THE LOCALLY LINEAR CAIRNS–BLAKE–DOWD MODEL: A NOTE ON DELTA–NUGA HEDGING OF LONGEVITY RISK

BY

YANXIN LIU AND JOHNNY SIU-HANG LI

ABSTRACT

Although longevity risk arises from both the variation surrounding the trend in future mortality and the uncertainty about the trend itself, the latter is often left unmodeled. In this paper, we address this problem by introducing the locally linear CBD model, in which the drifts that govern the expected mortality trend are allowed to follow a stochastic process. This specification results in median forecasts that are more consistent with the recent trends and more robust relative to changes in the data sample period. It also yields wider prediction intervals that may better reflect the possibilities of future trend changes. The treatment of the drifts as a stochastic process naturally calls for nuga hedging, a method proposed by Cairns (2013) to hedge the risk associated with changes in drifts. To improve the existing nuga-hedging method, we propose a new hedging method which demands less stringent assumptions. The proposed method allows hedgers to extract more hedge effectiveness out of a hedging instrument, and is therefore useful when there are only a few traded longevity securities in the market.

KEYWORDS

Nuga hedging, q-forwards, state-space models, the CBD model, trend changes.

1. INTRODUCTION

Depending on the context and intention, longevity risk can be defined in different ways. From a statistical viewpoint, Coughlan *et al.* (2013) provided the following concise yet complete definition: *"It is a combination of (i) uncertainty surrounding the trend increases in life expectancy and (ii) variations around this uncertain trend that is the real problem. This is what is meant by longevity risk and it arises as a result of unanticipated changes in mortality rates".* Despite the risk encompasses two components, most existing stochastic mortality models capture only the latter. For instance, in the Lee–Carter model (Lee and Carter,

Astin Bulletin 47(1), 79-151. doi: 10.1017/asb.2016.33

1992), the evolution of mortality over time is typically captured by an autoregressive integrated moving average process, with the special case — random walk with drift — being used the most often. What this modeling method captures is "diffusion risk", which arises from the variations around a fixed drift that determines the expected trend, but not the uncertainty associated with the trend itself. The collection of models (Models M1–M8) considered by Cairns *et al.* (2009, 2011a) and Dowd *et al.* (2010a,b) are subject to the same limitation.

Although often left unmodeled, the risk associated with the trend in mortality improvement does exist. There is profound empirical evidence for trend changes in mortality, exemplified by the findings of Gallop (2006) concerning the mortality improvement in the United Kingdom over the past century. It was found that the average rate of mortality improvement for British males was gently fluctuating around 0.7% per annum over the period of 1930–32 to 1970–72, after which the rate increased substantially, reaching 2.0% per annum over the period of 1990–92 to 2001–03. Similar trend changes are also observed in other developed countries (see, e.g., Kannisto *et al.*, 1994; Vaupel, 1997). A number of researchers have further confirmed the trend changes in mortality by rigorous statistical tests, including Li *et al.* (2011) who considered Zivot and Andrew's test, Ahmadi and Li (2014) who used a non-parametric change-point test, and O'Hare and Li (2015) who utilized the CUSUM test that is based on cumulative sums of standardized residuals.

Researchers have also developed mortality models that incorporate trend changes in the past (Renshaw and Haberman, 2003; Coelho and Nunes, 2011; Li *et al.*, 2011; Ahmadi and Li, 2014, O'Hare and Li, 2015; van Berkum *et al.*, 2014), but these models at best can only reflect how historical trend changes may affect the best estimate forecast. They do not capture trend changes as a risk, as they do not allow historical trend changes to recur at random future time points with random extents. For example, in the broken-trend stationary model proposed by Li *et al.* (2011), it is assumed that the future trend in mortality improvement is always the same as the historical trend after the detected structural break point. The model proposed recently by van Berkum *et al.* (2014) is less stringent in the sense that multiple trend changes are permitted during the data sample period, but still the drift is assumed to be fixed beyond the last estimated structural break point. An alternative way that has been used to deal with trend changes is to optimize the calibration window, so that the model is fitted to the period of time during which the trend is the most linear (Booth *et al.*, 2002; Denuit and Goderniaux, 2005; Li *et al.*, 2015a). However, excluding a portion of the data does not address the random nature of trend changes.

On the modeling front, we may address the risk associated with trend changes by permitting the drift term(s) in the assumed mortality model to be stochastic. A previous attempt to introduce a stochastic drift was made by Milidonis *et al.* (2011), who modeled the time-varying factor in the Lee–Carter model using a regime-switching log-normal process with two regimes. The drift term in each regime is permitted to be different, so that the drift of the process may vary as the system switches between the two regimes under an assumed

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

Markov chain. However, as indicated in the estimation results of Milidonis *et al.* (2011), it is the volatility rather than the drift that separates the two regimes. Therefore, the regime-switching process may be suitable for capturing short-term catastrophic mortality events which are generally accompanied with high mortality volatility, but may not be adequate for capturing the risk arising from drift changes. Another attempt was made by Sweeting (2011), who considered a piece-wise linear regression. Although the slope of the regression line is permitted to change in the future, the probability and extent of future slope changes are calculated in an ad hoc manner. In particular, the probability is taken as the ratio of the number of observed break points to the total number of data points, while the extent is estimated using the root mean square of the annual changes in the underlying dynamic factor. Alternative mortality models which more explicitly and rigorously address the risk associated with trend changes are yet to be developed.

The first objective of this paper is to develop a stochastic mortality model that permits the user to quantify not only "diffusion risk" but also "drift risk". Our proposed model is based on the original Cairns–Blake–Dowd (CBD) model (Cairns *et al.*, 2006), under which the dynamics of mortality are driven by two time-varying factors. As usual, the evolution of the time-varying factors is modeled by a bivariate random walk with drift, but on top of that, we permit the drifts themselves to follow another bivariate random walk. This modeling method is justified by the results of Nyblom and Mäkeläinen's (1983) test for random walk coefficients. We call our proposed model the locally linear CBD model (thereafter the LLCBD model), because the drifts governing the linear increments in the two CBD time-varying factors are different in different time steps. To enable estimation, we first formulate our proposed model in a state-space representation, just as how Mavros *et al.* (2014), Hári *et al.* (2008), Pedroza (2006), de Jong and Tickle (2006) and Carter (1996) specified the models they considered. We then use the Kalman filtering technique (Kalman, 1960; Kalman and Bucy, 1961) to estimate the unknown model parameters and retrieve the hidden states (i.e., the two CBD time-varying factors and the two varying drifts) in a recursive manner. The method we use can estimate all parameters in the proposed model in one single stage.

Although the extension of the Lee–Carter model proposed by Hári *et al.* (2008) possesses a time-varying drift, it is different from our proposed model in various aspects. In terms of objectives, the extension of Hári *et al.* (2008) was not designed with a motivation to quantify trend risk, and possibly for this reason, it assumes that the drift would fluctuate around its long-term mean. In contrast, with a goal of assessing drift risk, we postulate the dynamics of the drift vector as a random walk. The use of a random walk is in part because of the support from the random walk coefficient test and in part because we have no a priori knowledge about the value and more importantly the existence of the mean of the drift vector. In terms of model structure, the extension of Hári *et al.* (2008) is built on a reformulation introduced by Girosi and King (2005) whereby the vector of log central death rates are structured to follow a multidimensional

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

random walk, whose drifts are driven by a vector of latent factors which follows a stationary multivariate autoregressive moving average process. To maximize comparability with the original CBD model, our state-space model requires no reformulation and has an observation equation that preserves the parameterization of the original CBD model. Compared to the extension of Hári *et al.* (2008) in which the latent factors are not straightforward to interpret, our proposed model contains richer demographical intuitions as all four hidden states in it have their own physical meanings. In spite of the mentioned differences, the estimation results for both models indicate one thing in common: The permission of time-varying drift(s) results in mortality projections that are more robust with respect to changes in the data sample period. Both models can therefore ameliorate the well-known problem that the usual estimator of the drift(s) of a random walk is highly sensitive to (and indeed completely dependent on) the first and last observations of the data series (see, e.g., Li and Chan, 2005; Zhou and Li, 2013).

Having addressed drift risk in modeling, our second objective is to develop a method to manage it. To our knowledge, Cairns (2013) was the first to study how an index-based longevity hedge may be constructed to reduce the hedger's exposure to drift risk. He extended the existing "delta" hedging method (Cairns, 2011; Luciano *et al.*, 2012; Zhou and Li, 2014) to the "delta–nuga" hedging method, in which the sensitivities of the liability being hedged and the portfolio of hedging instruments to changes in drifts are matched. While the delta–nuga hedging method has some appeals, it is subject to a few limitations. First, in deriving the delta–nuga hedging strategies, it is assumed that the future values of the time-varying factors in the underlying model are related to the current values in a deterministically linear manner. Hence, if the linear relation does not hold, the hedging results would be sub-optimal. Second, when applied to the CBD/LLCBD model, the delta–nuga hedging method requires exactly four hedging instruments. This stringent requirement may render the method impractical in the early stages of market development when the market does not provide the required number of standardized hedging instruments. Third, as we are going to demonstrate in Section 5.5, the delta–nuga hedging method is subject to the singularity problem, which means that the solution to the hedging strategies does not exist when certain combinations of hedging instruments are used. To mitigate these limitations, we propose in this paper the generalized state-space hedging method.

In the generalized state-space hedging method, we waive the linearity assumption by considering the sensitivities of the liability being hedged and the portfolio of hedging instruments to all future hidden states that are relevant. Similar to the work of Cairns *et al.* (2014), we derive the hedging strategies by variance minimization. We regard our proposed method as a generalization, because it degenerates to the traditional delta and delta–nuga hedging methods when all future hidden states are deterministically linearly related to the current ones. Based on the sensitivities to all relevant hidden states, the proposed method may also be seen as a complement to several existing static hedging

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

methods, including the methods of q-duration (Coughlan, 2009) and key q-duration (Li and Hardy, 2011; Li and Luo, 2012) that are based on the sensitivities to the death probabilities at selected ages and the method of key K-duration (Tan *et al.*, 2014) that is based on the sensitivities to the CBD mortality indexes at selected time points. In comparison to the traditional delta and delta–nuga hedging methods, the proposed method is more flexible in terms of the number and type of hedging instruments. Our empirical results indicate that when population basis risk is assumed to be absent, the proposed hedging method can lead to a greater than 85% hedge effectiveness even if only one hedging instrument is used.

The rest of this paper is organized as follows. In Section 2, we explore statistical evidence for stochastic drifts. In Section 3, we detail the specification of the proposed LLCBD model. In the same section, we also analyze the performance of the proposed model by estimating it to some real mortality data. In Section 4, we discuss several additional issues about modeling. In Section 5, we present the generalized state-space hedging method and compare it with the traditional delta and delta–nuga hedging methods. In Section 6, we illustrate the proposed hedging method with a hypothetical example. Finally, in Section 7, we conclude with some suggestions for future research.

2. Evidence for stochastic drifts

Let $q_{x,t}$ be the *crude* probability that an individual dies between time $t-1$ and t (during year t), given that he/she has survived to age x at time $t-1$. We calculate $q_{x,t}$ by the following approximation:

$$q_{x,t} \approx \frac{1}{1+0.5m_{x,t}},$$

where $m_{x,t}$ is the crude central rate of death at age x and in year t.[1] The simplest version of the CBD model can be expressed as

$$\ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_1(t) + \kappa_2(t)(x-\bar{x}) + \epsilon_{x,t},$$

where $\bar{x}$ represents the average over the sample age range to which the model is fitted, $\epsilon_{x,t}$'s are the sampling errors, which are assumed to be i.i.d. normally distributed with a zero mean and a constant variance of σ_ϵ^2, and $\kappa_1(t)$ and $\kappa_2(t)$ are time-varying stochastic factors, of which the dynamics are assumed to follow a bivariate random walk with constant drifts C_1 and C_2.

We estimate the above model to the mortality data from Canadian male population over a sample age range of 50 to 89 (40 ages) and a sample period of 1941 to 2010 (70 years), using the method of least squares; that is, the time-varying

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

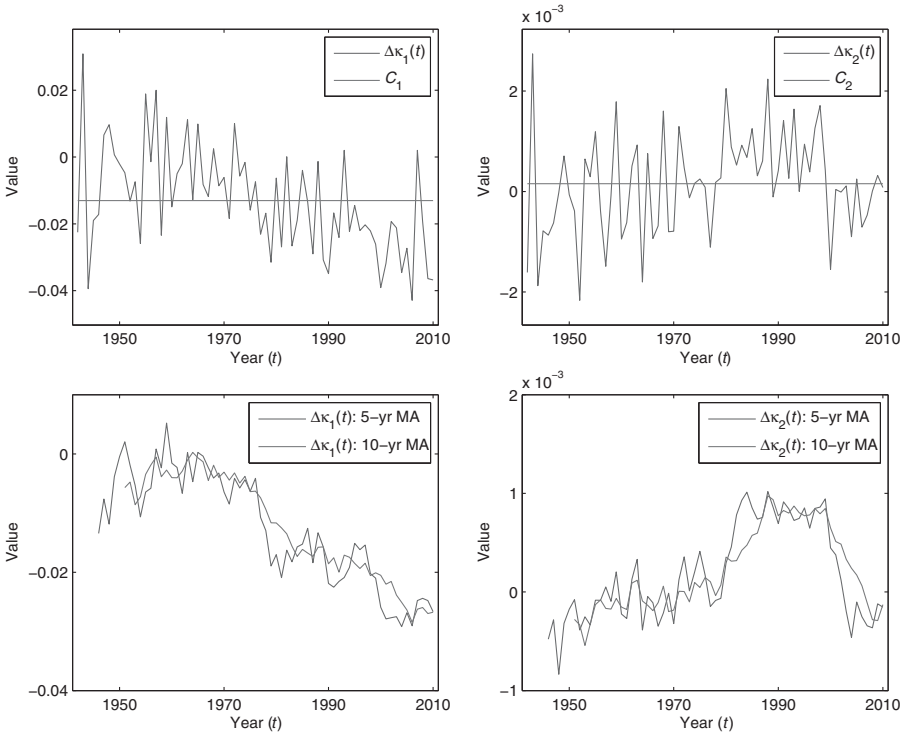

FIGURE 1: The estimated values of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$ and their respective means C_1 and C_2 (the upper panels), and the 5-year and 10-year moving averages of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$ (the lower panels).

stochastic factors are obtained using the following objective function:

$$\min_{\kappa_1(t),\kappa_2(t)} \sum_{x=50}^{89} \left(\ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) - \kappa_1(t) - \kappa_2(t)(x-\bar{x}) \right)^2, \quad t = 1941, \ldots, 2010.$$

The data are obtained from the Human Mortality Database (2014). Let Δ be the difference operator. If the assumption of constant drifts holds, then the estimated values of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$ should fluctuate around their respective sample means $\hat{C}_1$ and $\hat{C}_2$. However, as shown in upper panels of Figure 1, the estimates of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$ do not seem to follow the expected pattern.

