possibly a different autoregressive order. The switch from one regime to another depends on the past values of the time-series, and therefore the VETAR process is often described as being 'self-exciting'.

Let $\{\mathbf{Z}_t\}$ be a vector time-series. The general VETAR process can be expressed as

$$\mathbf{Z}_t = \begin{cases} \boldsymbol{\phi}^{(1)} + \sum_{j=1}^{p_1} \boldsymbol{\Phi}_j^{(1)} \mathbf{Z}_{t-j} + \mathbf{a}_t^{(1)}, & y_{t-d} \leq r_1 \\ \boldsymbol{\phi}^{(2)} + \sum_{j=1}^{p_2} \boldsymbol{\Phi}_j^{(2)} \mathbf{Z}_{t-j} + \mathbf{a}_t^{(2)}, & r_1 < y_{t-d} \leq r_2 \\ \qquad\qquad \vdots & \qquad\qquad \vdots \\ \boldsymbol{\phi}^{(k)} + \sum_{j=1}^{p_k} \boldsymbol{\Phi}_j^{(k)} \mathbf{Z}_{t-j} + \mathbf{a}_t^{(k)}, & r_{k-1} < y_{t-d} \end{cases}, \tag{4.1}$$

where $-\infty < r_1 < \ldots < r_{k-1} < +\infty$. In the gth regime ($1 \leq g \leq k$), $\mathbf{Z}_t$ follows an order p_g vector autoregression, which has constant vector $\boldsymbol{\phi}^{(g)}$ and autoregressive matrices $\boldsymbol{\Phi}_j^{(g)}$, $j = 1, \ldots, p_g$. The innovation vector in regime g is given by $\mathbf{a}_t^{(g)}$, which satisfies the equation $\mathbf{a}_t^{(g)} = \boldsymbol{\Sigma}_g^{1/2} \mathbf{e}_t$, where $\boldsymbol{\Sigma}_t^{1/2}$ is a symmetric positive definite matrix and $\{\mathbf{e}_t\}$ is a sequence of serially uncorrelated normal random vectors with a zero mean vector and an identity covariance matrix.

The change of regime is governed by the threshold random variable, y_{t-d}, which is a function of the history of the time-series in question up to and including time $t - d$, where d is the delay parameter identified during the estimation process. The value of y_{t-d} relative to the fixed threshold parameters $r_1, \ldots, r_{k-1}$ determines which regime the system is currently in.

4.2 Using the VETAR Process for Semi-Coherent Mortality Forecasting

To produce semi-coherent mortality forecasts, we use the model structure specified in equation (3.3), that is, the Lee-Carter structure with common age response parameters for both populations. However, rather than the collection of processes specified in equations (3.4) to (3.6), a VETAR process is used to model the evolution of the period effects $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$.

In particular, we use a three-regime VETAR process to model $\mathbf{Z}_t = (\Delta\kappa_t^{(1)}, \Delta\kappa_t^{(2)})'$, where Δ is the difference operator and the prime sign denotes transpose. We can understand $\Delta\kappa_t^{(i)}$ as the overall mortality improvement for population i from time $t-1$ to t. In different regimes, the dynamics of the improvement rates of the two populations being modeled are different.

Because we intend to have the change in regime driven by the difference in mortality between the two populations, we define the threshold random variable as

$$y_t = \frac{1}{\ell}\sum_{i=0}^{\ell-1}(\kappa_{t-i}^{(1)} - \kappa_{t-i}^{(2)}), \tag{4.2}$$

which can be interpreted to mean the average difference in the general mortality levels between the two populations over a lookback period of ℓ years. The averaging is to avoid changes in regime from being triggered by short-term mortality jumps that are due to, for example, flu epidemics. The results presented in this paper are based on $\ell = 5$, which seems reasonable because a value of ℓ that is too large will substantially reduce the effective number of observations used in estimation while a value of ℓ that is too small may lead to problems associated with short-term mortality jumps. Given the way in which the threshold variable is defined, the threshold parameters r_1 and r_2 can be considered as the lower and upper limits of the tolerance corridor, respectively.

Let

$$(\mu^{(g)}(1), \mu^{(g)}(2))' = (I - \sum_{j=1}^{p_g} \mathbf{\Phi}_j^{(g)})^{-1}\boldsymbol{\phi}^{(g)},$$

where I is the 2×2 identity matrix. We can regard $\mu^{(g)}(1)$ and $\mu^{(g)}(2)$ as the steady-state expected rates of change in the period effects for populations 1 and 2, respectively, in the gth regime. It is expected that the estimates of $\mu^{(g)}(1)$ and $\mu^{(g)}(2)$ are negative for $g = 1, 2, 3$, because of the continuous reduction in mortality observed over the data sample period. Moreover, it is expected that

$$\hat{\mu}^{(1)}(1) > \hat{\mu}^{(1)}(2) \quad \text{and} \quad \hat{\mu}^{(3)}(1) < \hat{\mu}^{(3)}(2), \tag{4.3}$$

where the $\hat{}$ sign denotes an estimate, if intermittent mean reversions are captured by the VETAR process.

It is not difficult to see that semi-coherence is achieved through this three-regime VETAR process. In regime 2, in which the mortality differential (measured by the threshold variable

y_{t-d}) lies within the tolerance corridor $(r_1, r_2]$, the period effects of the two populations gradually diverge, as they change at different expected speeds ($\hat{\mu}^{(2)}(1)$ and $\hat{\mu}^{(2)}(2)$). However, an indefinite divergence is avoided by mean-reversion, which becomes effective when y_{t-d} exceeds $(r_1, r_2]$. In more detail, if population 1's mortality improves faster than population 2's for a prolonged period of time so that y_{t-d} becomes smaller than the lower limit r_1, then the process switches to regime 1, in which $\kappa_t^{(1)}$ reduces less rapidly than $\kappa_t^{(2)}$ (i.e., $\mu^{(1)}(1)$ is less negative than $\mu^{(1)}(2)$). If the opposite happens, then the process switches to regime 3, in which the reduction in $\kappa_t^{(2)}$ is slower than in $\kappa_t^{(1)}$. Note that the VETAR process does not strictly require mean-reversion to be enforced immediately at the forecast origin, because the system can start in any one of the three regimes, depending on the initial value of the threshold variable.

In general, $|\hat{\mu}^{(1)}(1) - \hat{\mu}^{(1)}(2)|$ is not equal to $|\hat{\mu}^{(3)}(1) - \hat{\mu}^{(3)}(2)|$, which means that the forces of mean reversion in regimes 1 and 3 are different. This property allows the process to capture the potential asymmetry in population basis risk.

4.3 Testing VETAR-Type Nonlinearity and Estimating the VETAR Process

In this sub-section, we explain and implement Tsay's (1998) method for testing VETAR-type nonlinearity and estimating VETAR processes. We shall restrict the discussion to points necessary for describing the relevant applications in this paper. Further computational details can be found in Tsay (1998, 2010).

Conceptually, a three-regime VETAR process that is well-fitted to historical values of $\mathbf{Z}_t = (\Delta\kappa_t^{(1)}, \Delta\kappa_t^{(2)})'$ and satisfies the inequalities specified in equation (4.3) will produce semi-coherent mortality forecasts. However, in practice, a VETAR process should not be assumed without checking the actual data. To justify the appropriateness of a VETAR process for $\mathbf{Z}_t = (\Delta\kappa_t^{(1)}, \Delta\kappa_t^{(2)})'$, we employ Tsay's (1998) formal statistical procedure for testing the null hypothesis that $\mathbf{Z}_t$ follows an ordinary linear vector autoregressive model (VAR) versus the alternative hypothesis that it follows the non-linear VETAR process specified in equation (4.1).

Consider a three-regime VETAR process with $p = \max\{p_1, p_2, p_3\}$ and $d \leq p$, fitted to historical values of $\mathbf{Z}_t = (\Delta\kappa_t^{(1)}, \Delta\kappa_t^{(2)})'$. Let T be the number of historical values of $\mathbf{Z}_t$ used in the fitting and (τ) be the time index of the τth smallest observations in the set of $\{y_{t-d}\}$. Tsay (1998) proposed the following arranged autoregression formulation that is based on the increasing order of the threshold variable y_{t-d}:

$$\mathbf{Z}_{(\tau)+d} = \mathbf{\Phi}'\mathbf{X}_{(\tau)+d} + \mathbf{a}_{(\tau)+d} \qquad \text{for } \tau = 1, \ldots, T-p, \tag{4.4}$$

where $\mathbf{X}_{(\tau)+d} = (\mathbf{1}, \mathbf{Z}_{(\tau)+d-1}, \ldots, \mathbf{Z}_{(\tau)+d-p})'$ is a $(3p+1)$-dimensional autoregressor and $\mathbf{\Phi} = (\boldsymbol{\phi}, \mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_p)'$ is the corresponding parameter matrix. If $\mathbf{Z}_t$ is linear, then the non-linear VETAR process collapses into a linear VAR process with $(\boldsymbol{\phi}^{(g)}, \mathbf{\Phi}_1^{(g)}, \ldots, \mathbf{\Phi}_p^{(g)})' = \mathbf{\Phi}$ for all $g = 1, 2, 3$. In this case, the recursive least squares estimator of the arranged autoregression in equation (4.4) is consistent and the corresponding predictive residuals follow a vector white noise process. On the

Table 1: Tests for VETAR-type non-linearity

p	d	Test statistic	degrees of freedom (λ)	p-value
1	1	8.33	6	0.2147
2	1	26.78	10	0.0028
2	2	18.98	10	0.0405
3	1	32.23	14	0.0037
3	2	25.95	14	0.0262
3	3	29.56	14	0.0088

other hand, if $\mathbf{Z}_t$ is threshold-type non-linear, then the predictive residuals move away from the white noise pattern after the time-series crosses the threshold(s). Tsay (1998) made use of this property to derive a chi-square test for VETAR-type non-linearity.

We now apply Tsay's (1998) testing procedure to the time-series of $\mathbf{Z}_t = (\Delta\kappa_t^{(1)}, \Delta\kappa_t^{(2)})'$ for $t = 1902, \ldots, 2010$, obtained from the mortality data described in Section 2 and the base model structure specified by equation (3.3). It should be noted that the number of parameters in the VETAR process increases exponentially with p. Given the number of data points we have, it is practically impossible to consider VETAR processes with $p > 3$. For this reason, we only consider VETAR models with $p \leq 3$. Under the null hypothesis of linearity, Tsay's test statistic follows asymptotically a chi-squared distribution with $\lambda = (4p + 2)$ degrees of freedom. Table 1 shows the nonlinearity test results for various combinations of p and $d \leq p$. At 5% level of significance, the test results indicate VETAR-type non-linearity in the data for $p \geq 2$. On the basis of the test results, we choose $p = 2$ and $d = 1$, because this combination gives the smallest p-value among all of the combinations of p and d considered.

The next step is to estimate the threshold values, r_1 and r_2. If both r_1 and r_2 are known, then the arranged autoregression in equation (4.4) can be divided into three regimes. Let π_1 denote the largest value of (τ_1) such that $\{-\infty < y_{(\tau_1)} \leq r_1\}$ and π_2 denote the largest value of (τ_2) such that $\{r_1 < y_{(\tau_2)} \leq r_2\}$. The first regime contains $n_1 = \pi_1$ data points with $(\tau) = 1, \ldots, \pi_1$, the second regime contains $n_2 = \pi_2 - \pi_1$ data points with $(\tau) = \pi_1 + 1, \ldots, \pi_2$, and the third regime contains $n_3 = T - p - \pi_2$ data points with $(\tau) = \pi_2 + 1, \ldots, T - p$. For fixed values of r_1 and r_2, the least-squares estimate of the parameter matrix $\mathbf{\Phi}^{(g)}$ in the gth regime can be computed using the ordinary multivariate least-squares method. The corresponding regime residual variance-covariance matrix $\hat{\mathbf{\Sigma}}_g$ can also be obtained readily.

The values of r_1 and r_2 are then estimated on the basis of the Akaike information criterion (AIC). The AIC for a three-regime VETAR process is defined as

$$\text{AIC} = \sum_{g=1}^{3} \left\{ n_g \left| \hat{\mathbf{\Sigma}}_g \right| + 6(3p + 1) \right\}. \tag{4.5}$$

Given the chosen values of p and d, we use a grid search method to select the values of r_1 and r_2

Table 2: VETAR model estimation results

$\hat{\boldsymbol{\phi}}^{(g)}$	$\hat{\boldsymbol{\Phi}}_1^{(g)}$	$\hat{\boldsymbol{\Phi}}_2^{(g)}$	$\hat{\boldsymbol{\Sigma}}_g$

(a) First regime ($g = 1$)

$$\begin{pmatrix} -0.38 \\ (0.33) \\ \\ -0.62 \\ (0.12) \end{pmatrix} \quad \begin{pmatrix} 0.01 & 0.05 \\ (0.28) & (0.44) \\ \\ 0.13 & -0.21 \\ (0.10) & (0.16) \end{pmatrix} \quad \begin{pmatrix} 0.71 & -0.99 \\ (0.27) & (0.41) \\ \\ 0.14 & -0.41 \\ (0.10) & (0.15) \end{pmatrix} \quad \begin{pmatrix} 1.8863 & 0.4888 \\ & \\ & 0.2500 \\ & \end{pmatrix}$$

(b) Second regime ($g = 2$)

$$\begin{pmatrix} -0.45 \\ (0.15) \\ \\ -0.32 \\ (0.11) \end{pmatrix} \quad \begin{pmatrix} -0.45 & 0.16 \\ (0.11) & (0.18) \\ \\ 0.04 & -0.43 \\ (0.08) & (0.13) \end{pmatrix} \quad \begin{pmatrix} -0.35 & 0.40 \\ (0.10) & (0.17) \\ \\ -0.07 & 0.16 \\ (0.07) & (0.13) \end{pmatrix} \quad \begin{pmatrix} 0.9880 & 0.2331 \\ & \\ & 0.5233 \\ & \end{pmatrix}$$

(c) Third regime ($g = 3$)

$$\begin{pmatrix} -0.74 \\ (0.23) \\ \\ -0.33 \\ (0.28) \end{pmatrix} \quad \begin{pmatrix} -0.45 & 0.51 \\ (0.22) & (0.21) \\ \\ 0.49 & -0.28 \\ (0.27) & (0.26) \end{pmatrix} \quad \begin{pmatrix} -0.32 & 0.36 \\ (0.23) & (0.24) \\ \\ 0.46 & -0.56 \\ (0.28) & (0.29) \end{pmatrix} \quad \begin{pmatrix} 0.5794 & 0.1251 \\ & \\ & 0.8553 \\ & \end{pmatrix}$$

Notes: The estimated threshold values are $(\hat{r}_1, \hat{r}_2) = (-0.7668, 1.5471)$.
Standard errors of the estimates are given in parentheses.

that minimize the AIC defined in equation (4.5). In performing the grid search, it is assumed that $r_1 \in [P_{10}(y_{t-d}), P_{45}(y_{t-d})]$ and $r_2 \in [P_{55}(y_{t-d}), P_{90}(y_{t-d})]$, where $P_\alpha(y_{t-d})$ denotes the empirical αth percentile of y_{t-d}. The grid search result indicates that the overall minimum AIC is achieved at $(\hat{r}_1, \hat{r}_2) = (-0.7668, 1.5471)$.

Summing up, we have specified a three-regime VETAR process with $p = 2$, $d = 1$, $\hat{r}_1 = -0.7668$ and $\hat{r}_2 = 1.5471$ for the time-series of $\mathbf{Z}_t = (\Delta\kappa_t^{(1)}, \Delta\kappa_t^{(2)})'$ under consideration. The least squares estimation results of the fitted three-regime VETAR process are provided in Table 2.

To guard against an incorrect model specification, we perform two diagnostic checks for the estimated model. The first check is based on the sample cross-correlation matrices (SCCMs) of the standardized residuals. Following Tsay (2010, p.407), we report in Table 3 the simplified SCCMs of the standardized residuals for lags 1 to 5. In each simplified SCCM, '+' indicates a

Table 3: Simplified sample cross-correlation matrices (SCCMs) for the standardized residuals of the estimated VETAR process

Lag	1	2	3	4	5
Simplified SCCM	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$

value greater than twice the estimated standard error, '$-$' denotes a value less than minus twice the estimated standard error and '$\cdot$' represents an insignificant value. It can be seen that no element in the SCCMs is significant, indicating that the estimated model adequately captures the auto- and cross-correlations that exist in the historical data. In the second check, we use the multivariate portmanteau test to examine if there is any remaining auto- or cross-correlation in the standardized residuals (Tsay, 2010, p.397). The value of the test statistic, which follows a chi-squared distribution with 16 degrees of freedom under the null hypothesis, is $Q_2(5) = 11.54$. The corresponding p-value is 0.7752, suggesting no significant auto- or cross-correlation remains in the standardized residuals. The results of both tests indicate that the estimated VETAR process is adequate.[3]

In Section 5.2, we will compare this fitted process with several other processes that are fitted to the same sequence of historical period effects.

5 Evaluating the Estimated Model

5.1 Reasonableness of the Projection Results

In this sub-section, we assess the reasonableness of the projections resulting from the estimated VETAR process. Figure 5 shows the sample paths of $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ and the threshold variable y_t simulated from the estimated VETAR process. It can be seen that the simulated values of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ progress naturally from the historical values. Also, as expected, mean-reversion can be observed in the simulated sample paths of y_t.

Table 4 shows the estimates $\mu^{(g)}(1)$ and $\mu^{(g)}(2)$, the steady-state rates of reduction in $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ in regime g. We have $\hat{\mu}^{(1)}(1) > \hat{\mu}^{(1)}(2)$ and $\hat{\mu}^{(3)}(1) < \hat{\mu}^{(3)}(2)$, which, as explained in Section 4.2, ensure that the resulting mortality forecasts are semi-coherent.

[3]Admittedly, some of the estimated parameters in Table 2 are not significant at a 5% significance level. We have re-estimated the model by restricting the insignificant parameters to zero. However, a diagnostic checking of the residuals revealed that the restricted model is not adequate. For this reason, the unrestricted model presented in Table 2 is used for the analyses presented in the rest this paper. For the readers' information, the parameter estimates and the results of the diagnostic checks for the restricted model are provided in the Appendix.

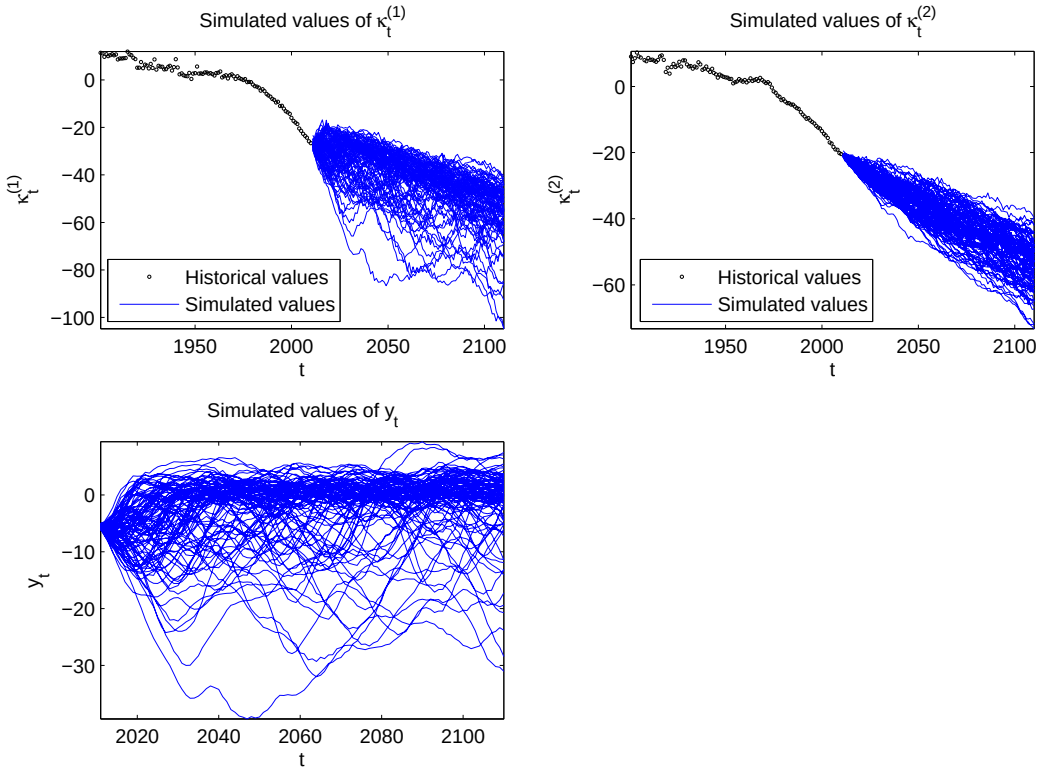

Figure 5: 100 sample paths of $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ and the threshold variable y_t simulated from the estimated VETAR model.

The property of semi-coherence can also be seen from Figure 6, which shows the simulated sample paths in each regime of the VETAR model, given certain fixed starting values. In particular, in regime 1 when $\kappa_t^{(1)} - \kappa_t^{(2)}$ is too small, the sample paths of $\kappa_t^{(1)} - \kappa_t^{(2)}$ tend to be upward sloping, thereby enlarging the expected difference between $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. Likewise, in regime 3 when $\kappa_t^{(1)} - \kappa_t^{(2)}$ is too large, the sample paths of $\kappa_t^{(1)} - \kappa_t^{(2)}$ tend to be downward sloping, so that the expected difference between $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ will shrink.

5.2 A Comparison with Alternative Modeling Approaches

In this sub-section, we compare our proposed modeling method with several alternative approaches. For a fair comparison, we only consider mortality models that do not involve cohort effects. The collection of models under consideration is summarized below.