The lower panels of Figure 1 depict the 5- and 10-year moving averages of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$. These moving averages may be considered as proxies for the drifts at different time points. It is clear that the moving averages are time-varying and exhibit random patterns. The observations we made from Figure 1 lead to the question as to whether the drifts themselves are stochastic.

We further investigate the stochastic nature of the drift terms by applying a statistical test for random walk coefficients. The following description focuses on C_1, but the test for C_2 can be conducted in a similar manner. Suppose that

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

$\kappa_1(t)$ follows a random walk, with a drift $C_1(t)$ that also follows a random walk itself. The dynamics of $\kappa_1(t)$ can be expressed by the following system of equations:

$$\begin{cases} \Delta\kappa_1(t) = C_1(t) + \xi(t), \\ C_1(t) = C_1(t-1) + \upsilon(t), \quad t = t_a + 1, t_a + 2, \ldots, t_b, \end{cases} \tag{2.1}$$

where $t_a = 1941$ and $t_b = 2010$ represent the beginning and end points of the calibration window, respectively. It is assumed that $\xi(t)$'s and $\upsilon(t)$'s are i.i.d. normally distributed with a zero mean and constant variances, and that $\xi(t)$ and $\upsilon(t)$ are mutually independent.

We let σ_ξ^2 and σ_υ^2 be the variances of $\xi(t)$ and $\upsilon(t)$, respectively. If $\sigma_\upsilon^2 = 0$, then $C_1(t)$ is constant over time rather than following a random walk. It is thus obvious that our goal is to test the null hypothesis $H_0 : \sigma_\upsilon^2 = 0$ against the alternative hypothesis $H_1 : \sigma_\upsilon^2 > 0$. However, as pointed out by LaMotte and McWhorter (1978), it impossible to evaluate the power of such a test. For this reason, they recommended basing the test on the ratio $\sigma_\upsilon^2/\sigma_\xi^2$, which makes the computation of the test's power possible.

We use the locally most powerful invariant (LMPI) test developed by Nyblom and Mäkeläinen (1983) to test $H_0 : \sigma_\upsilon^2/\sigma_\xi^2 = 0$ against $H_1 : \sigma_\upsilon^2/\sigma_\xi^2 > 0$. It follows from Equation (2.1) that we can rewrite $\Delta\kappa_1(t)$ as a sum of past innovations:

$$\Delta\kappa_1(t) = C_1(t_a) + \sum_{i=t_a+1}^{t} \upsilon(i) + \xi(t), \qquad t = t_a + 1, t_a + 2, \ldots, t_b.$$

Consequently,

$$\operatorname{cov}(\Delta\kappa_1(s), \Delta\kappa_1(t)) = \sigma_\upsilon^2 \cdot (\min(s, t) - t_a) + \delta_{s,t}\sigma_\xi^2,$$

where $\delta_{s,t}$ equals 1 if $s = t$ and 0 otherwise, and $s, t = t_a + 1, t_a + 2, \ldots, t_b$. It immediately follows that $\Delta\vec{\kappa}_1 = (\Delta\kappa_1(t_a + 1), \ldots, \Delta\kappa_1(t_b))' \sim \text{MVN}\left(XC_1(t_a), \sigma_\xi^2(\mathrm{I}_{t_b - t_a} + \sigma_\upsilon^2/\sigma_\xi^2 \cdot V)\right)$, where

$$X = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad V = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & \cdots & t_b - t_a \end{pmatrix},$$

MVN represents a multivariate normal distribution, and I_k denotes a k-by-k identity matrix.

Using the generalized Neyman–Pearson lemma, the LMPI test rejects when the LMPI test statistic is greater than some constant that defines the rejection region. On the basis of our set-up, the LMPI test statistic under the null

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

hypothesis can be expressed as

$$
\begin{aligned}
L_{\mathrm{LMPI}} &= \frac{(\Delta\vec{\kappa}_1 - X\hat{C}_1)' V (\Delta\vec{\kappa}_1 - X\hat{C}_1)}{(\Delta\vec{\kappa}_1 - X\hat{C}_1)'(\Delta\vec{\kappa}_1 - X\hat{C}_1)} \\
&= \frac{\sum_{s,t=t_a+1}^{t_b}\sum^{t_b} (\min(s,t) - t_a)\left(\Delta\kappa_1(s) - \hat{C}_1\right)\left(\Delta\kappa_1(t) - \hat{C}_1\right)}{\sum_{t=t_a+1}^{t_b}\left(\Delta\kappa_1(t) - \hat{C}_1\right)^2} \\
&= \frac{\sum_{t=t_a+1}^{t_b}\left(\sum_{s=t}^{t_b}(\Delta\kappa_1(s) - \hat{C}_1)\right)^2}{\sum_{t=t_a+1}^{t_b}\left(\Delta\kappa_1(t) - \hat{C}_1\right)^2}.
\end{aligned}
$$

The derivation of the distribution of L_{LMPI} is based on the property of invariance in translation. In particular, the denominator of L_{LMPI} can be transformed into a χ^2 variable, while the numerator can be transformed into a linear combination of independent χ^2 variables. Using the results of Nyblom and Mäkeläinen (1983), L_{LMPI} follows the same distribution as

$$
\frac{\sum_{k=1}^{t_b-t_a-1}\lambda_{k,t_b-t_a}\left(1+\lambda_{k,t_b-t_a}\sigma_\upsilon^2/\sigma_\xi^2\right)u_k^2}{\sum_{k=1}^{t_b-t_a-1}\left(1+\lambda_{k,t_b-t_a}\sigma_\upsilon^2/\sigma_\xi^2\right)u_k^2},
$$

where u_k's are i.i.d. standard normal random variables and $\lambda_{k,t_b-t_a}^{-1} = 2(1 - \cos(\pi k/(t_b - t_a)))$ for $k = 1, 2, \ldots, (t_b - t_a - 1)$. At a significance level of α, the rejection region c_α for the LMPI test can be constructed by solving following equation:

$$
\alpha = \Pr\left(\frac{L_{\mathrm{LMPI}}}{t_b - t_a - 1} > c_\alpha\right) = \Pr\left(\sum_{k=1}^{t_b-t_a-1}\left(\frac{\lambda_{k,t_b-t_a}}{t_b - t_a - 1} - c_\alpha\right)u_k^2 > 0\right).
$$

The value of c_α can be solved numerically by applying Imhof's (1961) method of inversing the characteristic functions. It is found that at $\alpha = 0.05$, the value of $c_{0.05}$ for our tests is 0.4689.

The calculated values of $L_{\mathrm{LMPI}}/(t_b - t_a - 1)$ for testing stochastic drifts in $\kappa_1(t)$ and $\kappa_2(t)$ are 2.0949 and 0.5087, respectively. Because these values are strictly greater than $c_{0.05} = 0.4689$, the null hypotheses for both tests are rejected at a 5% level of significance. The test results recommend modeling both $\kappa_1(t)$ and $\kappa_2(t)$ with stochastic drifts.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

3. THE LLCBD MODEL

3.1. Model specification

Motivated by the results of the LMPI test, we propose the LLCBD model whereby the drifts in $\kappa_1(t)$ and $\kappa_2(t)$ are stochastic. For notational convenience, we define

$$y_{x,t} := \ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right).$$

As with the original CBD model, the LLCBD model assumes that

$$y_{x,t} = \kappa_1(t) + \kappa_2(t)(x-\bar{x}) + \epsilon_{x,t}, \quad \text{for} \quad x = x_c, x_c+1, \ldots, x_d, \tag{3.1}$$

where $[x_c, x_d]$ represents the age range to which the model is being applied, $\bar{x}$ represents the average over the sample age range $[x_a, x_b]$ to which the model is fitted, and $\epsilon_{x,t}$'s are the sampling errors, which are assumed to be i.i.d. normally distributed with a zero mean and a constant variance of σ_ϵ^2. Note that $[x_c, x_d]$ can be wider than the sample age range $[x_a, x_b]$, because the smooth age-interaction function $x-\bar{x}$ in the model permits extrapolation across the age dimension. Note also that the quantity being modeled ($q_{x,t}$) is the *crude* conditional death probability, so $\epsilon_{x,t}$ in Equation (3.1) measures the uncertainty due to variation in the actual number of deaths, provided that the unobserved underlying death probability

$$\tilde{q}_{x,t} = \frac{e^{\kappa_1(t)+(x-\bar{x})\kappa_2(t)}}{1+e^{\kappa_1(t)+(x-\bar{x})\kappa_2(t)}} \tag{3.2}$$

is known. The variability of $\epsilon_{x,t}$'s depends critically on the size of the population being modeled. Other things equal, the larger the population size is, the smaller the variance of $\epsilon_{x,t}$ (i.e., σ_ϵ^2) is.

The difference between the original CBD and the LLCBD models lies in the way in which the time-varying stochastic factors $\kappa_1(t)$ and $\kappa_2(t)$ are modeled. Specifically, for $i = 1, 2$, $\kappa_i(t)$ in the LLCBD model follows the following system of stochastic processes:

$$\begin{cases} \kappa_i(t) = C_i(t-1) + \kappa_i(t-1) + \eta_i(t), \\ C_i(t) = C_i(t-1) + \eta_{i+2}(t). \end{cases}$$

Equivalently speaking, $\kappa_i(t)$ follows a random walk with a stochastic drift, which itself follows another random walk. We describe this extension of the CBD model as "locally linear", since the drifts in $\kappa_1(t)$ and $\kappa_2(t)$ at different discrete time steps are different. We further assume that the vector of innovations $\vec{\eta}_t = (\eta_1(t), \eta_2(t), \eta_3(t), \eta_4(t))'$ possesses no serial correlation and follows a multivariate normal distribution with a zero mean vector and a covariance matrix Q. Note that $\vec{\eta}_t$ measures the uncertainty surrounding the *unobserved* underlying death probability (i.e., systematic longevity risk). This piece of uncertainty exists even if the number of persons-at-risk is infinitely large.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

By design, all four hidden states, $\kappa_1(t)$, $\kappa_2(t)$, $C_1(t)$ and $C_2(t)$, in the LLCBD model are stochastic. Therefore, Q in the LLCBD model must be positive definite, so that the multivariate normal distribution which $\vec{\eta}_t$ follows is non-degenerate. Let $Q_{i,j}$ be the (i, j)th element in matrix Q. We permit the off-diagonal elements in Q (i.e., $Q_{i,j}$ for $i \neq j$) to be non-zero so that the innovations can be statically correlated with one another.

All four hidden states in the LLCBD model are interpretable. As in the original CBD model, $\kappa_1(t)$ and $\kappa_2(t)$, respectively represent the level and slope of the mortality curve (the curve of $q_{x,t}$ in year t) after a logit transformation. Hence, a reduction in $\kappa_1(t)$ indicates an overall mortality improvement, while an increase in $\kappa_2(t)$ means that mortality (in logit scale) at younger ages (below the mean $\bar{x}$ of the sample age range) improves more rapidly than at older ages. Because $C_1(t)$ and $C_1(t)$, respectively govern the rates of change in $\kappa_1(t)$ and $\kappa_2(t)$, we can interpret $C_1(t)$ to mean the (local) pace of mortality improvement and $C_2(t)$ to mean the (local) change in the age distribution of mortality improvements.

To facilitate estimation and analyses, it is more convenient to express the LLCBD model as a linear Gaussian state-space model comprising of an observation equation and an unobservable state process. We let $\vec{y}_t = (y_{x_c,t}, y_{x_c+1,t}, \ldots, y_{x_d,t})'$ be the vector of observations at time t. The observation equation is given by

$$\vec{y}_t = B\vec{\alpha}_t + \vec{\epsilon}_t,$$

where

$$B = \begin{pmatrix} 1 & x_c - \bar{x} & 0 & 0 \\ 1 & x_c + 1 - \bar{x} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_d - \bar{x} & 0 & 0 \end{pmatrix},$$

$\vec{\alpha}_t = (\kappa_1(t), \kappa_2(t), C_1(t), C_2(t))'$ is the vector of unobservable states at time t and $\vec{\epsilon}_t = (\epsilon_{x_c,t}, \epsilon_{x_c+1,t}, \ldots, \epsilon_{x_d,t})'$ is the vector of error terms at time t. Given the distributional assumptions we made, $\vec{\epsilon}_t \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, \sigma_\epsilon^2 \cdot \text{I}_{x_d - x_c + 1})$.

The unobservable state process can be expressed as

$$\vec{\alpha}_t = A\vec{\alpha}_{t-1} + \vec{\eta}_t,$$

where

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and $\vec{\eta}_t$, as previously defined, is the time-t innovation vector which follows MVN(0, Q).

The state-space specification above is quite general and can be adapted easily to yield different model variants. Most notably, we can recover the original CBD

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

model by setting

$$Q = \begin{pmatrix} Q_{1,1} & Q_{1,2} & 0 & 0 \\ Q_{2,1} & Q_{2,2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where $Q_{i,j}$ for $i, j = 1, 2$ are free parameters, so that the drift terms are forced to be constant. In this special case, $\vec{\eta}_t$ follows a degenerate multivariate normal distribution. To maintain the stochastic nature of $\kappa_1(t)$ and $\kappa_2(t)$, the sub-matrix

$$Q^* := \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \tag{3.3}$$

must be positive definite.

3.2. Estimation

To illustrate, we fit the LLCBD model to the mortality data of Canadian males over a calibration window of 1941–2010 and an age range of 50–89. As a comparison, we also fit the original CBD model to the same data set. The estimation of unknown parameters and retrieval of the hidden states are accomplished by the EM algorithm and the Kalman filter, the details of which are provided in the Appendix.