- Model NC

  Model NC is built on equation (3.3), with the evolutions of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ being modeled by two random walks with independent innovations and different drift terms. This model yields non-coherent mortality forecasts.

- Model SC

Table 4: The estimated values of $\mu^{(g)}(1)$ and $\mu^{(g)}(2)$, the steady-state rates of reduction in $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ in the gth regime, for $g = 1, 2, 3$.

Regime (g)	$\hat{\mu}^{(g)}(1)$	$\hat{\mu}^{(g)}(2)$
1	-0.046367	-0.390444
2	-0.325994	-0.244268
3	-0.691088	-0.553801

Model SC is our proposed model, which is built on equation (3.3), with the evolutions of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ being modeled by the VETAR process detailed in Section 4. The mortality forecasts generated by this model are semi-coherent.

- Model FC1

  Model FC1 is the augmented common factor model introduced by Li and Lee (2005). It can be expressed as

  $$\ln(m(x,t,i)) = \alpha_x^{(i)} + \beta_x \kappa_t + \beta_x^{(i)} \kappa_t^{(i)}, \qquad i = 1, 2, \tag{5.1}$$

  where $\alpha_x^{(i)}$, β_x and $\beta_x^{(i)}$ are age-specific parameters, and κ_t and $\kappa_t^{(i)}$ are time-varying parameters. The evolution of κ_t is captured by a random walk with drift,

  $$\kappa_t = c + \kappa_{t-1} + \epsilon_t,$$

  while that of $\kappa_t^{(i)}$ is modeled by a first order autoregressive process,

  $$\kappa_t^{(i)} = \phi_0^{(i)} + \phi_1^{(i)} \kappa_{t-1}^{(i)} + \zeta_t^{(i)}, \quad i = 1, 2, \tag{5.2}$$

  where c and $\phi_0^{(i)}$ are constants, $\phi_1^{(i)}$ is another constant whose absolute value is strictly less than 1, and ϵ_t and $\zeta_t^{(i)}$ are mutually independent normal variables with zero mean, constant variance and no serial correlation. This model implies fully-coherent mortality forecasts.

- Model FC2

  Model FC2 is the two-population extension of the Lee-Carter model considered by Cairns et al. (2011). As previously described, this model is built on equation (3.3), with the evolutions of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ being modeled equations (3.4) to (3.6). This model also produces fully-coherent mortality forecasts.

- Model FC3

  Model FC3 is identical to Model FC2, except that the evolutions of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are jointly modeled by a special vector error correction model (VECM) that is configured to ensure that the resulting forecasts are fully coherent. Further details regarding Model FC3 can be found in the paper by Zhou et al. (2014).

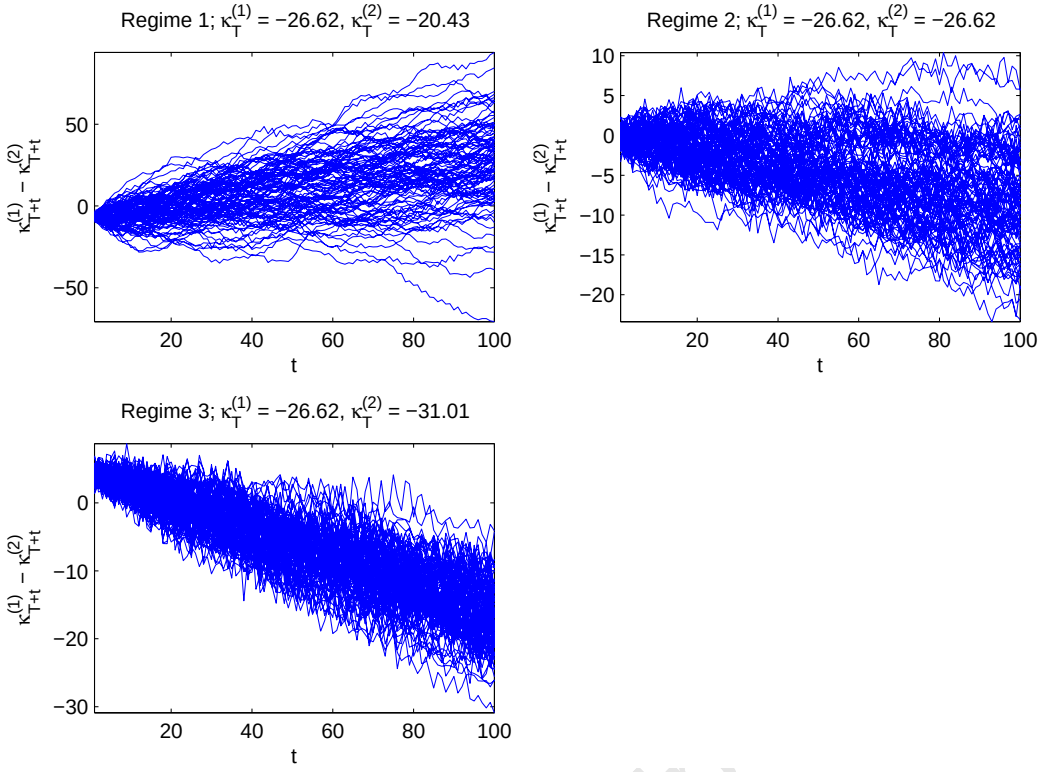

Figure 6: 100 simulated sample paths of $\kappa_t^{(1)} - \kappa_t^{(2)}$, $t = T+1, \ldots, T+100$, in each of the three regimes of the estimated VETAR model. The initial values (i.e., $\kappa_T^{(1)}$ and $\kappa_T^{(1)}$) used are displayed on the top of the diagrams.

Note that Models NC, SC, FC2 and FC3 are built on the same base model structure (equation (3.3)). Therefore, we can evaluate their relative performance by comparing the goodness-of-fit of the period effect processes encompassed in these models. In Table 5 we display the AIC values of the estimated period effect processes in Models NC, SC, FC2 and FC3.[4] The VETAR process yields the smallest AIC value, which indicates that it provides the best fit to the historical values of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, taking into account the number of parameters contained in the process.

We now expand our model evaluation exercise to include Model FC1, which is constructed from a different base model structure. First, we perform an additional goodness-of-fit analysis, which allows us to compare all five candidate models. The analysis is based on one-step ahead in-sample forecasts, an approach that is often used in evaluating the goodness-of-fit of structural time-series models (see, e.g., Harvey, 1989). In particular, we consider the mean absolute percentage error

[4]The AIC values for the period effect processes in Models NC, FC2 and FC3 are computed by the following formula: $AIC = n_d \ln(|\hat{\Sigma}|) + 2n_p$, where n_d is the effective number of observations, n_p is the number of parameters and $\hat{\Sigma}$ is the covariance matrix of the estimated residuals. On the basis of this definition, a model with a smaller AIC value is preferred.

Table 5: The AIC values of the estimated period effect processes in Models NC, SC, FC2 and FC3.

	Period effect process(es)	AIC
Model NC	Two independent random walks	42.741
Model SC	A VETAR process	−38.772
Model FC2	A random walk plus a first order autoregression	35.266
Model FC3	A specially parameterized VECM	15.237

Table 6: The values of the MAPE produced by Models NC, SC, FC1, FC2 and FC3.

	Model NC	Model SC	Model FC1	Model FC2	Model FC3
MAPE	0.0437	0.0412	0.0413	0.0436	0.0425

(MAPE):

$$\text{MAPE} = \frac{1}{31 \times 109 \times 2} \sum_{x=55}^{85} \sum_{t=1902}^{2010} \sum_{i=1}^{2} \left| \frac{\tilde{m}(x,t,i) - \text{E}(m(x,t,i)|\mathcal{F}_{t-1})}{\tilde{m}(x,t,i)} \right|,$$

where $\tilde{m}(x,t,i)$ is the realized value of $m(x,t,i)$ and $\text{E}(m(x,t,i)|\mathcal{F}_{t-1})$ is the predicted value of $m(x,t,i)$ on the basis of the estimated model and the information up to and including time $t-1$. The MAPE for each of the five candidate models is reported in Table 6. Model SC produces the smallest MAPE, indicating that it yields a better goodness-of-fit compared to the other four candidate models (including Model FC1).

Next, we evaluate the properties of the forecasts generated from all five candidate models. Figure 7 shows the fan charts for the projection of future log central death rates at age 75. Each fan chart shows the 10% prediction interval with the heaviest shading, surrounded by the 20%, 30%, ..., 90% prediction intervals with progressively lighter shadings.[5] The solid line in the fan indicates the central (median) projection. The full- and semi-coherent forecasts differ from the non-coherent forecasts in that at least one of the two central projected trajectories are pulled either upward or downward so that they do not diverge indefinitely. The degrees of forecast uncertainty produced by the five models are quite similar, but Model SC is the only model that gives asymmetric prediction intervals. The results for other ages are very similar and are therefore not shown.

The distinctions among different forms of coherence are more apparent in the forecasts of future mortality differentials. Figure 8 displays the fan charts of $\ln(m(75,t,1)) - \ln(m(75,t,2))$

[5]The $(1-\alpha)$% prediction interval of $\ln(m(x,t,i))$ is computed by the $\alpha/2$th and $(1-\alpha/2)$th percentiles of the simulated empirical distribution of $\ln(m(x,t,i))$.

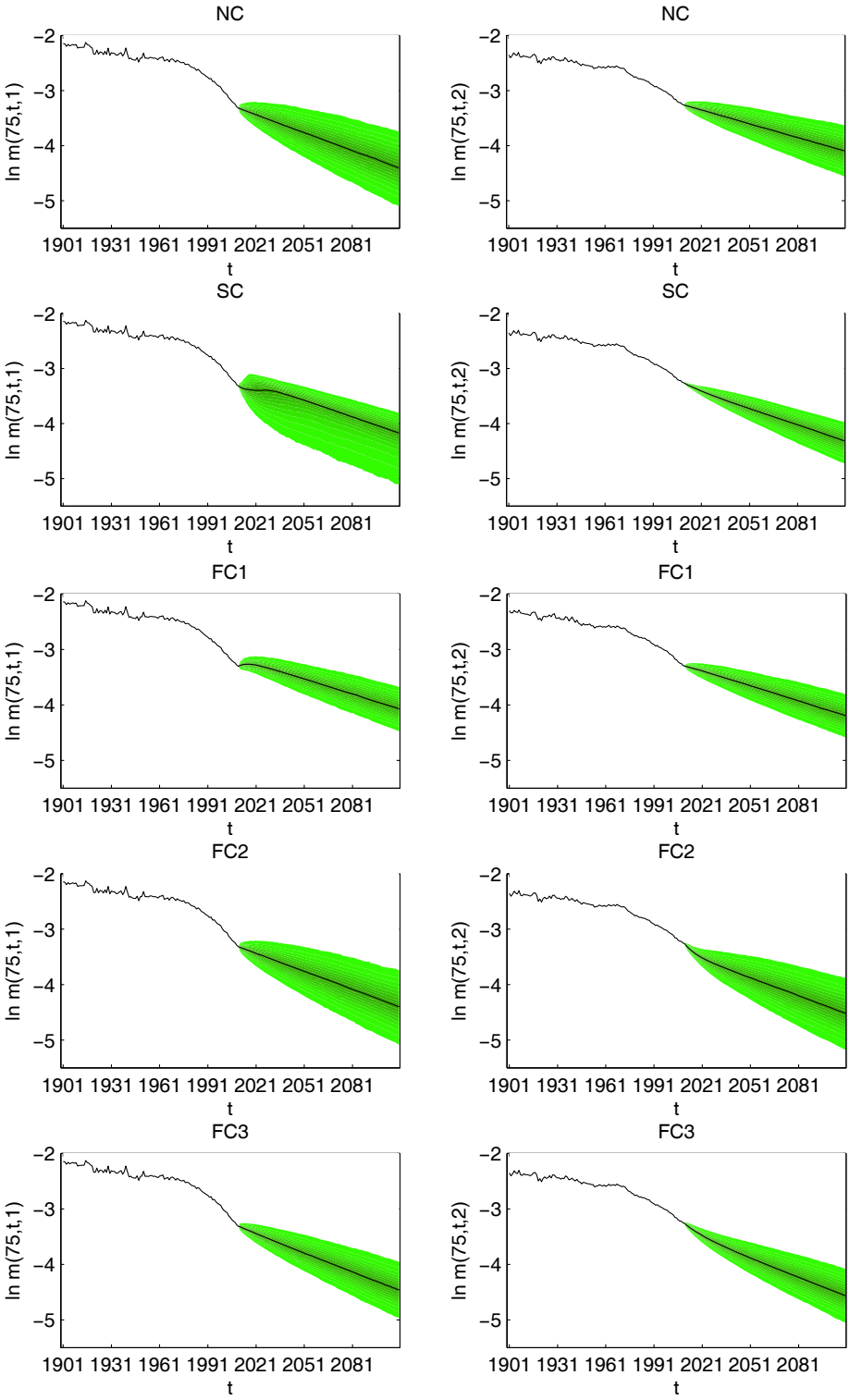

Figure 7: Fan charts showing the mean and interval forecasts of $\ln m(75,t,1)$ and $\ln m(75,t,2)$, generated from Models NC, SC, FC1, FC2 and FC3.

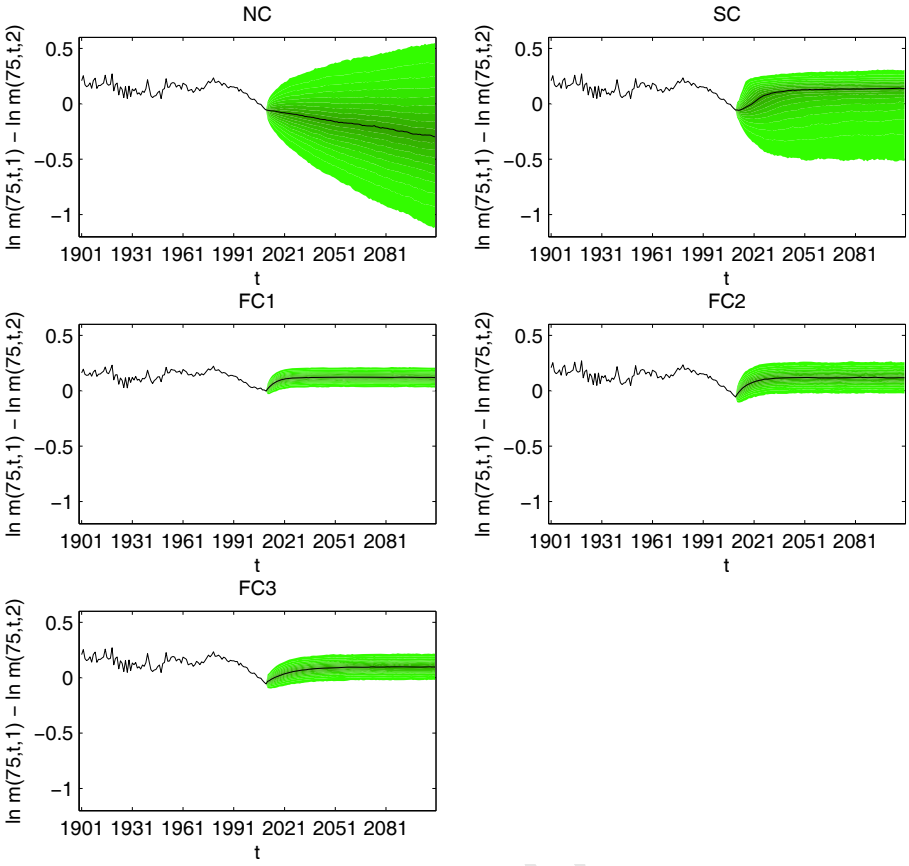

Figure 8: Fan charts showing the mean and interval forecasts of $\ln m(75,t,1) - \ln m(75,t,2)$, generated from Models NC, SC, FC1, FC2 and FC3.

generated from the non-, semi- and fully-coherent models under consideration. The indefinite divergence resulting from the non-coherence assumption can be seen clearly from the downward sloping linear central projections produced by Model NC. The problem of an indefinite divergence is avoided in both semi- and fully-coherent modeling approaches. For Models SC, FC1, FC2 and FC3, the central projections as well as the darker shades of the fans are (approximately) flat over the long run.

The fan charts in Figure 8 also indicate some shortcomings of the fully-coherent modeling approaches. First, for the central projections generated from the fully-coherent models (Models FC1, FC2 and FC3), the reversion to the long-term equilibrium is very quick, beginning immediately at the forecast origin. The abrupt change in the trend of the mortality differential is not easy to justify. By contrast, the central projection produced by the semi-coherent model (Model SC) appears to be a more natural extension of the historical values. Second, the fully-coherent models appear to understate the level of uncertainty (i.e., population basis risk). Specifically, they provide almost no provision for the downside risk (the risk that the mortality differential becomes more negative). In stark contrast, Model SC is advantageous of being able to capture

the potential asymmetry in population basis risk, producing a fan chart that also covers scenarios in which the mortality differential becomes even more negative in future.

6 The Impact on Longevity Risk Hedging

In this section, we apply the models described in Section 5.2 to assessing the effectiveness of several hypothetical longevity hedges. Our primary goal is to investigate the potential effect of *hypothesis risk*, the risk arising from the possibility that the assumed coherence hypothesis turns out to be wrong. Of our particular interest is how much hedge effectiveness may be lost if the usual assumption of full-coherence does not hold.

6.1 The Set-up

Let us suppose that the liability being hedged is a pension plan that pays each pensioner \$1 at the end of each year until death or reaching age 85, whichever is the earliest. It is assumed that no pension is payable beyond age 85, because the upper limit of the age range to which the models are fitted is 85. This assumption may be waived if one assumes instead a parametric curve to extrapolate death probabilities beyond age 85.

It is further assumed that all individuals covered by the pension plan are aged 60 on the valuation date. We consider only a single cohort of individuals, because it may not be appropriate to value a pension plan involving multiple birth cohorts with the models under consideration, which are based on the basic Lee-Carter structure that does not take cohort effects into account. The pensioners are assumed to have exactly the same mortality experience as English and Welsh males born in the same year. Small sample risk (a.k.a. sampling risk or Poisson risk) is not taken into consideration.

We consider the following three static hedge portfolios, all of which are composed of securities that are linked to the future mortality of the United States male population.

- Hedge Portfolio 1

  The first hedge portfolio is composed of a longevity bond that is issued on the valuation date and has a maturity of 25 years. Similar to the one that was jointly announced by BNP Paribas and the European Investment Bank in 2004 (see Blake et al., 2006), the longevity bond we consider is an annuity bond (i.e., there is no principal repayment) with annual coupon payments. Each coupon payment is proportional to the realized survival rate of a specific cohort of individuals at the time when the coupon is paid. It is assumed that the coupons are linked to the cohort of United States males aged 60 on the date of inception.

- Hedge Portfolio 2

The second hedge portfolio is composed of five q-forwards, maturing 6, 11, 16, 21 and 26 years from the valuation date, respectively. Each q-forward is a zero-coupon swap that exchanges on the maturity date a fixed amount for a random amount that is proportional to the single-year death probability of the reference population at a certain reference age, realized during the calendar year immediately before the maturity date. It is assumed that the five q-forwards in the hedge portfolio are all linked to the cohort of United States males aged 60 on the date of valuation, so that their reference ages are respectively 65, 70, 75, 80 and 85. We refer readers to Coughlan (2009) for further details concerning q-forwards.

- Hedge Portfolio 3

  The third hedge portfolio is composed of two S-forwards, both of which are issued on the valuation date. Each S-forward is a zero-coupon swap that exchanges on the maturity date a fixed amount for a random amount that is proportional to the realized probability that an individual who belongs to a certain birth cohort in the reference population survives from the inception date to the maturity date. The two S-forwards have maturities of 5 and 20 years, respectively, and are linked to the cohort of United States males aged 60 on the inception date. Further details about S-forwards can also be found in Coughlan (2009).

It is clear that population basis risk exists in all three hypothetical longevity hedges, as the hedging instruments and the liability being hedged are linked to different populations.

To measure hedge effectiveness, the following metric is used:

$$R = 1 - \frac{\sigma_h^2}{\sigma_u^2},$$

where σ_u^2 denotes the variance of the present values of all payments made by the pension plan when there is no longevity hedge, and σ_h^2 represents the variance of the present values of all net cash flows when a longevity hedge is in place. A value of R that is close to one indicates a high hedge effectiveness, while a value of R that is close to zero indicates the opposite. This metric has also been used by other researchers, including Cairns (2013b), Cairns et al. (2014) and Li and Hardy (2011). Note that the calculation of R does not require any pricing information, that is, the time-0 price of the longevity bond in Hedge Portfolio 1, the fixed legs of the q-forwards in Hedge Portfolio 2 and the fixed legs of the S-forwards in Hedge Portfolio 3. These quantities are therefore not computed.

In our calculations, the date of valuation is set to January 2011, because the last year for which the mortality data are available is 2010. It is assumed that the interest rate at which cash flows are discounted is 2%.

6.2 The Best Achievable Hedge Effectiveness

We first study the best achievable hedge effectiveness: the maximum value of R that a longevity hedge can possibly achieve irrespective of how the hedge is calibrated. The best achievable hedge

Table 7: The best attainable hedge effectiveness under different model assumptions.

	Model NC	Model SC	Model FC1	Model FC2	Model FC3
Hedge Portfolio 1	0.0003	0.7610	0.8016	0.8089	0.8227
Hedge Portfolio 2	0.0010	0.7682	0.8181	0.8995	0.8139
Hedge Portfolio 3	0.0003	0.7649	0.8058	0.8572	0.8170

effectiveness is calculated with the following procedure.

1. Simulate 10,000 sample paths of mortality rates of from 2011 to 2025 from the assumed two-population stochastic mortality model, fitted to the historical mortality data from the United States and English and Welsh populations.