Table 1 displays the estimated parameters (i.e., σ_ϵ^2 and Q) for both the original CBD model and the LLCBD model. Also shown in the table are the parameters' confidence intervals, which are computed by bootstrapping (Stoffer and Wall, 1991; Cavanaugh and Shumway, 1997). We observe that the permission of stochastic drifts leads to only a minimal change in σ_ϵ^2, but results in rather significant reductions in $Q_{1,1}$ and $Q_{2,2}$. The latter observation is because in the LLCBD model, part of the volatilities of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$ is captured by $Q_{3,3}$ and $Q_{4,4}$.

In Figure 2, we show the values of the hidden states in both estimated models. The values for years 1941 to 2010 are retrieved from the historical data, whereas those for years 2011 and onwards are forecasted. The degree of forecast uncertainty for each hidden state can be seen from the corresponding fan chart, which shows the central 10% prediction interval with the heaviest shading, surrounded by the 20%, 30%, …, 90% prediction intervals with progressively lighter shadings. The line in the middle of each fan chart represents the corresponding median forecast.

Let us first focus on $C_1(t)$ and $C_2(t)$. Under the LLCBD model, the retrieved values of $C_1(t)$ and $C_2(t)$ vary considerably over the calibration window, providing another piece of evidence against the assumption of constant drifts; the median forecasts of $C_1(t)$ and $C_2(t)$ are in line with the retrieved values in the recent past, and are surrounded by ample forecast uncertainty. In sharp contrast, under the original CBD model, the retrieved and forecasted values of $C_1(t)$ and $C_2(t)$ are constant over time, and are roughly equal to the average values of

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 1

THE ESTIMATED VALUES OF σ_ϵ^2 AND Q AND THE CORRESPONDING 95% CONFIDENCE INTERVALS, THE ORIGINAL CBD MODEL AND THE LLCBD MODEL.

Parameter	Estimate	95% Confidence Interval	
The Original CBD Model			
σ_ϵ^2	2.2462×10^{-3}	$(2.1291 \times 10^{-3},$	$2.3664 \times 10^{-3})$
$Q_{1,1}$	2.0164×10^{-4}	$(1.1644 \times 10^{-4},$	$3.0781 \times 10^{-4})$
$Q_{1,2}$	8.3876×10^{-7}	$(-3.5692 \times 10^{-6},$	$5.0612 \times 10^{-6})$
$Q_{2,2}$	7.0298×10^{-7}	$(3.4742 \times 10^{-7},$	$1.1430 \times 10^{-6})$
The LLCBD Model			
σ_ϵ^2	2.2109×10^{-3}	$(2.1237 \times 10^{-3},$	$2.3735 \times 10^{-3})$
$Q_{1,1}$	6.6911×10^{-5}	$(2.4100 \times 10^{-5},$	$1.3649 \times 10^{-4})$
$Q_{1,2}$	3.2010×10^{-6}	$(6.1500 \times 10^{-7},$	$6.1600 \times 10^{-6})$
$Q_{1,3}$	-7.0967×10^{-6}	$(-1.6900 \times 10^{-5},$	$1.3950 \times 10^{-5})$
$Q_{1,4}$	-1.0880×10^{-6}	$(-2.0000 \times 10^{-6},$	$1.6600 \times 10^{-7})$
$Q_{2,2}$	1.7979×10^{-7}	$(4.9750 \times 10^{-8},$	$4.8300 \times 10^{-7})$
$Q_{2,3}$	-1.3259×10^{-7}	$(-7.7000 \times 10^{-7},$	$1.1300 \times 10^{-6})$
$Q_{2,4}$	-6.4419×10^{-8}	$(-1.0600 \times 10^{-7},$	$8.0700 \times 10^{-9})$
$Q_{3,3}$	4.4151×10^{-6}	$(2.0650 \times 10^{-7},$	$1.1300 \times 10^{-5})$
$Q_{3,4}$	-4.3974×10^{-8}	$(-2.7950 \times 10^{-7},$	$2.8450 \times 10^{-7})$
$Q_{4,4}$	2.5681×10^{-8}	$(2.6850 \times 10^{-9},$	$5.5850 \times 10^{-8})$

$C_1(t)$ and $C_2(t)$ in the LLCBD model retrieved over the calibration window; the forecasted values are apparently biased high.

The patterns of the retrieved values of $C_1(t)$ and $C_2(t)$ in the LLCBD model are informative. The trend in $C_1(t)$ appears to fluctuate around a constant prior to the 1970s, but then reduces rapidly over the next two decades. The rapid reduction in $C_1(t)$ echoes the observations made by Kannisto *et al.* (1994) and Vaupel (1997) that the rates of mortality improvement in the developed world significantly accelerated in the 1970s. The trend in $C_2(t)$ also seems stable prior to the 1970s, but then the stability ceases. The pattern of $C_2(t)$ suggests that the age distribution of mortality improvements for Canadian males has undergone rapid changes over the past four decades.

Next, we turn to the patterns of $\kappa_1(t)$ and $\kappa_2(t)$ over time. The dynamics of these two hidden states are of our particular interest, because the death probabilities for all ages at time t are determined entirely by the values of $\kappa_1(t)$ and $\kappa_2(t)$. For $\kappa_1(t)$ and $\kappa_2(t)$, the two models yield similar retrieved values, but highly different forecasts. For the LLCBD model, the gradients of the median forecasts and the retrieved values in the recent past are quite consistent with each other, but this consistency does not apply to the original CBD model. These observations are the consequences of the aforementioned differences in the patterns of $C_1(t)$ and $C_2(t)$ — which determine the expected speed at which $\kappa_1(t)$ and $\kappa_2(t)$ vary — generated from the two models. It is also noteworthy to compare

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

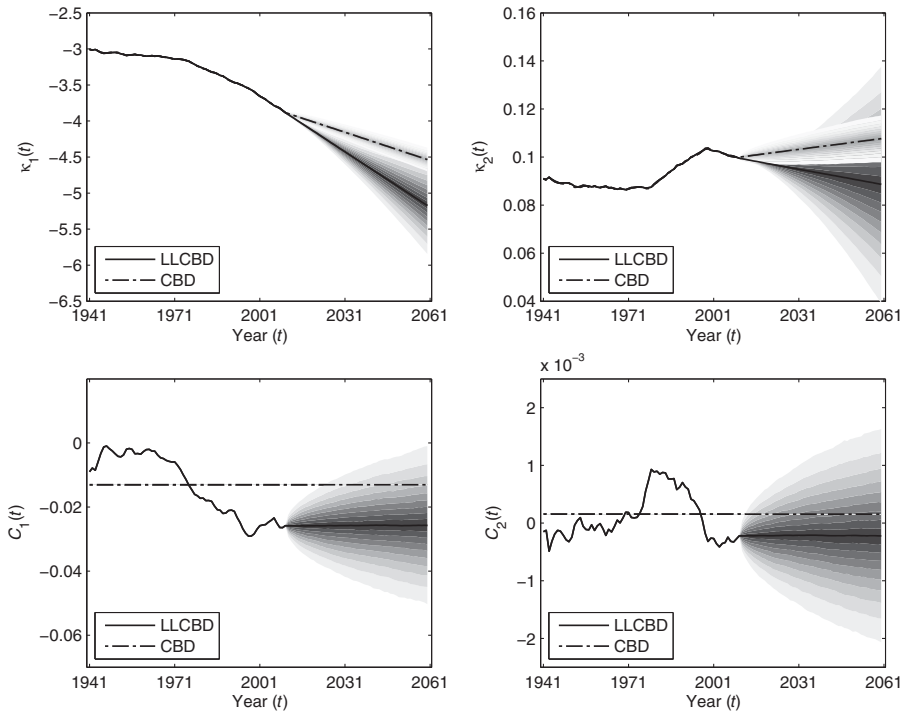

FIGURE 2: The retrieved (1941–2010) and forecasted (2011–2060) values of the hidden states, $\kappa_1(t)$, $\kappa_2(t)$, $C_1(t)$ and $C_2(t)$, in the original CBD model and the LLCBD model.

the levels of forecast uncertainty. By permitting stochastic drifts, the LLCBD model results in more conservative prediction intervals for $\kappa_1(t)$ and $\kappa_2(t)$ (and hence for $q_{x,t}$) in the long run. This outcome is not surprising, because the randomness associated with the drifts contributes to the uncertainty surrounding the forecasts of $\kappa_1(t)$ and $\kappa_2(t)$.

Finally, we remark that the hidden states retrieved over 1941–2010 are subject to uncertainty. Figure 3 shows the 95% confidence intervals for the retrieved hidden states in the LLCBD model. Following Shumway and Stoffer (2006), the 95% confidence interval for $\kappa_1(t)$ is calculated as $\kappa_1(t) \pm 1.96\sqrt{\text{Var}(\kappa_1(t))}$, and those for the other hidden states are calculated in a similar manner. The variances of the retrieved states are computed using the Kalman filter and Kalman smoother, which are detailed in the Appendix.

3.3. Goodness-of-fit

We first evaluate the fit of the LLCBD model to the historical data with a test suggested by Harvey (1990). The test is based on the model's vector of prediction

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

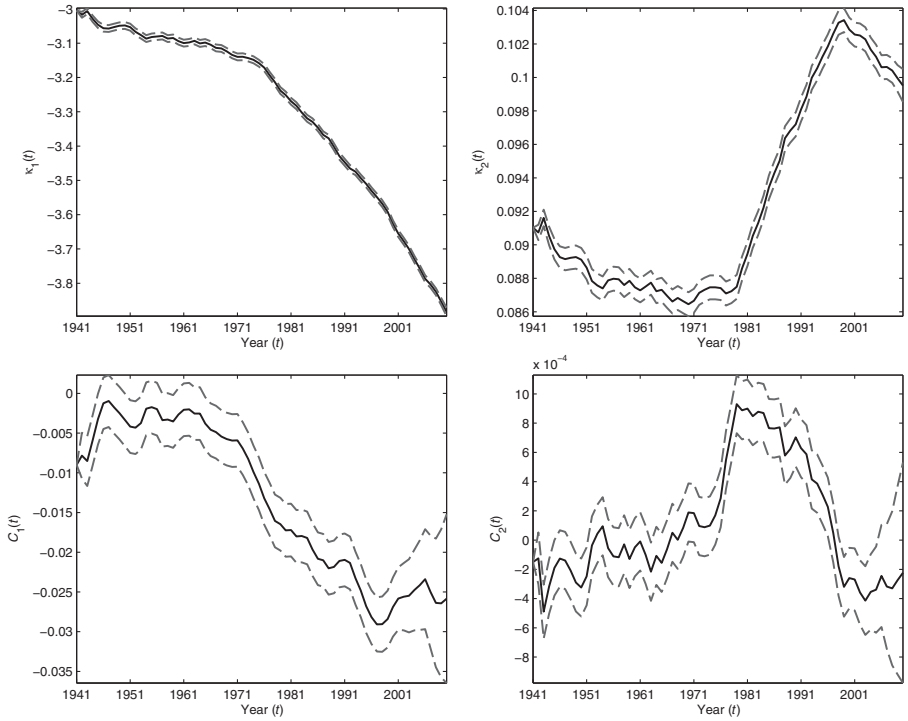

FIGURE 3: The retrieved hidden states (solid lines) in the LLCBD model and their 95% confidence intervals (dashed lines), 1941–2010.

errors, $\vec{w}_t = (w_{x_a,t}, w_{x_a+1,t}, \ldots, w_{x_b,t})'$, which can be computed as

$$\vec{w}_t = \vec{\tilde{y}}_t - BA\mathrm{E}[\vec{\alpha}_{t-1}], \qquad t = t_a + 1, t_a + 2, \ldots, t_b,$$

where $\vec{\tilde{y}}_t$ denotes the realization of $\vec{y}_t$, $\mathrm{E}[\vec{\alpha}_{t-1}]$ represents the retrieved vector of hidden states at time $t-1$, and $BA\mathrm{E}[\vec{\alpha}_{t-1}]$ is the one-step ahead predicted value of $\vec{y}_t$. According to Harvey (1990), the goodness-of-fit of a state-space model can be evaluated through the prediction error variance σ_p^2 and the prediction error mean deviation $\mathcal{D}$, which can approximated as

$$\sigma_p^2 \approx \frac{1}{(x_b - x_a + 1)(t_b - t_a)} \sum_{x=x_a}^{x_b} \sum_{t=t_a+1}^{t_b} w_{x,t}^2$$

and

$$\mathcal{D} \approx \frac{1}{(x_b - x_a + 1)(t_b - t_a)} \sum_{x=x_a}^{x_d} \sum_{t=t_a+1}^{t_b} |w_{x,t}|,$$

respectively. If the sample size is large and the model is specified correctly, then $w_{x,t}$'s are i.i.d. normal random variables with a zero mean and a constant

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 2

THE VALUES OF $\mathcal{N}$, $\ln(\hat{\mathcal{L}})$ AND AIC FOR THE ORIGINAL CBD MODEL AND THE LLCBD MODEL.

Model	$\ln(\hat{\mathcal{L}})$	AIC	$\mathcal{N}$
CBD	4465.05	−8922.11	4
LLCBD	4487.89	−8953.78	11

variance σ_p^2. Under this condition,[2] the prediction error mean deviation $\mathcal{D}$ would converge in probability to $(2/\pi)^{0.5}\sigma_p$. Therefore, if the model provides an adequate fit to the historical data, then the value of $(2/\pi)\sigma_p^2/(\mathcal{D}^2)$ should be close to 1. For the fitted LLCBD model, $\sigma_p^2 = 0.0023$, $\mathcal{D} = 0.0370$ and thus $(2/\pi)\sigma_p^2/(\mathcal{D}^2) = 1.07$. The calculated value of $(2/\pi)\sigma_p^2/(\mathcal{D}^2)$ indicates an adequate fit.