2. On the basis of the 10,000 sample paths, calculate the value of σ_u^2.

3. On the basis of the 10,000 sample paths, calculate the minimum attainable value of σ_h^2 as

$$\min_{w_1,\ldots,w_k} \text{Var}\left(L - \sum_{i=1}^{k} w_i H_i\right), \tag{6.1}$$

   where k is the total number of instruments in the hedge portfolio, w_i is the notional amount of the ith hedging instrument, and L and H_i are the random present values of the liability being hedged and the ith hedging instrument, respectively.

4. Using the values of σ_u^2 and σ_h^2 calculated in the previous two steps, calculate the maximum attainable value of R.

For Hedge Portfolio 1 in which there is only one instrument, the minimum of σ_h^2 is attained when

$$w_1 = \frac{\text{Cov}(L, H_1)}{\text{Var}(H_1)}.$$

For Hedge Portfolios 2 and 3 in which there are multiple instruments, the solution to the minimization in expression (6.1) can be obtained by numerical methods.

Table 7 displays the calculated best attainable hedge effectiveness under different model assumptions. The results indicate that if the actual underlying model is Model NC, then none of the longevity hedges yield a meaningful hedge effectiveness. By contrast, if the actual underlying model is fully-coherent, then there is a potential to achieve a hedge effectiveness of 80 to 90%. The hedge effectiveness reduces to approximately 76% if the actual underlying model is Model SC, which is only semi- rather than fully-coherent. The significance of hypothesis risk can be seen from the variation in the hedge effectiveness produced under different coherence hypotheses.

6.3 Hedge Effectiveness Arising from a Specific Longevity Hedging Strategy

The use of equation (6.1) to estimate the required notation amount(s) is computationally intensive. In practice, it is more feasible to calibrate a longevity hedge with a hedging strategy that requires less computation effort.

We now investigate the potential impact of hypothesis risk when a specific strategy is used to calibrate the longevity hedge. The strategy considered is the method of key q-duration proposed by Li and Luo (2012). The main idea behind this method is to match the sensitivities of the values of the hedge portfolio and the liability being hedged to small changes in the underlying mortality probabilities.[6]

To keep the analysis simple, we focus on Hedge Portfolio 3, which is composed of five q-forwards that are respectively linked to the death probabilities for age 65 in 2016, age 70 in 2021, age 75 in 2026, age 80 in 2031 and age 85 in 2036.

The first step in implementing the method of key q-duration is to define the key mortality rates that collectively represent the underlying mortality curve. We set the key mortality rates to the death probabilities to which the five q-forwards in the hedge portfolio are linked.

Consider a portfolio with a random time-0 value $P(\mathbf{q})$, which depends on the underlying mortality curve $\mathbf{q}$. The jth key q-duration of the portfolio is defined as

$$KQD(P(\mathbf{q}), j) = \lim_{\delta(j)\to 0} \frac{P(\tilde{\mathbf{q}}(\delta(j))) - P(\mathbf{q})}{\delta(j)},$$

where $\delta(j)$ denotes a small change in the jth key mortality rate and $\tilde{\mathbf{q}}(\delta(j))$ represents the underlying mortality curve that is subject to a change of $\delta(j)$ in the jth key mortality rate. The estimation of key q-duration was explained in detail by Li and Luo (2012). Each key q-duration measures the sensitivity of the time-0 value of the portfolio with respect to a small change in a certain key mortality rate. Our goal is to match the key q-durations of the liability being hedged and the hedge portfolio.

Let us first consider the ideal case in which both the liability being hedged and the portfolio of q-forwards are linked to the same underlying mortality curve $\mathbf{q}$ (i.e., there is no population basis risk). To match the key q-durations of the liability being hedged and the portfolio of q-forwards, we require a notional amount of

$$w_j = \frac{KQD(L(\mathbf{q}), j)}{KQD(H_j(\mathbf{q}), j)}$$

for the jth q-forward contract, where $L(\mathbf{q})$ and $H_j(\mathbf{q})$ respectively represent the time-0 values of L and H_j when the underlying mortality curve is $\mathbf{q}$.

In real life, population basis risk does exist. Let $\mathbf{q}_1$ and $\mathbf{q}_2$ respectively represent the underlying mortality curves for the populations associated with the liability being hedged (population 1) and

[6] Similar hedging strategies have been proposed by Cairns (2011), Cairns (2013b), Lin and Tsai (2014), Tan et al. (2014) and Wang et al. (2010).

the hedging portfolio (population 2). In our application, $\mathbf{q}_1$ and $\mathbf{q}_2$ are the mortality curves for the cohort of individuals, aged 60 in 2010, from the male populations of England and Wales and the United States, respectively. To match the key q-durations of the liability being hedged and the portfolio of q-forwards, the notional amount of the jth q-forward is

$$w_j = \frac{KQD(L(\mathbf{q_1}), j)}{KQD(H_j(\mathbf{q_1}), j)} = \frac{KQD(L(\mathbf{q_1}), j)}{KQD(H_j(\mathbf{q_2}), j)} \frac{\partial q^{(1)}_{x_j,t_j}}{\partial q^{(2)}_{x_j,t_j}},$$

where $q^{(i)}_{x_j,t_j}$ is the jth key mortality rate for population i, and x_j and t_j are the age and year that are associated with the jth key mortality rate, respectively. The partial derivative on the right-hand-side of the equation may be regarded as an adjustment term for population basis risk.

Throughout this analysis, the values of w_j for $j = 1, \ldots, 5$ are computed on the basis of Model FC1 (the augmented common factor model). Under Model FC1, the adjustment term is given by

$$\frac{\partial q^{(1)}_{x,t}}{\partial q^{(2)}_{x,t}} = \frac{q^{(1)}_{x,t}(1 + 0.5m^{(2)}_{x,t})^2 A(x,t,1)}{q^{(2)}_{x,t}(1 + 0.5m^{(1)}_{x,t})^2 A(x,t,2)}, \tag{6.2}$$

where

$$A(x,t,i) = \beta_x c + \beta^{(i)}_x (\phi^{(i)}_1)^{t-t_0} (\phi^{(i)}_0 + (\phi^{(i)}_1 - 1)\kappa^{(i)}_{t_0}), \quad i = 1, 2,$$

and $t_0 = 2010$ denotes the calendar year immediately before the longevity hedge is set up. A proof of the above equation can be found in the paper by Li and Hardy (2011).

We evaluate hedge effectiveness by using the models described in Section 5.2 to simulate future mortality scenarios. In this way, we can assess the performance of the longevity hedge – calibrated on the basis of Model FC1 – when the actual underlying model is also Model FC1 or not.

The results of this analysis are presented in Table 8. As expected, the hedge effectiveness is the highest (66%) when the simulation model is Model FC1, which is identical to the model on which the key q-durations are based. The hedge effectiveness (63-55%) obtained when the simulation model is Model FC2 or FC3 is comparably high, indicating that the hedging strategy is reasonably robust with respect to model risk, provided that the simulation model is built on the hypothesis of full coherence. Nevertheless, the impact of hypothesis risk is significant. When evaluated using mortality scenarios simulated from the semi-coherent model, the strategy only yields a hedge effectiveness of approximately 30%. When the simulation model is non-coherent, the resulting hedge effectiveness is negative, indicating that the risk exposed to the hedger is increased rather than reduced. The negative hedge effectiveness is largely because the adjustment term in equation (6.2), specified by Model FC1, is very wrong when the true underlying model is non-coherent.

Table 8: The hedge effectiveness produced by the longevity hedge that is calibrated by the method of key q-durations, with all key q-durations calculated on the basis of Model FC1.

Simulation Model	Model NC	Model SC	Model FC1	Model FC2	Model FC3
Hedge Effectiveness	−0.2677	0.3065	0.6635	0.6611	0.6256

7 The Impact on Longevity Risk Pricing

7.1 The Illustrative Security

In this section, we investigate the impact of various coherence hypotheses on pricing by considering a hypothetical security whose structure is similar to that of the Kortis bond issued by Swiss Re in December 2010. Let us now briefly review the specification of the Kortis bond.[7]

The Kortis bond was sold at par with a principal of $50 million. Similar to typical catastrophe bonds, the Kortis bond pays quarterly coupons at a floating rate that is higher than the risk-free interest rate. In particular, the coupon rate was set to 5.0% above the London Interbank Offer Rate (LIBOR). The premium spread of 5.0% may be regarded as the investors' reward for taking on the risk associated with the bond.

The principal is scheduled to be repaid at maturity in January 2017. The amount of principal repayment is not guaranteed, but depends on the principal reduction factor (PRF), which is defined as

$$PRF = \max\left(\min\left(\frac{LDIV(2016) - ap}{ep - ap}, 1\right), 0\right),$$

where $LDIV(t)$ stands for the longevity divergence index value (LDIV) for year t, and ap and ep denote the attachment and exhaustion points, respectively. As such, the principal will be repaid in full if $LDIV(2016) < ap$, reduced if $ap < LDIV(2016) < ep$, and exhausted if $LDIV(2016) > ep$. The values of ap and ep are set to 3.4% and 3.9%, respectively.

The LDIV tracks the divergence in mortality rates between English and Welsh male population (population $i = 1$) and the United States male population (population $i = 2$). The LDIV for year t is calculated as

$$LDIV(t) = I_t^{(1)} - I_t^{(2)},$$

where

$$I_t^{(i)} = \frac{1}{1 + x_2^{(i)} - x_1^{(i)}} \sum_{x=x_1^{(i)}}^{x_2^{(i)}} \left(1 - \left(\frac{m_{x,t}^{(i)}}{m_{x,t-8}^{(i)}}\right)^{1/8}\right)$$

[7]The information about the Kortis bond is taken from Blake et al. (2013) and Hunt and Blake (2013).

is the average annualized mortality improvement rate observed across ages $x_1^{(i)}$ to $x_2^{(i)}$ over the period from year $t-8$ to t. The age ranges used for English and Welsh and the United States populations are $x_1^{(1)} = 75$ to $x_2^{(1)} = 85$ and $x_1^{(2)} = 55$ to $x_2^{(2)} = 65$, respectively.

The Kortis bond was designed to match the specific risk management needs of Swiss Re, which reinsures life insurance products from the United States and life annuity products from the United Kingdom. These two portfolios create an imperfect natural hedge, which is least effective if the future mortality of the United States population turns out to be too high and the future mortality of the United Kingdom population turns out to be too low. Given the bond's design, the adverse financial outcome arising from an overly large mortality differential between the populations of the United Kingdom and United States can be offset by a reduced principal repayment to the investors of the bond.

The structure of the hypothetical bond we consider is the same as that of the Kortis bond, except that its PRF is linked instead to the survival rate divergence index (SRDI), whose value in year t is given by

$$SRDI_t = S_t^{(1)} - S_t^{(2)},$$

where

$$S_t^{(i)} = \prod_{j=1}^{t} e^{-m_{65+j,2010+j}^{(i)}}$$

represents the approximate t-year survival rate for the cohort of individuals (in population i) aged 65 in year 2010.[8] We make this adaptation because the LDIV in the original Kortis bond involves individuals with different years of birth, so that the uncertainty surrounding the principal reduction depends on the uncertainty associated with the cohort effect for each population and the association between the cohort effects of the two populations. These two features, however, are not incorporated in the Lee-Carter structure on which our models are based. The adaptation waives the need for considering cohort effects, but does not affect the key issue – the impact of various coherence hypotheses – that we intend to illustrate.