Next, we compare the fit of the LLCBD model with that of the original CBD model. When formulated as a Gaussian state-space model, the original CBD model is nested in the LLCBD model. Therefore, we can evaluate the relative goodness-of-fit of the two models by the Akaike information criterion (AIC), which is defined as

$$\text{AIC} = 2(\mathcal{N} - \ln(\hat{\mathcal{L}})),$$

where $\mathcal{N}$ and $\hat{\mathcal{L}}$ represent the number of parameters and the maximized likelihood value, respectively. A model with a smaller AIC value is more preferred. We remark here that in a state-space formulation, $\kappa_1(t)$, $\kappa_2(t)$, $C_1(t)$ and $C_2(t)$ are regarded as hidden states rather than model parameters. Hence, for example, the total number of parameters in the LLCBD model is 11, encompassing σ_ϵ^2 and 10 distinct elements of matrix Q.

In Table 2, we report the values of $\mathcal{N}$, $\ln(\hat{\mathcal{L}})$ and AIC for each model we estimated. The results indicate that the LLCBD model provides a significantly better fit than the original CBD model, taken into account the additional parameters it contains.

3.4. Forecasting performance

We now perform two tests to evaluate the forecasting performance of the models under consideration. The first test is the "contracting horizon backtest" previously considered by Dowd *et al.* (2010b) and Lee and Miller (2001). The test is based on the accuracy of the projections of $\ln(q_{x,t}/1-q_{x,t})$ in year $t = 2010$, using models that are estimated to data over different calibration windows. In particular, the first forecast is derived from data over 1941–1971, the second forecast is derived from data over 1941–1972, and so on. As the end point of calibration, window becomes closer to 2010, the forecasted value of $\ln(q_{x,2010}/1 - q_{x,2010})$ should converge to the actual value. We may regard the forecasting performance of a model as good if the model yields projections that are close to the actual value, no matter what the calibration window is. The result of this test for $x = 60$

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

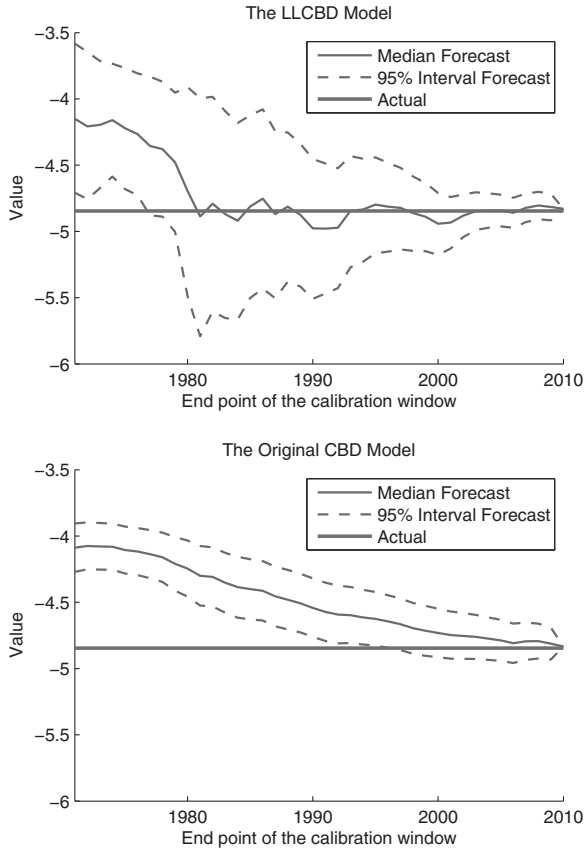

FIGURE 4: The median and 95% interval forecasts of $\ln(q_{60,2010}/1 - q_{60,2010})$, generated from the CBD and LLCBD models that estimated to data over different calibration windows. The starting point of the calibration windows is always 1941 but end points range from 1971 to 2009.

is shown in Figure 4. Except the first few, the median forecasts produced by the LLCBD model are fairly close to the actual value realized in 2010, but those generated from the original CBD model are consistently biased high. In addition, by comparing the proportions of the 95% prediction intervals that encompass the actual value, we may infer that the LLCBD model provides a more adequate provision of uncertainty. The results for other values of x in the sample age range are similar and are therefore not shown.

In the second test, we estimate the models to restricted calibration windows, and then compare the forecasted values produced by the models with the actual values that are not used in fitting the models. We consider 39 restricted sample periods, ranging from 1941–1971 to 1941–2009. The comparisons are made on the basis of two metrics: Mean Error (ME) and Mean Squared Error (MSE). The results of this test are tabulated in Table 3. For instance, the ME of −0.0675 for the LLCBD model with a calibration window of 1941–1995 is computed by

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 3

THE MEAN ERRORS (ME) AND MEAN SQUARED ERRORS (MSE) FOR THE FORECASTS OF $\ln(q_{x,t}/1-q_{x,t})$ PRODUCED BY THE ORIGINAL CBD MODEL AND THE LLCBD MODEL, USING DATA OVER DIFFERENT CALIBRATION WINDOWS.

Calibration	ME		MSE		Calibration	ME		MSE	
Window	LLCBD	CBD	LLCBD	CBD	Window	LLCBD	CBD	LLCBD	CBD
1941–1971	−0.1956	−0.2267	0.0927	0.1120	1941–1991	−0.0054	−0.1204	0.0195	0.0250
1941–1972	−0.1775	−0.2493	0.0797	0.1209	1941–1992	0.0009	−0.1116	0.0169	0.0225
1941–1973	−0.1925	−0.2563	0.0850	0.1230	1941–1993	−0.0534	−0.1267	0.0155	0.0256
1941–1974	−0.2409	−0.2661	0.1027	0.1262	1941–1994	−0.0528	−0.1215	0.0136	0.0238
1941–1975	−0.2223	−0.2611	0.0871	0.1203	1941–1995	−0.0675	−0.1239	0.0139	0.0239
1941–1976	−0.2025	−0.2623	0.0755	0.1192	1941–1996	−0.0616	−0.1183	0.0131	0.0222
1941–1977	−0.1264	−0.2434	0.0548	0.1119	1941–1997	−0.0638	−0.1163	0.0144	0.0216
1941–1978	−0.1151	−0.2338	0.0521	0.1059	1941–1998	−0.0587	−0.1107	0.0177	0.0210
1941–1979	−0.0567	−0.2046	0.0407	0.0909	1941–1999	−0.0507	−0.1057	0.0187	0.0201
1941–1980	−0.0691	−0.2049	0.0117	0.0826	1941–2000	−0.0095	−0.0842	0.0143	0.0145
1941–1981	−0.0232	−0.1848	0.0112	0.0683	1941–2001	0.0047	−0.0688	0.0107	0.0116
1941–1982	−0.0668	−0.1933	0.0146	0.0686	1941–2002	−0.0089	−0.0669	0.0079	0.0112
1941–1983	−0.0386	−0.1768	0.0135	0.0584	1941–2003	−0.0199	−0.0657	0.0069	0.0109
1941–1984	−0.0213	−0.1649	0.0146	0.0513	1941–2004	−0.0059	−0.0506	0.0055	0.0087
1941–1985	−0.0683	−0.1741	0.0171	0.0512	1941–2005	−0.0020	−0.0411	0.0057	0.0079
1941–1986	−0.0822	−0.1745	0.0159	0.0499	1941–2006	0.0244	−0.0143	0.0061	0.0064
1941–1987	−0.0389	−0.1545	0.0130	0.0409	1941–2007	0.0001	−0.0289	0.0059	0.0071
1941–1988	−0.0835	−0.1644	0.0224	0.0407	1941–2008	−0.0187	−0.0375	0.0062	0.0074
1941–1989	−0.0554	−0.1507	0.0190	0.0353	1941–2009	−0.0171	−0.0293	0.0059	0.0065
1941–1990	−0.0043	−0.1277	0.0176	0.0281					

averaging the errors (defined as the actual value less the forecasted value) made in the forecasts of $\ln(q_{x,t}/1-q_{x,t})$ for $x = 50, \ldots, 89$ and $t = 1996, \ldots, 2010$. It can be seen that on the basis of all three metrics, the LLCBD model consistently yields better forecast accuracy in comparison to the original CBD model.

3.5. Robustness

When modeling mortality dynamics, it is reasonable to incorporate the most recent data. However, there is no consensus among researchers as to what length of calibration window should be used. It has been demonstrated extensively that mortality forecasts produced by traditional projection models, in which the drift term(s) is/are assumed to be constant, are highly sensitive to the length of the calibration window used. While a longer calibration window permits us to incorporate more information from the historical data, it generally leads to a forecast that is not sufficiently consistent with the recent trend. This problem, as we are about to demonstrate, may be ameliorated by permitting stochastic drifts.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 4

THE LMPI TEST RESULTS (TEST STATISTIC AND CRITICAL VALUES AT 5% AND 10% SIGNIFICANCE LEVELS) FOR DIFFERENT CALIBRATION WINDOWS.

	Test Statistic		Critical Value	
Calibration Window	$\Delta\kappa_1(t)$	$\Delta\kappa_2(t)$	5%	10%
1941–2010	2.0949	0.5087	0.4689	0.3485
1951–2010	2.2383	0.3515	0.4709	0.3503
1961–2010	1.9063	0.2661	0.4719	0.3529
1971–2010	0.9425	0.4444	0.4728	0.3544

We consider four calibration windows which have the same end point but different starting points: 1941–2010, 1951–2010, 1961–2010 and 1971–2010. We first perform the LMPI test for the four calibration windows (see Table 4). For all four calibration windows, the null hypothesis of a constant drift in $\kappa_1(t)$ is rejected at the 5% significant level; and for all but only one calibration window, the null hypothesis of a constant drift in $\kappa_2(t)$ is rejected at the 10% significant level.

Then, we estimate both the original CBD model and the LLCBD model to data over the four calibration windows. For each model, we examine how the resulting forecasts may change as the starting point of the calibration window moves.

Figure 5 depicts the forecasts of the hidden states in both models on the basis of the four different calibration windows. For the original CBD model, the four calibration windows lead to noticeably different estimates of $C_1(t)$ and $C_2(t)$, and hence considerably different rates of change in $\kappa_1(t)$ and $\kappa_2(t)$. The four median forecasts of $\kappa_1(t)$ and $\kappa_2(t)$ are clearly diverging, while the four fan charts are far from being overlapping one another. Despite the forecasts based on the calibration window starting in 1971 are somewhat consistent with the recent trends, the consistency diminishes significantly as the calibration window begins earlier.

Compared to the original CBD model, the LLCBD model produces median forecasts that are substantially more robust with respect to changes in the beginning point of the calibration window. Regardless of how long the calibration window is, the consistency of the median forecasts with the recent trends remains. These features may be attributed again to the permission of varying drifts, so that the projected rates of change in $\kappa_1(t)$ and $\kappa_2(t)$ are in line with the rates of change in the recent past rather than being close to the average rates of change over the calibration window.

The widths of the fan charts for $C_1(t)$ and $C_2(t)$ in the LLCBD model deserve a few comments. For $C_1(t)$, the longer the calibration window is, the wider the fan chart is. This relationship is expected, as more historical variations are incorporated into the model when the calibration window lengthens. However, the opposite is true for $C_2(t)$. This apparently non-intuitive relationship may be

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

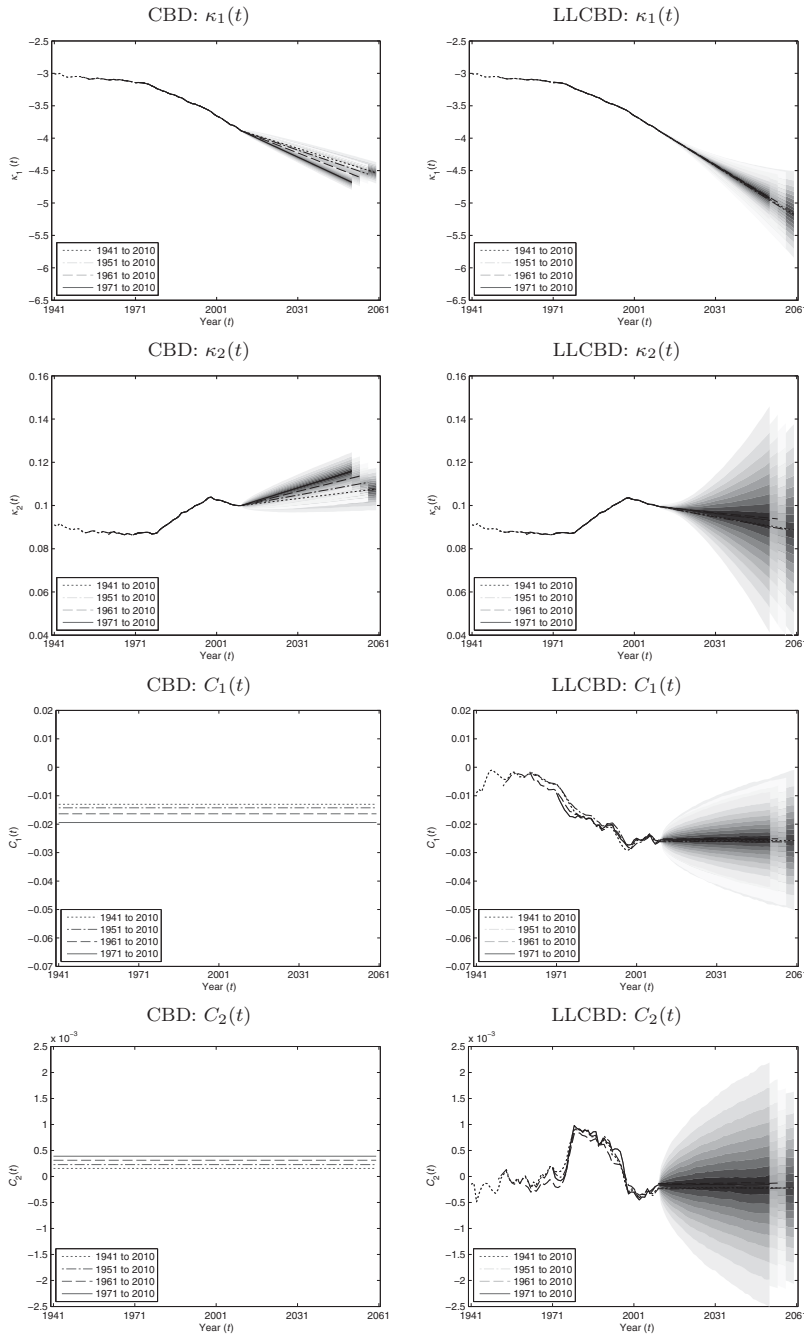

FIGURE 5: Forecasts of the hidden states in the original CBD model and the LLCBD model that are fitted to data over four calibration windows: 1941–2010, 1951–2010, 1961–2010, 1971–2010.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

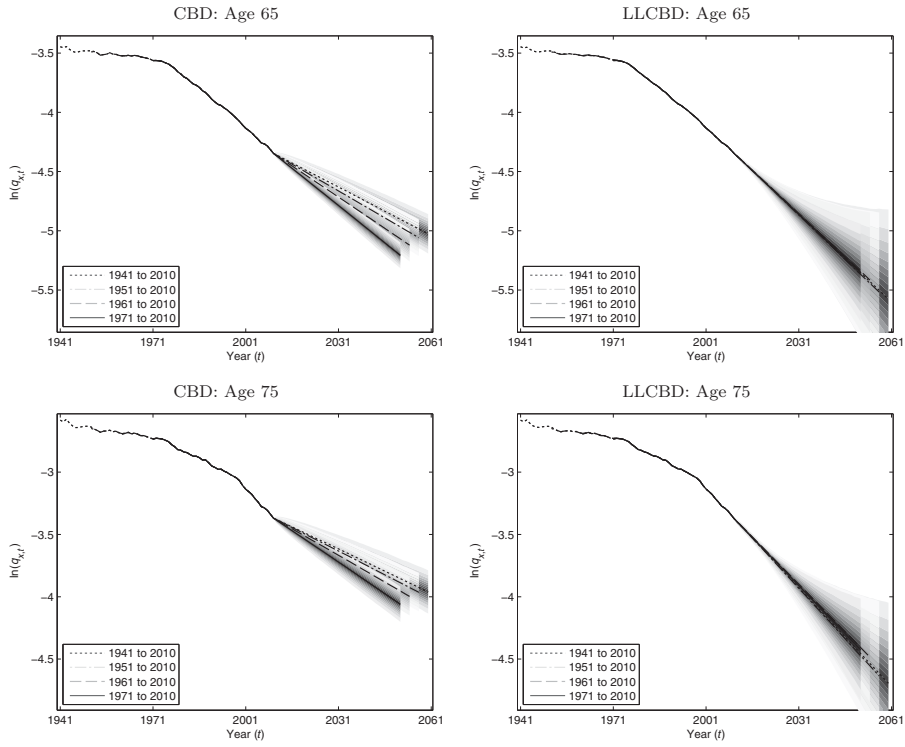

FIGURE 6: Forecasts of $\ln(q_{x,t})$ at $x = 65, 75$ produced by the original CBD model and the LLCBD model that are fitted to data over four calibration windows: 1941–2010, 1951–2010, 1961–2010, 1971–2010.

attributed to the fact that volatility of the retrieved values of $C_2(t)$ in the earlier decades is much smaller. As we begin the calibration window earlier, the calibration window covers a longer period of low $C_2(t)$ volatility while the period of high $C_2(t)$ volatility covered remains unchanged, so the (average) $C_2(t)$ volatility captured by estimated model becomes smaller.

Figure 6 shows, for both models, the forecasts of $q_{x,t}$ at ages 65 and 75 that are based on the four different calibration windows. In terms of $q_{x,t}$, the difference between the robustness of the two models is even more apparent.

Still, there exists small variation in the width of the LLCBD fan charts. As $\ln(q_{x,t})$ is a function of $\kappa_1(t)$ and $\kappa_2(t)$, the uncertainty surrounding $\ln(q_{x,t})$ depends on the uncertainty surrounding $C_1(t)$ and $C_2(t)$. In Figure 6, the two ages, $x = 65$ and $x = 75$, considered are close to $\bar{x} = 69.5$. When x is close to $\bar{x}$, the coefficient $(x - \bar{x})$ of $\kappa_2(t)$ is small, which means the uncertainty surrounding $C_2(t)$ has a relatively small impact on the uncertainty surrounding $\ln(q_{x,t})$. For this reason, the fan charts of $\ln(q_{x,t})$ have similar patterns to the fan charts of $C_1(t)$: The longer the calibration window is, the wider the fan chart is.

As in the forecasts of the hidden states, in the *long run* the interval forecasts of $q_{x,t}$ produced by the LLCBD model are more conservative than those

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

generated from the original CBD model. The fan charts derived from the LLCBD model encompass a larger collection of possible long-term mortality scenarios, ranging from a zero mortality improvement rate to improvement rates that are even greater that those realized in the recent past. As explained below, we view the provision of fan charts that are wider in the long run as an advantage.

When developing a mortality model, it is important to consider biological reasonableness, a concept that was first raised by Cairns *et al.* (2006) in the context of median mortality forecasts. Simply put, this concept means that the collective views of experts in mortality should be taken into account. For example, one should rule out a model that projects a strictly positive probability of immortality.

The concept of biological reasonableness should also be applicable to interval forecasts. In the context interval mortality forecasting, we believe that it is legitimate to interpret biological reasonableness as follows: If the collective view of experts is that future life expectancies should not exceed a certain range, then a biologically reasonable interval forecast should not be wider than the range possible of outcomes that experts agree on; in contrast, if experts have rather different opinions on the prospect of longevity, then a biologically reasonable interval forecast should take a shape that encompasses as much as possible the range of opinions. We use two examples to explain why we regard the interval mortality forecasts produced by our LLCBD model are more biologically reasonable.

The seminal work of Oeppen and Vaupel (2002) found that the trend in record life expectancy since 1840 is close to perfectly linear, showing no sign of deceleration. It was then argued that the linear trend would continue in the coming decades. To achieve a linear climb in life expectancy, age-specific death probabilities need to decline at an increasing pace. As shown in Figure 6, the plausibility of an increasing rate of mortality recline can be captured in the fan charts derived from the LLCBD model, but not in those generated from the original CBD model.

The recent reports produced by the Society of Actuaries (2014) and the Canadian Institute of Actuaries (2014) suggest that (at least part of) the actuarial profession in North America believes that mortality improvement rates will reduce to 0–1% after a transitional period of some 20–30 years. The profession's view means that trajectories of future mortality rates will become flat or almost flat a few decades from now. Such an outcome does not seem to be captured by the fan charts generated using the original CBD model, but may possibly be contained in the LLCBD fan charts whose widths increase with time more quickly.

3.6. Excluding variation in death counts

As the quantity being modeled ($q_{x,t}$) by the LLCBD model is the *crude* conditional death probability, the fan charts in Figure 6 incorporate both systematic

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

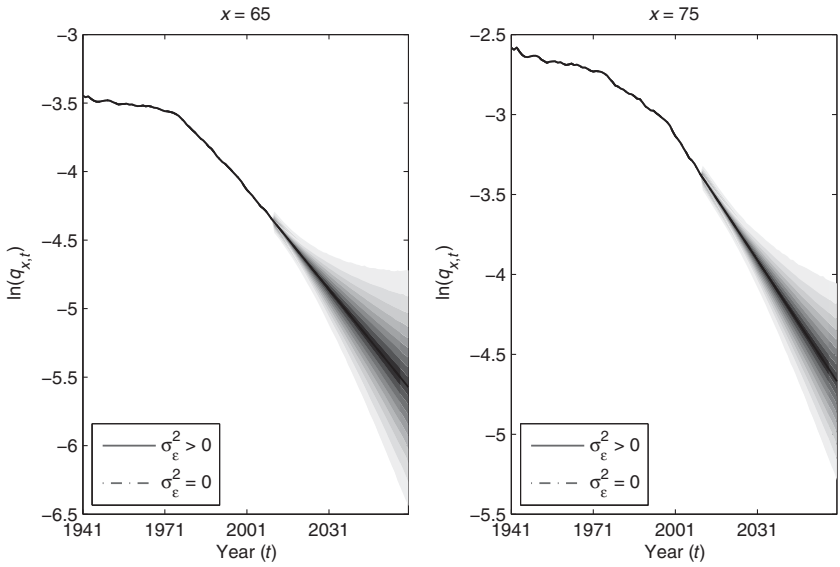

FIGURE 7: Forecasts of $\ln(q_{x,t})$ at $x = 65, 75$ produced by the LLCBD models with $\sigma_\epsilon^2 = 0$ (excluding variation in death counts) and with $\sigma_\epsilon^2 > 0$ (including variation in death counts).

longevity risk and variation in the actual number of deaths. They reflect the level of uncertainty surrounding the future *crude* death probabilities, assuming that the population size remains stable in the future.

To exclude the uncertainty due to variation in the actual number of deaths, we can generate fan charts of future death probabilities under the assumption that $\sigma_\epsilon^2 = 0$. Figure 7 compares the fan charts that incorporate only systematic longevity risk with those that incorporate both sources of uncertainty. It can be observed that systematic longevity risk accounts for most of the total uncertainty. This outcome is not overly surprising, because the population being modeled is a national population with a reasonably large number of persons-at-risk at each age over the age range under consideration.

3.7. Further comments on the dynamics of $C_1(t)$ and $C_2(t)$

We assume that $C_i(t)$, $i = 1, 2$ follows a random walk rather than a mean-reverting stationary process (e.g., an AR(1)), on grounds that we have no a priori knowledge about the value and more importantly the existence of the mean of the drift vector. To substantiate the random walk assumption with statistical evidence, we consider the Dickey–Fuller test, which tests the null hypothesis of

$$z(t) = z(t-1) + \epsilon_z(t)$$

against the alternative hypothesis of

$$z(t) = \mu_z + \phi_z z(t-1) + \epsilon_z(t)$$

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 5

THE RESULTS OF THE DICKEY–FULLER TESTS FOR A RANDOM WALK AGAINST AN AR(1), APPLIED TO THE 5- AND 10-YEAR MOVING AVERAGES OF $\Delta\kappa_1(t)$ AND $\Delta\kappa_2(t)$ IN THE ORIGINAL CBD MODEL AND THE RETRIEVED VALUES OF $C_1(t)$ AND $C_2(t)$ IN THE LLCBD MODEL. A 5% LEVEL OF SIGNIFICANCE IS USED.

p-value	Test Statistic	Critical Value	p-value	Test Statistic	Critical Value
5-Year Moving Averages of $\Delta\kappa_1(t)$			5-Year Moving Averages of $\Delta\kappa_2(t)$		
0.6293	−1.2349	−2.9097	0.1596	−2.3529	−2.9097
10-Year Moving Averages of $\Delta\kappa_1(t)$			10-Year Moving Averages of $\Delta\kappa_2(t)$		
0.9572	0.0316	−2.9141	0.5815	−1.3441	−2.9141
Retrieved Values of $C_1(t)$			Retrieved Values of $C_2(t)$		
0.9802	0.3685	−2.9054	0.7115	−1.0467	−2.9054

for a generic time-series $\{z(t)\}$, where μ_z is a constant and ϕ_z is another constant with an absolute value that is strictly smaller than 1. The test is applied to the following:

1. 5-year moving averages of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$ in the original CBD model, estimated using the least squares method;
2. 10-year moving averages of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$ in the original CBD model, estimated using the least squares method;
3. the retrieved values of $C_1(t)$ and $C_2(t)$ in the LLCBD model.

As mentioned in Section 2, (a) and (b) may be regarded as proxies for the drifts at different time points. The results of all tests performed (see Table 5) are in favour of a random walk.

4. OTHER MODELING CONSIDERATIONS

4.1. A comparison with models with additional dynamic factors and/or age effect structures

One may wonder if the benefits of using a stochastic drift process can be achieved by using models with additional dynamic factors and/or different age effect structures. For this reason, we further compare the proposed LLCBD model against the following four discrete-time stochastic mortality models, which have been considered extensively in the literature (see, e.g., Cairns *et al.*, 2009, 2011a; Dowd *et al.*, 2010a,b).[3]

- The original Lee–Carter model (Model M1):

$$\ln(m_{x,t}) = a(x) + b(x)\kappa(t) + \epsilon_{x,t},$$

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

where $a(x)$ and $b(x)$ are age-specific parameters, $\kappa(t)$ is a time-varying dynamic factor, and $\epsilon_{x,t}$ is the error term. Compared to the CBD/LLCBD model, Model M1 has one fewer dynamic factor and a different age effect structure (specified by parameters $a(x)$ and $b(x)$).

- The Renshaw–Haberman model (Model M2):

$$\ln(m_{x,t}) = a(x) + b_1(x)\kappa(t) + b_2(x)\gamma(t-x) + \epsilon_{x,t},$$

where $a(x)$, $b_1(x)$ and $b_2(x)$ are age-specific parameters, $\kappa(t)$ is a time-varying dynamic factor, $\gamma(t-x)$ is a cohort-varying dynamic factor, and $\epsilon_{x,t}$ is the error term. Compared to the CBD/LLCBD model, Model M2 has the same number of dynamic factors, but one of which is cohort-related rather than time-related. It has also a different age effect structure.

- The CBD model with a cohort effect (Model M6):

$$\ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_1(t) + \kappa_2(t)(x-\bar{x}) + \gamma(t-x) + \epsilon_{x,t},$$

where $\bar{x}$ is the average of the age range $[x_a, x_b]$ to which the model is calibrated, $k_1(t)$ and $k_2(t)$ are time-varying dynamic factors, $\gamma(t-x)$ is a cohort-varying dynamic factor, and $\epsilon_{x,t}$ is the error term. Compared to the CBD/LLCBD model, Model M6 contains one extra dynamic factor, which varies with year-of-birth.

- The CBD model with quadratic and cohort effects (Model M7):

$$\ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)}((x-\bar{x})^2 - \hat{\sigma}_x^2) + \gamma(t-x) + \epsilon_{x,t},$$

where $\hat{\sigma}_x^2$ is the mean of $(x-\bar{x})^2$ over $[x_a, x_b]$. Compared to the CBD/LLCBD model, Model M7 contains two extra dynamic factors, one of which varies with time and the other of which varies with year-of-birth.

As what we did for the CBD and LLCBD models in Section 3.4, we evaluate the forecasting performance of the four additional models by

i. using the "contracting horizon backtest" considered in Dowd *et al.* (2010b), and
ii. fitting the models to restricted calibration windows and then comparing the resulting forecasts with the actual values that are not used in fitting the models.