Other than replacing the LDIV with the SRDI, we also modify the inception and maturity dates to reflect the availability of more recent mortality data. The attachment and exhaustion points are chosen to commensurate with the value of the SRDI at inception. The parameters of the hypothetical bond are summarized in Table 9.

7.2 The Pricing Methodology

Our goal is to determine the premium spread over LIBOR on the basis of the various models described in Section 5.2. Although this goal can be accomplished by various pricing methodologies,

[8] The approximation is exact if the force of mortality between two consecutive integer ages is constant.

Table 9: Parameters of the hypothetical bond used for illustration purposes.

Inception date	December 2010
Maturity date	January 2019
Value to which the PRF is linked	$SRDI(2018)$
Attachment point (ap)	0.045
Exhaustion point (ep)	0.055
Coupon frequency	Quarterly
Coupon rate	3-month LIBOR plus a premium spread

we consider Lane's (2000) risk-cubic pricing methodology, which has been used in practice by the Lane Financial Corporation to price various catastrophe securities. This pricing method has also been used by academics, including Wills and Sherris (2010) and Chen and Cummins (2010).

In the risk-cubic pricing methodology, the premium spread is computed as the sum of the expected loss (EL) and the expected excess return (EER); that is,

$$\text{Premium spread} = EL + EER. \tag{7.1}$$

The expected excess loss is modeled by a Cobb-Douglas function as

$$EER = \gamma_1 (PFL)^{\gamma_2} (CEL)^{\gamma_3}, \tag{7.2}$$

where PFL is the annualized probability of the first dollar loss, CEL is the conditional expected loss (i.e., the expected loss severity given that a loss has occurred) and γ_1, γ_2 and γ_3 are model parameters that are calibrated to historical securities transaction data.

To implement the risk-cubic pricing method, we first estimate parameters γ_1, γ_2 and γ_3 by fitting the following log-linear regression to historical securities transaction data:

$$\ln(EER) = \ln(\gamma_1) + \gamma_2 \ln(PFL) + \gamma_3 \ln(CEL).$$

The transaction data used include the values of EER, PFL and CEL that are associated with 18 mortality-linked securities transacted over the period of 2003 Q1 to 2010 Q1. The data are obtained from trade notes published by the Lane Financial Corporation. The resulting estimates of γ_1, γ_2 and γ_3 are displayed in Table 10.

Then the premium spread for the hypothetical bond is calculated with the following procedure.

1. Simulate 10,000 mortality paths from the assumed two-population mortality model.
2. For each simulated sample mortality path, calculate $SRDI(2018)$ and the principal reduction.

Table 10: Estimates of the parameters in the Cobb-Douglas function for modeling the expected excess return.

Parameter	$\ln(\gamma_1)$	γ_2	γ_3
Estimate	3.0268	1.0661	1.4119
Standard error	2.3907	0.2912	1.6516

Note: The value of R^2 is 0.553 and the p-value of the F-test for testing $\gamma_2 = \gamma_3 = 0$ is 0.00536.

3. Calculate the annualized values of PFL, EL and CEL as follows:

$$\begin{aligned} PFL &= \frac{\text{Number of mortality paths with } SRDI(2018) > ap}{10,000 \times 8}; \\ EL &= \frac{\text{Average principal reduction}}{8}; \\ CEL &= \frac{EL}{PFL}. \end{aligned}$$

4. Compute EER and the premium spread using equations (7.1) and (7.2), and the parameter estimates in Table 10.

7.3 The Baseline Pricing Results

The portion labeled 'baseline results' in Table 11 displays the estimated premium spread for the previously described hypothetical bond under each model assumption. For readers' information, the corresponding estimated values of EL, PFL, CEL and EER are also shown. It can be seen that the impact of hypothesis risk on pricing is highly significant.

For all three fully-coherent models, the implied values of PFL and PE for the hypothetical bond are zero: the bond's principal will always be repaid in full. Therefore, when these fully-coherent models are assumed, the hypothetical bond is literally risk-free and the investors should receive no spread above the risk-free interest rate. Note that when $PFL = PE = 0$, the values of CEL and EER are undefined. In this case, the risk-cubic pricing methodology does not give a meaningful premium spread estimate.

Although both Models NC and SC imply non-zero values of PFL and PE, the premium spreads calculated from these two models are highly different. The premium spread computed from Model NC is more than three times that implied by Model SC.

The impact of hypothesis risk can be better understood by observing the calculated density functions of $SRDI(2018)$, depicted in the upper panel of Figure 9. The distributions of $SRDI(2018)$ calculated from the fully-coherent models have small dispersions, leaving no probability density beyond the attachment point $ap = 0.045$. The hypothetical bond is thus riskless if it is evaluated by any of these fully-coherent models. However, this conclusion would be entirely wrong if it turns out that the hypothesis of full coherence does not hold. The distributions of

$SRDI(2018)$ computed from the non- and semi-coherent models are much more dispersed, with some probability density beyond the attachment point. The hypothetical bond is therefore no longer riskless under the two alternative coherence hypothesis. Over the range of $ap = 0.045$ to $ep = 0.055$, the magnitudes of the density functions calculated from Models NC and SC are quite different, offering an explanation as to why these two models yield rather different premium spread estimates.

7.4 Sensitivity Tests

Depending on the model assumption, the central estimate of $SRDI(2018)$ ranges from 0.06% to 1.70%. The hypothetical bond with $ap = 0.045$ and $ep = 0.055$ can therefore be considered as out-of-the-money. We now investigate how the pricing results may change if the security is more out-of-the-money (with $ap = 0.055$ and $ep = 0.065$) or less out-of-the-money (with $ap = 0.035$ and $ep = 0.045$).

The portions labeled 'Sensitivity test 1' and 'Sensitivity test 2' in Table 11 respectively show the pricing results that are based on the two alternative sets of attachment and exhaustion points. As in the baseline results, the values of PFL and EL implied by the fully-coherent models are zero. Hypothesis risk seems to be particularly severe when $ap = 0.055$ and $ep = 0.065$. For this combination of attachment and exhaustion points, the premium spread implied by Model NC is more than 24 times that computed from Model SC. Again, this outcome can be explained by considering the density functions of $SRDI(2018)$ (Figure 9, the upper panel). It can be seen that over the range of 0.055 to 0.065, the density function generated by Model SC is close to zero, but that generated by Model NC still has an observable magnitude.

One criticism of the original Kortis bond is that its maturity was too short to reflect the long-term nature of the underlying risk (Hunt and Blake, 2014). Examining how the pricing results change if the time to maturity is lengthened is therefore warranted. We now re-perform the pricing work by considering two alternative maturity dates, January 2027 and January 2031, which represent times to maturity of 16 and 20 years, respectively.

The portions labeled 'Sensitivity test 3' and 'Sensitivity test 4' in Table 11 respectively display the pricing results that are based on the two alternative maturity dates. For Models NC and SC, the estimated premium spreads are significantly larger than before, indicating the increased likelihood of reaching the attachment and detachment points as the projection of SRDI is made further into the future. Nevertheless, the fully-coherent models still imply zero values of PFL and EL, suggesting that the assumption of full coherence may significantly understate the possible range of mortality differentials.

The middle and lower panels of Figure 9 display the calculated density functions of $SRDI(2026)$ and $SRDI(2030)$, respectively. These two random index values respectively determine the principal reductions of the bonds maturing in January 2027 and January 2031. The

Table 11: The calculated values of EL, PFL, CEL, EER and the premium spread on the basis of the baseline parameters and four alternative combinations of maturity and moneyness.

	Model NC	Model SC	Model FC1	Model FC2	Model FC3
Baseline results					
EL	0.19 %	0.07 %	0.00 %	0.00 %	0.00 %
PFL	0.44 %	0.20 %	0.00 %	0.00 %	0.00 %
CEL	44.68 %	35.47 %	–	–	–
EER	2.01 %	0.64 %	–	–	–
Premium spread	2.20 %	0.71 %	–	–	–
Sensitivity test 1: $ap = 0.055$, $ep = 0.065$; other things equal					
EL	0.04 %	0.00 %	0.00 %	0.00 %	0.00 %
PFL	0.08 %	0.01 %	0.00 %	0.00 %	0.00 %
CEL	47.05 %	18.78 %	–	–	–
EER	0.35 %	0.01 %	–	–	–
Premium spread	0.39 %	0.02 %	–	–	–
Sensitivity test 2: $ap = 0.035$, $ep = 0.045$; other things equal					
EL	0.98 %	0.57 %	0.00 %	0.00 %	0.00 %
PFL	1.79 %	1.15 %	0.00 %	0.00 %	0.00 %
CEL	54.85 %	49.50 %	–	–	–
EER	12.14 %	6.51 %	–	–	–
Premium spread	13.12 %	7.08 %	–	–	–
Sensitivity test 3: Maturity date = January 2027; other things equal					
EL	1.66 %	1.06 %	0.00 %	0.00 %	0.00 %
PFL	1.91 %	1.24 %	0.00 %	0.00 %	0.00 %
CEL	87.08 %	85.25 %	–	–	–
EER	24.92 %	15.32 %	–	–	–
Premium spread	26.58 %	16.38 %	–	–	–
Sensitivity test 4: Maturity date = January 2031; other things equal					
EL	1.33 %	0.78 %	0.00 %	0.00 %	0.00 %
PFL	1.51 %	0.88 %	0.00 %	0.00 %	0.00 %
CEL	88.27 %	88.45 %	–	–	–
EER	19.77 %	11.17 %	–	–	–
Premium spread	21.10 %	11.95 %	–	–	–