The result of (i) is shown in Figure 8. By comparing Figure 8 with the upper panel of Figure 4, we can conclude that none of the four additional models can produce forecasts with the desirable properties (more accurate median forecasts and more adequate provisions of uncertainty) possessed by the LLCBD forecasts.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

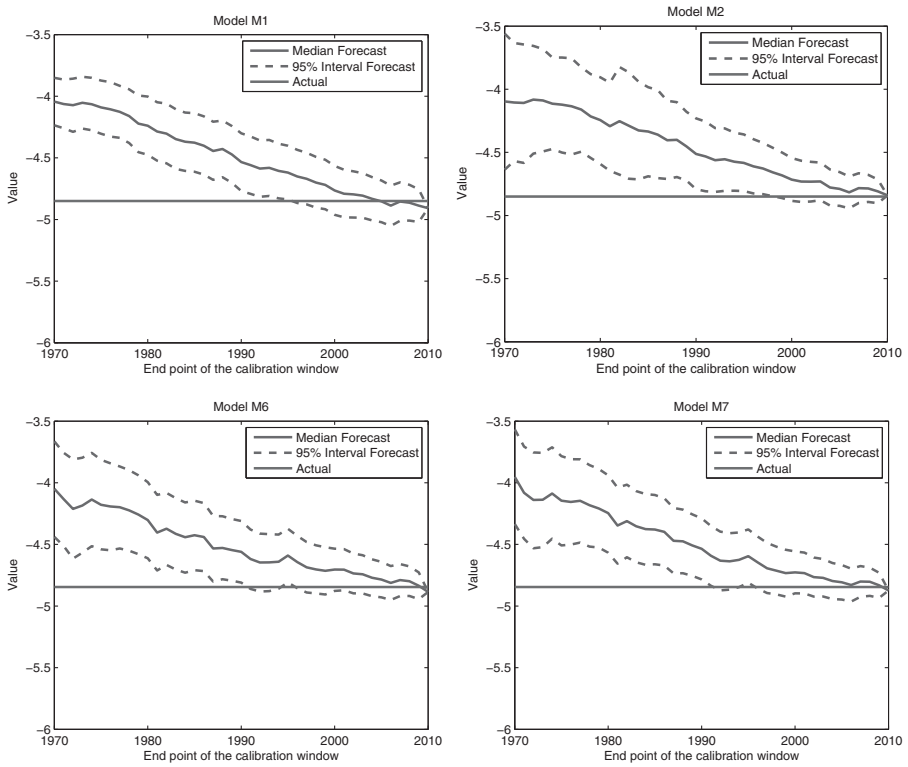

FIGURE 8: The median and 95% interval forecasts of $\ln(q_{60,2010}/1 - q_{60,2010})$, generated from Models M1, M2, M6 and M7 that are estimated to data over different calibration windows. The starting point of the calibration windows is always 1941 but end points range from 1971 to 2009.

The result of (ii) is reported in Figure 9. The LLCBD model generally yields ME and MSE with smaller magnitudes compared to any one of the alternative models. The benefit of using the LLCBD model is the most apparent when the end point of the calibration window is between 1980 and 1990.

The additional evaluation work indicates that the benefit of using a stochastic drift process cannot be obtained simply by using more dynamic factors or tweaking the age-effect structures. Of course, it may be possible to further improve the forecasting performance by adding more dynamic factors (e.g., a cohort effect) to the LLCBD model. These possible extensions are left for future research.

4.2. Sensitivity to the choice of age range

The baseline estimation result is based on an age range of $[x_a, x_b] = [50, 89]$. This age range is chosen for the following reasons.

First, the age range of [50, 89] is often used in the literature (see, e.g., Cairns *et al.*, 2009) to calibrate stochastic mortality models for pension and annuity

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

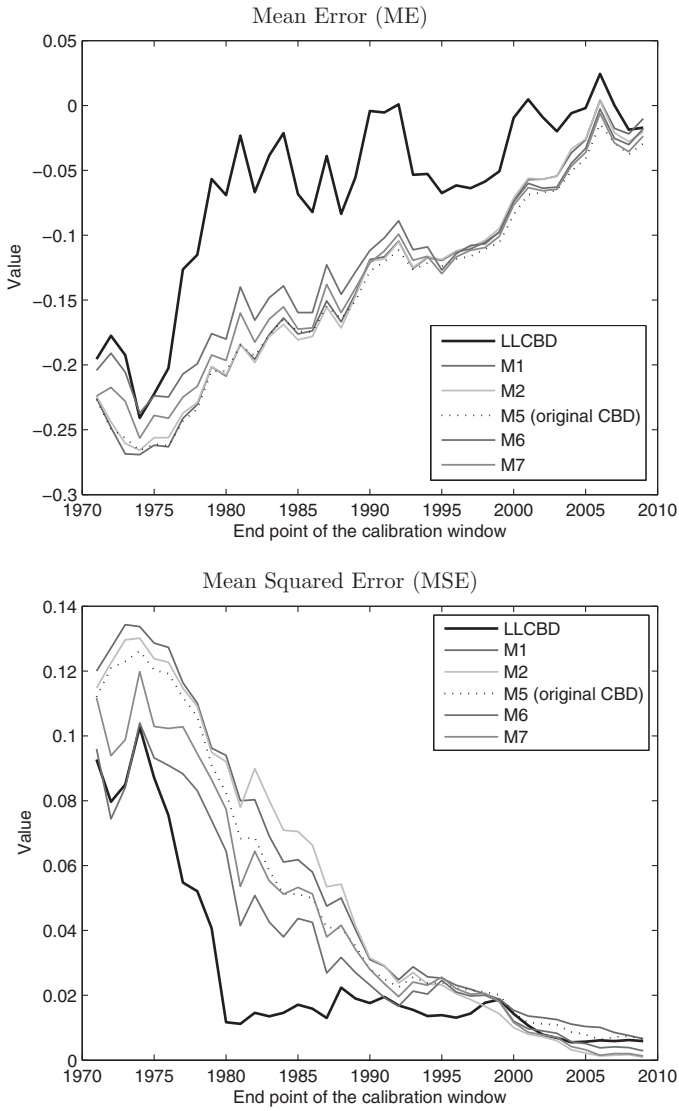

FIGURE 9: The Mean Error (ME) and Mean Squared Error (MSE) for the forecasts of $\ln(q_{x,t}/1-q_{x,t})$ produced by the LLCBD model and Models M1, M2, M5 (the original CBD), M6 and M7.

valuations. Using this age range enables readers to compare the estimation results in this and other papers more readily.

Second, according to the Human Mortality Database documentation (Wilmoth *et al.*, 2005), raw population counts for Canadians are available up to age 89 only. Population counts beyond age 89 are not "real" but estimated using the extinct cohort method.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

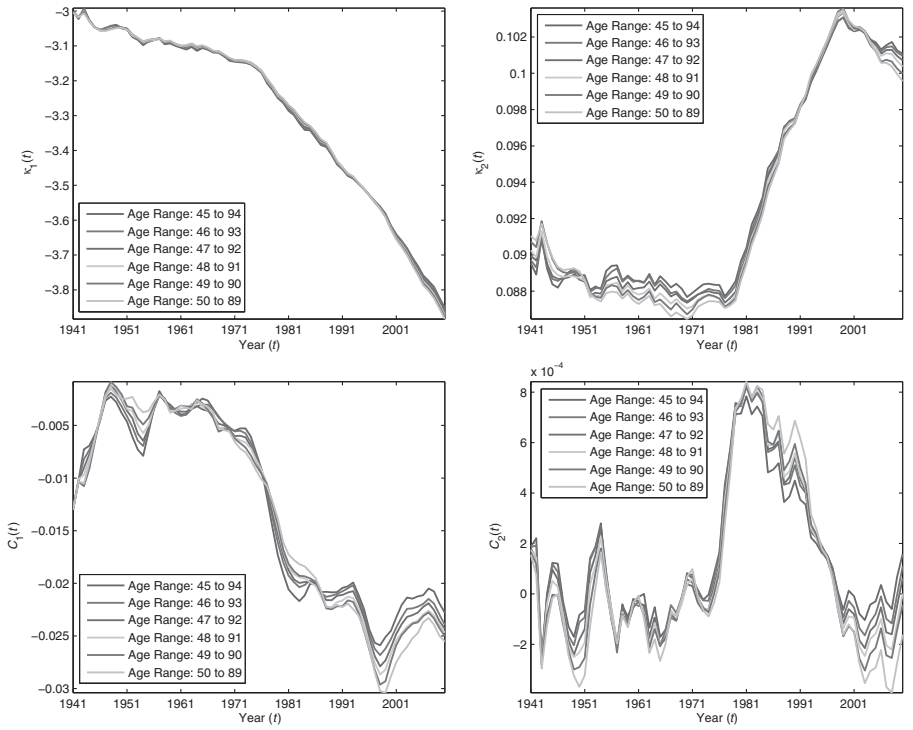

FIGURE 10: The retrieved values of the hidden states, $\kappa_1(t)$, $\kappa_2(t)$, $C_1(t)$ and $C_2(t)$, in the LLCBD model when different age ranges are used in estimation.

Third, as the models are built for modeling longevity risk at pensionable ages, beginning the sample age range at 50 prevents the models from being influenced by the (possibly different) mortality improvement dynamics at younger ages. Also, the age effect structure in the CBD/LLCBD model does not capture the accident hump at younger ages.

In this sub-section, we examine how the estimation results may be different if different age ranges are used. We consider the following six age ranges: 45–94, 46–93, 47–92, 48–91, 49–90 and 50–89 (the baseline age range). Note that the average age over each of the six age ranges is $\bar{x} = 69.5$. It is necessary to keep $\bar{x}$ fixed, because otherwise $\kappa_1(t)$ and $\kappa_2(t)$ would be scaled differently.

Figure 10 shows the retrieved values of the hidden states, $\kappa_1(t)$, $\kappa_2(t)$, $C_1(t)$ and $C_2(t)$, in the LLCBD model for each age range under consideration. For most of the time, the retrieved hidden states are robust relative to the choice of age range. Over the last 10 years of the calibration window, the retrieved drifts are arguably quite sensitive to the age range used. However, compared to the widths of the confidence intervals (Figure 3), the changes in the retrieved drifts due to changes in age range are small.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 6

THE LMPI TEST RESULTS (TEST STATISTIC AND CRITICAL VALUES AT 5% AND 10% LEVELS OF SIGNIFICANCE) FOR $\kappa_1(t)$ AND $\kappa_2(t)$ ESTIMATED FROM THE FOUR ADDITIONAL DATA SETS.

	Test Statistic		Critical Value	
Data Set	$\kappa_1(t)$	$\kappa_2(t)$	5%	10%
Dutch Male	1.4714	0.5775	0.4709	0.3503
English and Welsh Male	0.5909	0.0312	0.4689	0.3485
Japanese Unisex	0.0450	0.5258	0.4709	0.3503
Canadian Female	0.0210	0.0567	0.4689	0.3485

4.3. Application to other data sets

It would be interesting to see if stochastic drifts apply to the mortality dynamics of other populations, and if the LLCBD model still outperforms when it is fitted to other data sets. In this sub-section, we apply the testing and modeling methods to the following additional data sets:

Population	Age Range	Sample Period
Dutch Male	50 to 89	1951 to 2010
English and Welsh Male	50 to 89	1941 to 2010
Japanese Unisex	50 to 89	1951 to 2010
Canadian Female	50 to 89	1941 to 2010

These data sets cover different geographical locations (Europe, North America and Asia), genders (male, female and unisex) and sample periods (1941–2010 and 1951–2010). They are also obtained from the Human Mortality Database.

First, we apply the LMPI test to the least square estimates of the CBD dynamic factors $\kappa_1(t)$ and $\kappa_2(t)$. The test results are reported in Table 6. For Dutch males, the test results indicate that the drifts for both $\kappa_1(t)$ and $\kappa_2(t)$ are stochastic. More interestingly, the test results suggest that it is possible to have only one drift being stochastic ($\kappa_1(t)$ for English and Welsh males and $\kappa_2(t)$ for Japanese unisex). For Canadian females, the test results conclude that none of the two drifts is stochastic. This conclusion highlights a noteworthy fact: Just as cohort effect (which is highly significant in the United Kingdom but not so much in Asian populations) and jump effect (which can only be detected if the calibration window is long enough), stochastic drift is not a universal phenomenon. Whether a stochastic drift is needed depends critically on the data set considered.

Next, we use the AIC to compare the goodness-of-fit produced by the LLCBD models (with one and two stochastic drift(s)) and the original CBD model. The results are shown in Table 7. For Dutch males, the optimal model is the LLCBD model with two stochastic drifts; for English and Welsh males,

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 7

THE VALUES OF $\mathcal{N}$, $\ln(\hat{\mathcal{L}})$ AND AIC FOR THE MODELS FITTED TO THE FOUR ADDITIONAL DATA SETS. ("CBD": THE ORIGINAL CBD MODEL; "LLCBD": THE LLCBD MODEL WITH TWO STOCHASTIC DRIFTS; "LLCBD*": THE LLCBD MODEL WITH C_2 BEING CONSTANT; "LLCBD**": THE LLCBD MODEL WITH C_1 BEING CONSTANT.)

	Dutch Male			English and Welsh Male		
Model	$\ln(\hat{\mathcal{L}})$	AIC	$\mathcal{N}$	$\ln(\hat{\mathcal{L}})$	AIC	$\mathcal{N}$
CBD	3547.7047	−7087.4093	4	4146.5332	−8285.0664	4
LLCBD*	3556.3097	−7098.6194	7	4149.9301	**−8285.8603**	7
LLCBD**	3554.6960	−7095.3920	7	4149.4382	−8284.8764	7
LLCBD	3562.7838	**−7103.5676**	11	4153.3105	−8284.6209	11
	Japanese Unisex			Canadian Female		
Model	$\ln(\hat{\mathcal{L}})$	AIC	$\mathcal{N}$	$\ln(\hat{\mathcal{L}})$	AIC	$\mathcal{N}$
CBD	2662.9849	−5317.9698	4	3304.2798	**−6600.55951**	4
LLCBD*	2666.4540	−5318.9079	7	3305.5235	-6597.04696	7
LLCBD**	2671.2966	**−5328.5931**	7	3306.0631	−6598.12613	7
LLCBD	2671.7738	−5321.5476	11	3307.2052	−6592.41038	11

the optimal model is the LLCBD model with C_2 being constant; for Japanese unisex, the optimal model is the LLCBD model with C_1 being constant and for Canadian females, the optimal model is the original CBD model. These conclusions are in line with the LMPI test results.

For all of the four populations except Canadian females, we compare the forecasts generated by the original CBD model and the model chosen according to the AIC. Figure 11 displays the results of the "contracting horizon backtest" (used in Sections 3.4 and 4.1) for the three populations. In all cases, the LLCBD model (or its variant) yields a median forecast that is less biased and a 95% prediction interval that captures a larger proportion of the actual values.

Figure 12 shows the values of ME and MSE produced by the "forecasts" generated from models that are estimated to restricted calibration windows (1941–1971, …, 1941–2009 for English and Welsh males; 1951–1971, …, 1951–2009 for Dutch males and Japanese unisex). For Dutch males, the LLCBD model — which contains two stochastic drifts — yields significantly higher forecast accuracy compared to the original CBD model. For the other two populations, the LLCBD models — of which the drifts are only partially stochastic — still provide improved forecast accuracy, but the improvement is not that remarkable.

Finally, we evaluate the robustness of the models relative to the length of the calibration window used. The results are provided in Figure 13. As what we found in Sections 3.4 and 4.1, the LLCBD model (or its variant) is more robust than the original CBD model. It also yields forecasts that are more consistent with the recent trend.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

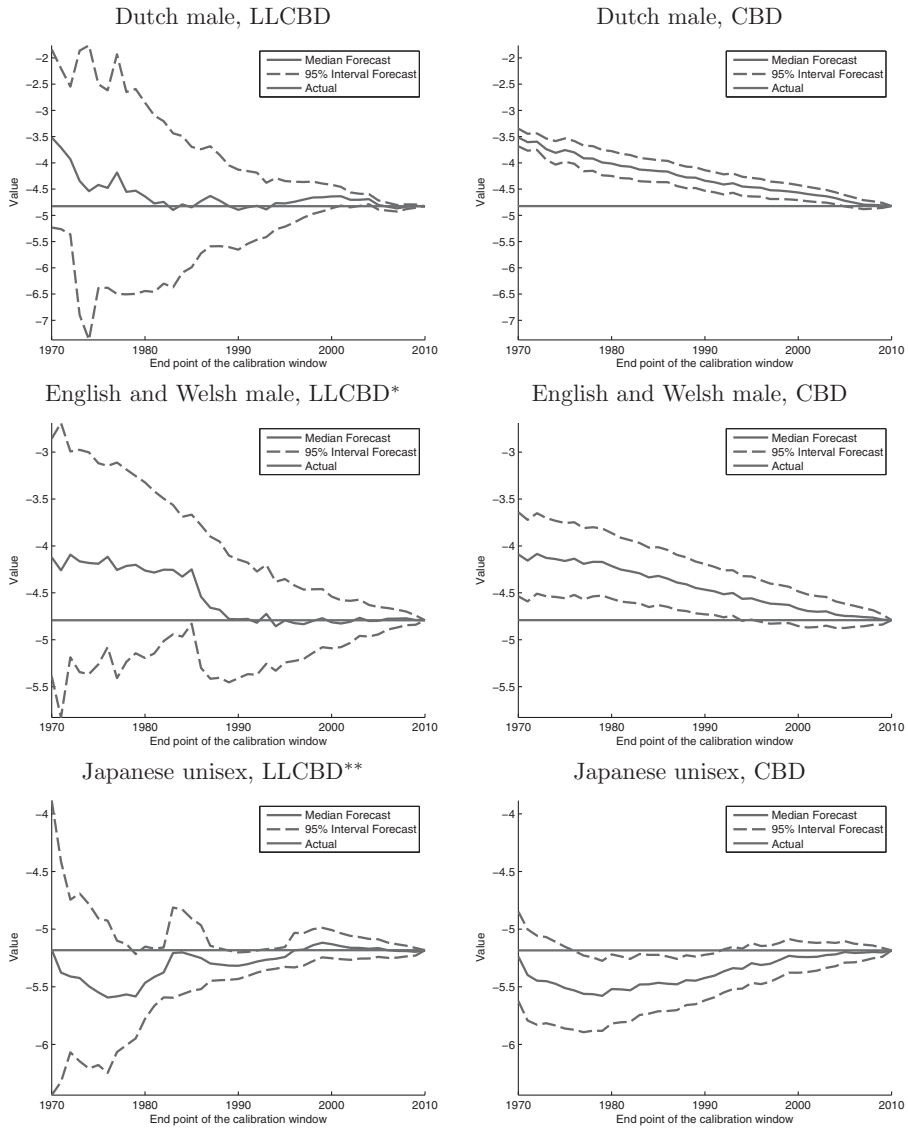

FIGURE 11: The median and 95% interval forecasts of $\ln(q_{60,2010}/1-q_{60,2010})$, generated from models that are estimated to data over different calibration windows. The starting point of the calibration windows is either 1951 (Dutch male and Japanese unisex) or 1941 (English and Welsh male), but end points range from 1971 to 2009.

5. HEDGING DRIFT AND DIFFUSION RISKS

Having modeled both drift and diffusion risks with the LLCBD model, in this section, we explain how we may hedge these risks by using standardized hedging

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

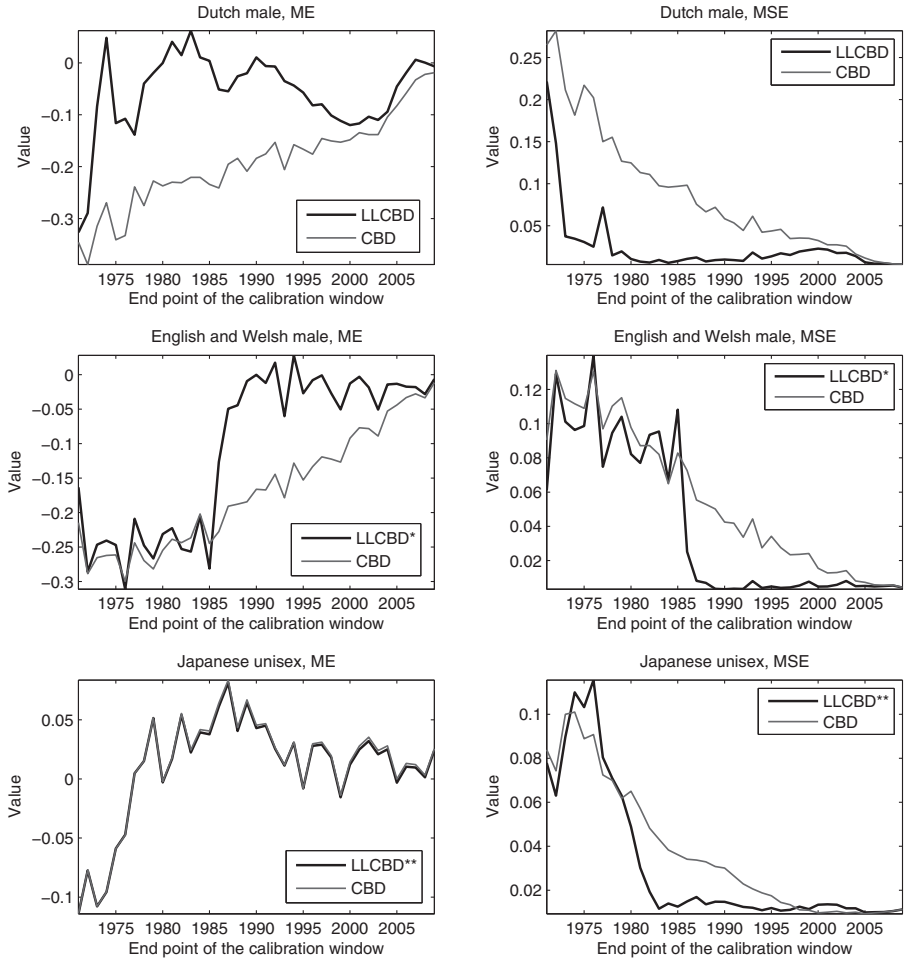

FIGURE 12: The Mean Error (ME) and Mean Squared Error (MSE) for the forecasts of $\ln(q_{x,t}/(1-q_{x,t})$ produced by the original CBD model and the optimal LLCBD model.

instruments. We start with a sub-section which details the assumptions made, followed by a review of the traditional delta and delta–nuga hedging methods. We then introduce our proposed "generalized state-space hedging method", and conclude with some comments about the distinctions between the traditional and newly proposed methods.

5.1. The set-up

Our goal is to hedge the longevity risk associated with a T-year temporary life annuity immediate that is just sold. Suppose that it is now time t_0 (i.e., the end of year t_0) and that the annuitant is now aged x_0. Ignoring sampling risk, the

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

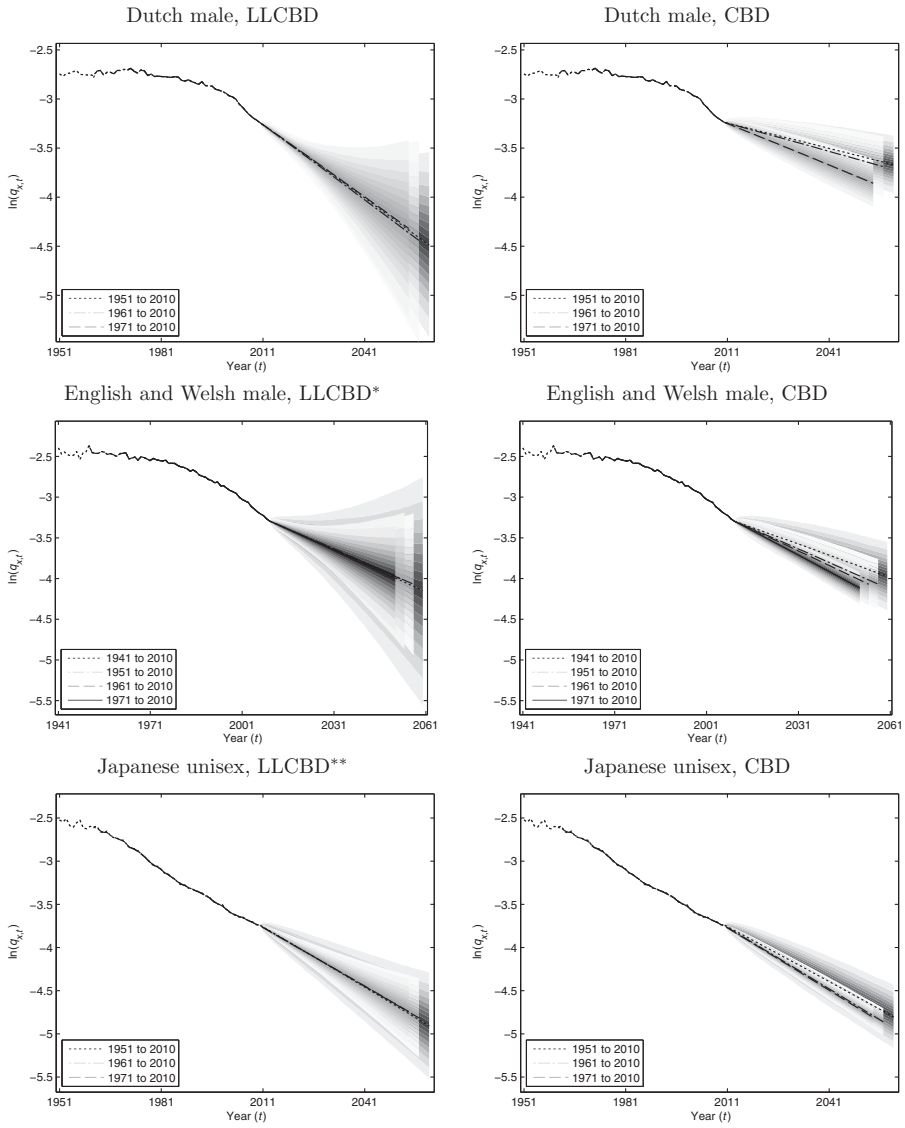

FIGURE 13: Forecasts of $\ln(q_{x,t})$ for $x = 75$ produced by the original CBD model and the optimal LLCBD model fitted to data over different calibration windows.

(random) present value of the liability being hedged is

$$L = \sum_{u=1}^{T} e^{-ru} \prod_{t=t_0+1}^{t_0+u} \tilde{p}_{x_0+t-t_0-1,t}$$

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

per contract, where r is the interest rate for discounting purposes,

$$\tilde{p}_{x,t} = 1 - \tilde{q}_{x,t}$$

and, as defined in Equation (3.2), $\tilde{q}_{x,t}$ is the underlying (unobserved) conditional probability of death in year t (between age x and $x+1$). Note that L is a function of $\kappa_1(t_0+1), \ldots, \kappa_1(t_0+T)$ and $\kappa_2(t_0+1), \ldots, \kappa_2(t_0+T)$, all of which are random as of the time when the longevity hedge is established.

The standardized hedging instruments used are q-forwards. A q-forward is a zero-coupon swap with its floating leg proportional to the realized death probability at a certain reference age in a certain reference year and its fixed leg proportional to the corresponding pre-determined forward mortality rate.

We consider a hedge portfolio of $m \geq 1$ q-forwards. Let x_j, t_j and $q^f_{x_j,t_j}$ be the reference age, reference year and forward mortality rate for the jth q-forward contract, respectively. We assume $t_0 < t_1 < \ldots < t_m$. We also assume that payment exchanges hands at the end of the reference year, so that $t_j - t_0$ is the time-to-maturity for the jth q-forward. For simplicity, it is assumed in the derivations that there is no population basis risk; that is, the q-forwards and the liability being hedged are associated with exactly the same population of individuals. In today's market, the LLMA's[4] LifeMetrics indexes, to which standardized q-forwards are linked, are based on smoothed (rather than crude) age-specific conditional death probabilities. For this reason, we assume that the q-forwards are linked to the underlying death probabilities that are not subject to the randomness of $\epsilon_{x,t}$. Under the mentioned assumptions, the (random) present value of the payoff from the jth q-forward contract is given by

$$H_j = e^{-r(t_j - t_0)}(q^f_{x_j,t_j} - \tilde{q}_{x_j,t_j}), \quad j = 1, \ldots, m.$$

It is obvious that H_j depends on $\kappa_1(t_j)$ and $\kappa_2(t_j)$, which are random variables as of the time when the longevity hedge is established.