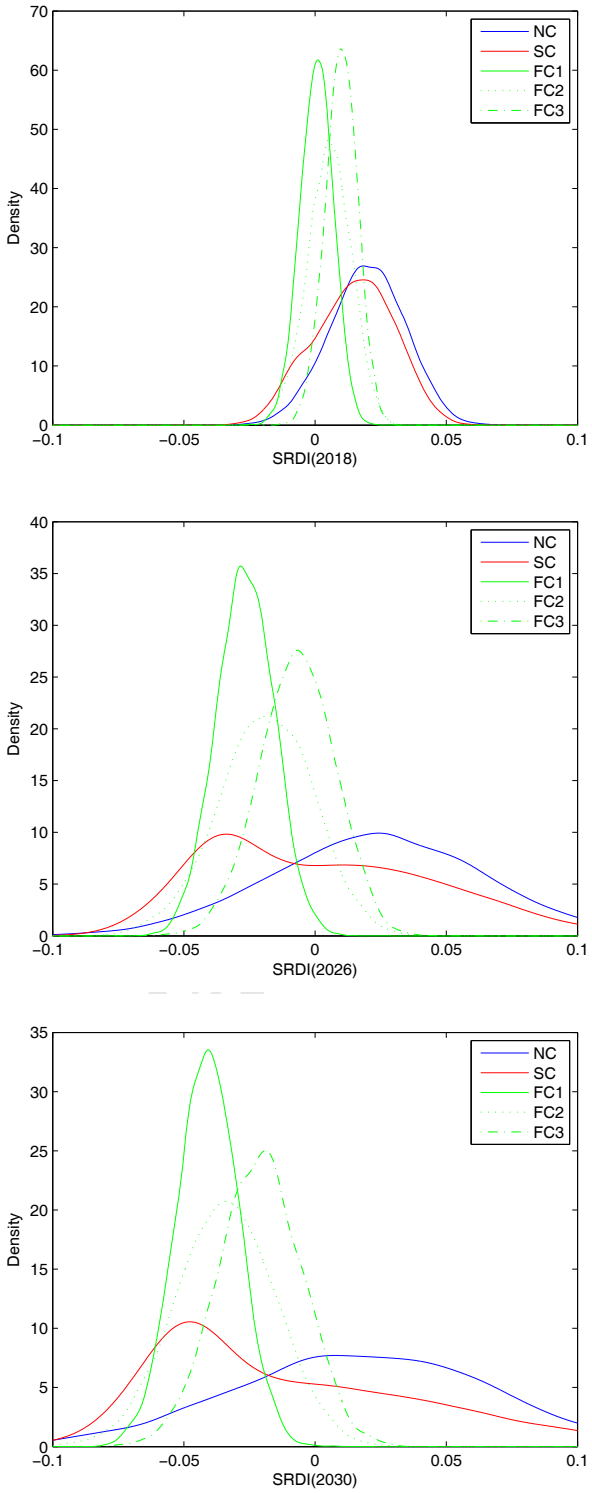

Figure 9: Density functions of $SRDI(2018)$, $SRDI(2026)$ and $SRDI(2030)$, simulated from Models NC, SC, FC1, FC2 and FC3.

dispersions of the distributions generated from the fully-coherent models are still not sufficient to yield any probability density over the range of 0.045 to 0.055. The means of the distributions computed from Model SC are roughly in line with those based on the fully-coherent models. What particularly make the premium spreads estimated from Model SC significant are the asymmetry and leptokurtosis possessed by the distributions generated from Model SC.

We conclude this section with a note on the parameters in the Cobb-Douglas function. Because there is very little publicly available information about the prices of past longevity risk transfers, we have to resort to using pricing data for catastrophic mortality bonds when calibrating γ_1, γ_2 and γ_3. We acknowledge that the calibrated values of γ_1, γ_2 and γ_3 may not be ideal for pricing longevity securities which generally have longer maturities and are related to survival rates rather than death rates. We are also aware that the standard errors of the parameter estimates are quite large, because the estimation of the Cobb-Douglas function is based on only 18 available data points.[9] Nevertheless, even if the applicable values of γ_1, γ_2 and γ_3 are different, the conclusions drawn in this section remain unchanged: in all scenarios under consideration, Model NC yields higher values of EL and EER than Model SC, and the values of EL and PFL produced by the fully-coherent models are always zero.

8 Discussion and Conclusion

In this paper, we have introduced a new concept called semi-coherence. Under this new concept, the projected mortality trajectories of two related populations are permitted to diverge over certain time periods. This new concept is more flexible than the commonly used full-coherence hypothesis, under which mean-reversion is always in effect, but is purposely more restrictive than non-coherence, which produces mortality forecasts that diverge indefinitely over time. We believe that models built on the proposed semi-coherence hypothesis can more realistically quantify the population basis risk associated with pairs of populations such as England and Wales and the United States, whose historical mortality trends, as demonstrated previously, exhibit persistent and significant divergence from time to time.

Another contribution of this paper is the implementation of the semi-coherence hypothesis by using a three-regime VETAR process, applied to the period effects encompassed in the Lee-Carter model structure. In the middle regime of the process, the mortality trajectories of the two populations being modeled are expected to diverge. However, mean-reversion kicks in when the process switches to another regime. The change in regime is determined by the threshold variable, which measures the recent divergence in mortality. Compared to the non- and fully-coherent alternatives, the proposed VETAR process provides a significantly better statistical fit as measured by the AIC. It also yields more reasonable stochastic projections of mortality differentials, with some (but not large) probabilistic provisions for the continuation of the recent

[9] Note that Wills and Sherris (2010) estimated the Cobb-Douglas function using even fewer (13) data points.

trend in mortality differentials for a few more decades.

We have also contributed a new notion called hypothesis risk, the risk arising from the possibility that the assumed coherence hypothesis turns out to be wrong. We have illustrated the impact of hypothesis risk using various pricing and hedging scenarios. In terms of pricing, it is found that different coherence hypotheses lead to highly different premium spread estimates for a security similar to the Kortis bond launched by Swiss Re in 2010. In terms of hedging, the most striking finding is that a hedging strategy developed from a fully-coherent model may yield an unsatisfactory hedge effectiveness if the hypothesis of full coherence does not hold. In our illustration, the strategy based on Model FC1 produces an over 60% hedge effectiveness as long as the true model is fully-coherent, but the effectiveness is reduced to only 30% when the true model is semi-coherent and even becomes negative when the true model is non-coherent.

We acknowledge that the proposed VETAR process is subject to a number of limitations. The most significant limitation is its high data requirement. With an order of $p = 2$, the VAR in each regime contains 13 parameters (2 in the constant vector, 8 in the autoregressive matrices and 3 in the covariance matrix for the innovation vector). As a rather large number of parameters have to be estimated, a long time-series of mortality data, preferably starting in the early 20th century, is required. On top of that, the VETAR process can only be fitted to data series of the same length. Therefore, if it happens that the data series from the two populations being modeled have different lengths, a portion of the longer data series must be discarded.

No different from typical time-series processes, the VETAR process is also subject to the problem of lacking robustness with respect to the presence of outliers. In our application, such outliers may arise from interruptive events such as wars and pandemics. Because outliers may affect parameter estimates and hence mortality forecasts, future research on how to robustify the VETAR process with respect to the presence of outliers is warranted. This research goal may be accomplished by following the lines of Croux and Joossens (2008) and Clements and Krolzig (1998), who consider robust estimation of vector autoregressive models and univariate threshold autoregressive models, respectively.

To focus on the properties of the VETAR process that are the most relevant to longevity risk securitization, we have downplayed the discussion of parameter risk. It is possible to quantify the impact of parameter risk associated with the VETAR process on longevity pricing and hedging by bootstrapping techniques (Koissi et al., 2006; Brouhns et al., 2005). We remark here that a two-stage method is used in estimating the model parameters. In particular, we first estimate the parameters in the Lee-Carter structure by maximum likelihood, and then estimate the parameters in the VETAR process on the basis of the first stage estimates of the historical period effects. Hence, when applying bootstrapping techniques to the VETAR process, the standard errors shown in Table 2 have to be adjusted to reflect the uncertainty surrounding the first stage estimates. We refer readers to Murphy and Topel (2002) for information regarding the necessary standard error adjustments.

The collection of models used in this paper can be generalized in different directions. For example, one may apply the VETAR process to the period and/or cohort effects in more sophisticated mortality models, including the Renshaw-Haberman model (Renshaw and Haberman, 2006) and various versions of the Cairns-Blake-Dowd model (Cairns et al., 2009). One may also extend our current set-up, which models only two populations at a time, to a set-up that can simultaneously handle more than two populations. Such an extension may possibly be accomplished by introducing additional threshold random variables to the VETAR process. More generally, we believe that the VETAR process is not the only process that can lead to semi-coherent mortality forecasts. It would be interesting in future research to explore what other, possibly better, processes exist that also satisfy the semi-coherence hypothesis.

In the study of the hypothetical longevity hedge, it is found that when the true underlying mortality model is the semi-coherent model, the best attainable hedge effectiveness is over 70%, but the key q-duration strategy only yields a hedge effectiveness of about 30%. These results indicate that there is room to improve the existing hedging strategies, making them more adaptable to a less strict form of coherence. Given the properties of semi-coherent mortality forecasts, state-dependent hedging strategies seem highly suitable when the true model is semi-coherent. It would therefore be interesting in future research to apply the recent developments in state-dependent hedging in finance (Lee, 2010; Alizadeh et al., 2008) to the context of longevity risk securitization. The development of state-dependent longevity hedging strategies would also facilitate us to investigate the hypothesis risk that may potentially arise from the situation when the hedging strategy is based on a semi-coherent model but the true model is fully coherent.

Finally, we believe that there is a need to develop a statistical hypothesis test that allows us to determine which one of the three forms of coherence the mortality dynamics of a certain pair of populations satisfy. Such a test will assist hedgers with choosing, from all reference populations available in the market, the reference population that is the most suitable for their hedging needs. Such a test can also help institutions issuing standardized mortality-linked securities to identify the reference populations that are the most likely to attract the greatest demand.

References

Ahmadi, S. and J.S.-H., Li. (2014). Coherent Mortality Forecasting with Generalized Linear Models: A Modified Time-Transformation Approach. Insurance: Mathematics and Economics 59: 149-221.

Alizadeh, A.H., Nomikos, N.K. and Pouliasis, P.K. (2008). A Markov Regime Switching Approach for Hedging Energy Commodities. Journal of Banking and Finance 32: 1970-1983.

Blake, D., Cairns, A., Coughlan, G., Dowd, K. and MacMinn, R. (2013). The New Life Market. Journal of Risk and Insurance 80: 501-557.

Blake, D., Cairns, A.J.G. and Dowd, K. (2006). Living with Mortality: Longevity Bonds and other Mortality-Linked Securities. British Actuarial Journal 12: 153-197.

Blake, D., MacMinn, R., Li, J.S.-H. and Hardy, M. (2014). Longevity Risk and Capital Markets: The 2012-2013 Update. North American Actuarial Journal 18: 501-557.

Boot, J.C.G., Feibes, W. and Lisman, J.H.C. (1967). Further Methods of Derivation of Quarterly Figures from Annual Data. Journal of the Royal Statistical Society. Series C (Applied Statistics) 16: 65-75.

Brouhns, N, Denuit, M. and Keilegom, I.V. (2005). Bootstrapping the Poisson Log-bilinear Model for Mortality Forecasting. Scandinavian Actuarial Journal 3: 212-224.

Bugg, R., Ismail, F., McWilliam, E. and Simpson, P. (2010). Longevity - The Opportunity and the Risks. Milliman White Paper. London: Milliman.

Cairns, A.J.G. (2011). Modelling and Management of Longevity Risk: Approximations to Survival Functions and Dynamic Hedging. Insurance: Mathematics and Economics 49: 438-453

Cairns, A.J.G. (2013a). Modelling and Management of Longevity Risk. Working Paper. Available at http://www.macs.hw.ac.uk/_andrewc/papers/ajgc70.pdf.

Cairns, A.J.G. (2013b). Robust hedging of longevity risk. Journal of Risk and Insurance 80: 621-648.

Cairns, A.J.G., Blake, D. and Dowd, K. (2006). Pricing death: Frameworks for the Valuation and Securitization of Mortality Risk. ASTIN Bulletin 36: 79-120.

Cairns, A.J.G., Blake, D., Dowd, K. and Coughlan, G.D. (2014). Longevity Hedge Effectiveness: A Decomposition. Quantitative Finance 14: 217-235.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., and 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: 1-35.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D. and Khalaf-Allah, M. (2011). Bayesian Stochastic Mortality Modelling for Two Populations. ASTIN Bulletin 41: 29-59.

Chen, H. and Cummins, J.D. (2010). Longevity Bond Premiums: The Extreme Value Approach and Risk Cubic Pricing. Insurance: Mathematics and Economics 46: 150-161.

Chen, H., MacMinn, R.D. and Sun, T. (2013). Multi-Population Mortality Models: A Factor Copula Approach. Paper presented at the Ninth International Longevity Risk and Capital Markets Solutions Conference, Beijing, China.

Clements, M.P. and Krolzig, H.-M. (1998). A Comparison of the Forecast Performance of Markov-switching and Threshold Autoregressive Models of US GNP. The Econometrics Journal 1: 47-75.

Coughlan, G. (2009). Longevity Risk Transfer: Indices and Capital Market Solutions. In Barrieu, P.M. and Albertini, L. (eds.) The Handbook of Insurance Linked Securities. London: Wiley.

Coughlan, G.D., Blake, D., MacMinn, R., Cairns, A.J.G. and Dowd, K. (2013). Longevity Risk and Hedging Solutions. In Dionne, G. (ed.) Handbook of Insurance. Springer.

Croux, C. and Joossens, K. (2008). Robust Estimation of the Vector Autoregressive Model by a Least Trimmed Squares Procedure. In Brito, P. (ed.) COMPSTAT 2008: Proceedings in Computational Statistics. Springer.

Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D. and Khalaf-Allah, M. (2010). Backtesting Stochastic Mortality Models: An Ex-Post Evaluation of Multi-Period-Ahead Density Forecasts. North American Actuarial Journal 14: 281-298.

Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D. and Khalaf-Allah, M. (2011). A Gravity Model of Mortality Rates for Two Related Populations. North American Actuarial Journal 15: 334-356.

Hardy, M.R. (2001). A Regime Switching Model of Long Term Stock Returns. North American Actuarial Journal 3: 185-211.

Harvey, A.C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press.

Hatzopoulos, P. and Haberman, S. (2013). Common Mortality Modeling and Coherent Forecasts. An Empirical Analysis of Worldwide Mortality Data. Insurance: Mathematics and Economics 52: 320-337.

Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), 2014 (www.mortality.org).

Hunt, A. and Blake, D. (2013). Projecting Mortality: Coherence and Co-integration in Multi-Population Projections. Paper presented at the Ninth International Longevity Risk and Capital Markets Solutions Conference, Beijing, China.

Jarner, S.F. and Kryger, E.M. (2011). Modelling Adult Mortality in Small Populations: The SAINT Model. ASTIN Bulletin 41: 377-418.

Koissi, M., Shapiro, A. and Hognas, G. (2006). Evaluating and Extending the Lee-Carter Model for Mortality Forecasting: Bootstrap Confidence Interval. Insurance: Mathematics and Economics 38: 120.

Lee, H.-T. (2010). Regime Switching Correlation Hedging. Journal of Banking and Finance 34: 2728-2741.

Li, J.S.-H. and Hardy, M.R. (2011). Measuring Basis Risk in Longevity Hedges. North American Actuarial Journal 15: 177-200.

Li, J.S.-H. and Luo, A. (2012). Key q-duration: A Framework for Hedging Longevity Risk. ASTIN Bulletin 42: 413-452.

Li, N. and Lee, R. (2005). Coherent Mortality Forecasts for a Group of Population: An Extension of the Lee-Carter Method. Demography 42: 575-594.

Lin, T. and Tsai, C.C.-L. (2014). Applications of Mortality Durations and Convexities in Natural Hedges. North American Actuarial Journal 18: 417-442.

Lin, Y., Liu, S. and Yu, J. (2013). Pricing Mortality Securities with Correlated Mortality Indexes. Journal of Risk and Insurance 80: 921-948.

Mackenbach, J.P., Bos, V., Andersen, O., Cardano, M., Costa, G., Harding, S., Reid, A., Hemstrom, O., Valkonen, T. and Kunst, A.E. (2003). Widening Socioeconomic Inequalities in Mortality in Six Western European Countries. International Journal of Epidemiology 32: 830-836.

Murphy, K.M. and Topel, R.H. (2002). Estimation and Inference in Two-Step Econometric Models. Journal of Business and Economic Statistics 20: 88-97.

Oeppen, J. and Vaupel, J.W. (2002). Broken Limits to Life Expectancy. Science 296: 1029-1031.

Renshaw, A.E. and Haberman, S. (2006). A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors. Insurance: Mathematics and Economics 38: 556-570.

Ribeiro, R. and di Pietro, V. (2009). Longevity Risk and Portfolio Allocation. Available at http://www.jpmorgan.com/pages/jpmorgan/investbk/solutions/lifemetrics/library.

Tan, C.I., Li, J., Li, J.S.-H. and Balasooriya, U. (2014). Parametric Mortality Indexes: From Index Construction to Hedging Strategies. Insurance: Mathematics and Economics 59: 285-299.

Tong, H. and Lim, K. S. (1980). Threshold Autoregression, Limit Cycles and Cyclical Data. Journal of the Royal Statistical Society. Series B (Methodological) 42: 245-292.

Tsay, R.S. (1998). Testing and Modeling Multivariate Threshold Models. Journal of the American Statistical Association 93: 1188-1202.

Tsay, R.S. (2010). Analysis of Financial Time-Series, 3rd Edition. New York: Wiley.

Tsay, R.S., Peña, D. and Pankratz, A.E. (2000). Outliers in Multivariate Time Series. Biometrika 87: 786-804.

Waldron, H. (2007). Trends in Mortality Differentials and Life Expectancy for Male Social Security-Covered Workers, by Average Relative Earnings. ORES Working Paper Series Number 108.

Wang, J.L., Huang, H.C., Yang, S.S. and Tsai, J.T. (2010). An Optimal Produce Mix for Hedging Longevity Risk in Life Insurance Companies: The Immunization Theory Approach. Journal of Risk and Insurance 77: 473-497.

White, K.M. (2002). Longevity Advances in High-income Countries, 1955-96. Population and Development Review 28: 59-76.

Wills, S. and Sherris, M. (2010). Securitization, Structuring and Pricing of Longevity Risk. Insurance: Mathematics and Economics 46: 173-185.

Wilmoth, J.R. (1993). Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change. Technical report. Department of Demography. University of California, Berkeley.

Wilson, C. (2001). On the Scale of Global Demographic Convergence 1950-2000. Population and Development Review 27: 593-600.

Yang, S.S. and Wang, C.W. (2013). Pricing and Securitization of Multi-Country Longevity Risk with Mortality Dependence. Insurance: Mathematics and Economics 52: 157-169.

Zhou, R., Li, J.S.-H. and Tan, K.S. (2013). Pricing mortality risk: A Two-Population Model with Transitory Jump Effects. Journal of Risk and Insurance 80: 733-774.

Zhou, R., Wang, Y., Kaufhold, K., Li, J.S.-H. and Tan, K.S. (2014). Modeling Period Effects in Multi-Population Mortality Models: Applications to Solvency II. North American Actuarial Journal 18: 150-167.

Appendix

Estimation and Diagnostic Checking Results of the Restricted Model

Because some of the estimated parameters in Table 2 are not significant at a 5% significance level, we attempt to re-estimate the model by restricting the insignificant parameters to zero. The estimated restricted model is presented in Table A.1.

Table A.1 Restricted VETAR model estimation results

$\hat{\boldsymbol{\phi}}^{(g)}$	$\hat{\boldsymbol{\Phi}}_1^{(g)}$	$\hat{\boldsymbol{\Phi}}_2^{(g)}$	$\hat{\boldsymbol{\Sigma}}_g$
(a) First regime ($g = 1$)			
$\begin{pmatrix} -0.41 \\ (0.25) \\ -0.59 \\ (0.11) \end{pmatrix}$	$\begin{pmatrix} -0- & -0- \\ -0- & -0- \end{pmatrix}$	$\begin{pmatrix} 0.71 & -0.99 \\ (0.20) & (0.34) \\ -0- & -0.27 \\ & (0.12) \end{pmatrix}$	$\begin{pmatrix} 1.8891 & 0.4870 \\ & 0.2725 \end{pmatrix}$
(b) Second regime ($g = 2$)			
$\begin{pmatrix} -0.48 \\ (0.15) \\ -0.35 \\ (0.10) \end{pmatrix}$	$\begin{pmatrix} -0.43 & -0- \\ (0.11) & \\ -0- & -0.51 \\ & (0.12) \end{pmatrix}$	$\begin{pmatrix} -0.32 & 0.33 \\ (0.10) & (0.16) \\ -0- & -0- \end{pmatrix}$	$\begin{pmatrix} 1.0020 & 0.2392 \\ & 0.5534 \end{pmatrix}$
(c) Third regime ($g = 3$)			
$\begin{pmatrix} -0.86 \\ (0.24) \\ -0.51 \\ (0.31) \end{pmatrix}$	$\begin{pmatrix} -0.49 & 0.43 \\ (0.22) & (0.22) \\ -0- & -0- \end{pmatrix}$	$\begin{pmatrix} -0- & -0- \\ -0- & -0.73 \\ & (0.30) \end{pmatrix}$	$\begin{pmatrix} 0.7281 & 0.0672 \\ & 1.0867 \end{pmatrix}$

Notes: The estimated threshold values are $(\hat{r}_1, \hat{r}_2) = (-0.7668, 1.5471)$.
Standard errors of the estimates are given in parentheses.
"$-0-$" denotes that the parameter has been restricted to zero.

The simplified SCCMs for the standardized residuals of the estimated restricted model are shown in Table A.2. It can be observed that some cross- and auto-correlations at lags 2 and 5

are significant, indicating that the restricted model does not provide an adequate fit. We further examine the estimated restricted model by applying the multivariate portmanteau test to its standardized residuals. The value of the test statistic is $Q_2(5) = 27.78$, which gives a p-value of 0.0336. The null hypothesis that there is no remaining auto- and cross-correlations is rejected at a 5% level of significance. The results of the multivariate portmanteau test also points to the conclusion that the restricted model is inadequate. Because of the inadequacy of the restricted model, the full model is used in the pricing and hedging work of this paper.

Table A.2 Simplified sample cross-correlation matrices (SCCMs) for the standardized residuals of the estimated restricted VETAR process

Lag	1	2	3	4	5
Simplified SCCM	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	$\begin{pmatrix} - & - \\ \cdot & - \end{pmatrix}$	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	$\begin{pmatrix} \cdot & \cdot \\ \cdot & + \end{pmatrix}$