We let N_j be the notional amount of the jth q-forward and

$$P = L - \sum_{j=1}^{m} N_j H_j$$

be the (random) present value of all cash flows when the longevity hedge is in place. If the longevity hedge is successful, then the variability in P would be significantly less than that in L. Using this reasoning, we assess hedge effectiveness with the following metric:

$$HE = 1 - \frac{\mathrm{Var}(P)}{\mathrm{Var}(L)}.$$

The value of HE is close to 1 if the longevity hedge is effective, and close to 0 if otherwise. This metric is also used by Cairns (2011, 2013), Cairns *et al.* (2014), Coughlan *et al.* (2011) and Li and Hardy (2011).

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

5.2. A review of traditional delta and delta–nuga hedging methods

5.2.1. *Delta hedging.* The idea behind the traditional delta hedging method, considered by researchers including Cairns (2011, 2013) and Zhou and Li (2014), is to match the sensitivities of the liability being hedged and the hedging portfolio with respect to changes in the time-varying factors in the assumed stochastic mortality model in year t_0. If the CBD model is assumed, then the sensitivities involved are represented by the following partial derivatives: $\partial L/\partial \kappa_i(t_0)$, $\partial H_j/\partial \kappa_i(t_0)$, $i = 1, 2$, $j = 1, \ldots, m$.

Because L and H_j are functions of $\kappa_1(t)$ and $\kappa_2(t)$ for $t > t_0$ rather than $\kappa_1(t_0)$ and $\kappa_2(t_0)$, the partial derivatives of L and H_j with respect to $\kappa_1(t_0)$ and $\kappa_2(t_0)$ cannot be computed straightforwardly. To make the estimation of sensitivities possible, the partial derivatives are calculated using the best estimates of L and H_j on the basis of $\kappa_1(t_0)$ and $\kappa_2(t_0)$:

$$\hat{L} = \sum_{u=1}^{T} e^{-ru} \prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0+t-t_0-1,t} \quad \text{and} \quad \hat{H}_j = e^{-r(t_j-t_0)}(q^f_{x_j,t_j} - \hat{q}_{x_j,t_j}), \tag{5.1}$$

where $\hat{q}_{x,t} = \frac{e^{\hat{\kappa}_1(t)+(x-\bar{x})\hat{\kappa}_2(t)}}{1+e^{\hat{\kappa}_1(t)+(x-\bar{x})\hat{\kappa}_2(t)}}$, $\hat{p}_{x,t} = 1 - \hat{q}_{x,t}$ and

$$\hat{\kappa}_i(t) = \kappa_i(t_0) + C_i(t_0) \times (t - t_0), \quad t > t_0, \quad i = 1, 2,$$

is the best estimate of $\kappa_i(t)$ given $\kappa_i(t_0)$ and $C_i(t_0)$.[5]

The partial derivatives of $\hat{L}$ and $\hat{H}_j$ with respect to $\kappa_1(t_0)$ and $\kappa_2(t_0)$ can be computed readily as follows:

$$\frac{\partial \hat{L}}{\partial \kappa_1(t_0)} = -\sum_{u=1}^{T} e^{-ru} \left(\sum_{t=t_0+1}^{t_0+u} \hat{q}_{x_0+t-t_0-1,t} \right) \left(\prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0+t-t_0-1,t} \right);$$

$$\frac{\partial \hat{L}}{\partial \kappa_2(t_0)} = -\sum_{u=1}^{T} e^{-ru} \left(\sum_{t=t_0+1}^{t_0+u} (x_0 + t - t_0 - 1 - \bar{x}) \hat{q}_{x_0+t-t_0-1,t} \right) \times \left(\prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0+t-t_0-1,t} \right);$$

$$\frac{\partial \hat{H}_j}{\partial \kappa_1(t_0)} = -e^{-r(t_j-t_0)} \hat{p}_{x_j,t_j} \hat{q}_{x_j,t_j};$$

$$\frac{\partial \hat{H}_j}{\partial \kappa_2(t_0)} = -e^{-r(t_j-t_0)} (x_j - \bar{x}) \hat{p}_{x_j,t_j} \hat{q}_{x_j,t_j}.$$

These derivatives are regarded as the "deltas" of the liability being hedged and the hedging instruments.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

We require exactly two hedging instruments to obtain a delta-neutral position. The notional amounts of the two q-forwards in the hedge portfolio should satisfy the following system of equations:

$$\begin{cases} \dfrac{\partial \hat{L}}{\partial \kappa_1(t_0)} = N_1 \times \dfrac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} + N_2 \times \dfrac{\partial \hat{H}_2}{\partial \kappa_1(t_0)} \\ \dfrac{\partial \hat{L}}{\partial \kappa_2(t_0)} = N_1 \times \dfrac{\partial \hat{H}_1}{\partial \kappa_2(t_0)} + N_2 \times \dfrac{\partial \hat{H}_2}{\partial \kappa_2(t_0)} \end{cases},$$

the solution to which is given by

$$\begin{pmatrix} N_1 \\ N_2 \end{pmatrix} = \begin{pmatrix} \dfrac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \frac{\partial \hat{H}_2}{\partial \kappa_1(t_0)} \\ \dfrac{\partial \hat{H}_1}{\partial \kappa_2(t_0)} & \frac{\partial \hat{H}_2}{\partial \kappa_2(t_0)} \end{pmatrix}^{-1} \begin{pmatrix} \dfrac{\partial \hat{L}}{\partial \kappa_1(t_0)} \\ \dfrac{\partial \hat{L}}{\partial \kappa_2(t_0)} \end{pmatrix},$$

provided that the inverse of the square matrix on the right-hand side exists. The invertibility of the square matrix is discussed in Section 5.5.

5.2.2. *Delta–nuga hedging.* Cairns (2013) acknowledged that the values of the liability being hedged and the hedge portfolio are also affected by changes in the estimated values of the drift terms in the processes for $\kappa_1(t)$ and $\kappa_2(t)$. To mitigate this piece of uncertainty, he proposed the delta–nuga hedging method in which the sensitivities of $\hat{L}$ and $\hat{H}_j$ to the changes in the drift terms are also considered. This method includes, additionally, the following four partial derivatives:

$$\frac{\partial \hat{L}}{\partial C_1(t_0)} = -\sum_{u=1}^{T} e^{-ru} \left(\sum_{t=t_0+1}^{t_0+u} (t-t_0) \hat{q}_{x_0+t-t_0-1,t} \right) \left(\prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0+t-t_0-1,t} \right);$$

$$\frac{\partial \hat{L}}{\partial C_2(t_0)} = -\sum_{u=1}^{T} e^{-ru} \left(\sum_{t=t_0+1}^{t_0+u} (t-t_0)(x_0+t-t_0-1-\bar{x}) \hat{q}_{x_0+t-t_0-1,t} \right) \times \left(\prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0+t-t_0-1,t} \right);$$

$$\frac{\partial \hat{H}_j}{\partial C_1(t_0)} = -e^{-r(t_j-t_0)} (t_j - t_0) \hat{p}_{x_j,t_j} \hat{q}_{x_j,t_j};$$

$$\frac{\partial \hat{H}_j}{\partial C_2(t_0)} = -e^{-r(t_j-t_0)} (t_j - t_0)(x_j - \bar{x}) \hat{p}_{x_j,t_j} \hat{q}_{x_j,t_j}.$$

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

These additional partial derivatives are considered as the "nugas" of the liability being hedged and the hedging instruments. To neutralize the deltas and nugas, we need exactly four hedging instruments. The notional amounts of the four hedging instruments should satisfy the following system of equations:

$$\begin{cases} \dfrac{\partial \hat{L}}{\partial \kappa_1(t_0)} = N_1 \times \dfrac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} + N_2 \times \dfrac{\partial \hat{H}_2}{\partial \kappa_1(t_0)} + N_3 \times \dfrac{\partial \hat{H}_3}{\partial \kappa_1(t_0)} + N_4 \times \dfrac{\partial \hat{H}_4}{\partial \kappa_1(t_0)}, \\ \dfrac{\partial \hat{L}}{\partial \kappa_2(t_0)} = N_1 \times \dfrac{\partial \hat{H}_1}{\partial \kappa_2(t_0)} + N_2 \times \dfrac{\partial \hat{H}_2}{\partial \kappa_2(t_0)} + N_3 \times \dfrac{\partial \hat{H}_3}{\partial \kappa_2(t_0)} + N_4 \times \dfrac{\partial \hat{H}_4}{\partial \kappa_2(t_0)}, \\ \dfrac{\partial \hat{L}}{\partial C_1(t_0)} = N_1 \times \dfrac{\partial \hat{H}_1}{\partial C_1(t_0)} + N_2 \times \dfrac{\partial \hat{H}_2}{\partial C_1(t_0)} + N_3 \times \dfrac{\partial \hat{H}_3}{\partial C_1(t_0)} + N_4 \times \dfrac{\partial \hat{H}_4}{\partial C_1(t_0)}, \\ \dfrac{\partial \hat{L}}{\partial C_2(t_0)} = N_1 \times \dfrac{\partial \hat{H}_1}{\partial C_2(t_0)} + N_2 \times \dfrac{\partial \hat{H}_2}{\partial C_2(t_0)} + N_3 \times \dfrac{\partial \hat{H}_3}{\partial C_2(t_0)} + N_4 \times \dfrac{\partial \hat{H}_4}{\partial C_2(t_0)}, \end{cases}$$

which implies

$$\begin{pmatrix} N_1 \\ \vdots \\ N_4 \end{pmatrix} = \begin{pmatrix} \dfrac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \cdots & \dfrac{\partial \hat{H}_4}{\partial \kappa_1(t_0)} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial \hat{H}_1}{\partial C_2(t_0)} & \cdots & \dfrac{\partial \hat{H}_4}{\partial C_2(t_0)} \end{pmatrix}^{-1} \begin{pmatrix} \dfrac{\partial \hat{L}}{\partial \kappa_1(t_0)} \\ \vdots \\ \dfrac{\partial \hat{L}}{\partial C_2(t_0)} \end{pmatrix}, \tag{5.2}$$

provided that the inverse of the square matrix on the right-hand side exists. The invertibility of the square matrix is discussed in Section 5.5.

5.3. The generalized state-space hedging method

The review presented in the previous sub-section exposes two limitations of the traditional delta and delta–nuga hedging methods. First, in deriving the deltas and nugas, it is assumed that $\kappa_1(t)$ and $\kappa_2(t)$ for $t > t_0$ are linear functions of $\kappa_1(t_0), \kappa_2(t_0), C_1(t_0)$ and $C_2(t_0)$. The resulting deltas and nugas therefore contain no information about the sensitivities of L and H_j to $\kappa_1(t)$ and $\kappa_2(t)$ for any $t > t_0$. Second, the traditional methods have very stringent requirements on the number of hedging instruments being used. The delta hedging method requires exactly two distinct hedging instruments whereas the delta–nuga hedging needs exactly four.

To overcome these limitations, we hereby introduce the generalized state-space hedging method which better utilizes the information contained in the hidden states at different times and is more flexible in terms of the number of hedging instruments required. In the generalized state-space hedging method, we work on L and H_j (rather than their best estimates) and preserve the fact that they are functions of $\kappa_1(t)$ and $\kappa_2(t)$ for $t > t_0$.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 8

A SUMMARY OF THE DISTINCTIONS AMONG L, $\hat{L}$, l, H_j, $\hat{H}_j$ AND h_j.

L	A Non-Linear Function of $\vec{\alpha}^*_{t_0+1}, \ldots, \vec{\alpha}^*_{t_0+T}$
$\hat{L}$	An Approximation of L, Obtained by Setting $\kappa_i(t) = \kappa_i(t_0) + C_1(t_0)(t - t_0)$ for $i = 1, 2$ and $t > t_0$; $\hat{L}$ is a Non-Linear Function of $\kappa_1(t_0)$ and $\kappa_2(t_0)$ (i.e., $\vec{\alpha}^*_{t_0}$)
l	An Approximation of L Based on a First-Order Taylor's Expansion Around $\hat{L}$; l is a Linear Function of $\vec{\alpha}^*_{t_0+1}, \ldots, \vec{\alpha}^*_{t_0+T}$
H_j	A Non-Linear Function of $\vec{\alpha}^*_{t_j}$
$\hat{H}_j$	An Approximation of H_j, Obtained by Setting $\kappa_i(t_j) = \kappa_i(t_0) + C_1(t_0)(t_j - t_0)$ for $i = 1, 2$; $\hat{H}_j$ is a Non-Linear Function of $\kappa_1(t_0)$ and $\kappa_2(t_0)$ (i.e., $\vec{\alpha}^*_{t_0}$)
h_j	An Approximation of H_j Based on a First-Order Taylor's Expansion Around $\hat{H}_j$; h_j is a Linear Function of $\vec{\alpha}^*_{t_j}$

We let $\vec{\alpha}^*_t = (\kappa_1(t), \kappa_2(t))'$ and $\vec{\alpha}^{**}_t = (C_1(t), C_2(t))'$. It follows that L is a function of the sequence of $\{\vec{\alpha}^*_{t_0+1}, \ldots, \vec{\alpha}^*_{t_0+T}\}$ and H_j is a function of $\vec{\alpha}^*_{t_j}$, $j = 1, \ldots, m$. The derivation of the generalized state-space hedging strategy involves the first-order Taylor approximations of L and H_j about all relevant vectors of hidden states. For L, the first-order approximation $l(\vec{\alpha}^*_{t_0+1}, \ldots, \vec{\alpha}^*_{t_0+T})$ is given by

$$L \approx l(\vec{\alpha}^*_{t_0+1}, \ldots, \vec{\alpha}^*_{t_0+T}) = \hat{L} + \sum_{i=t_0+1}^{t_0+T} \left(\frac{\partial L}{\partial \vec{\alpha}^*_i}\right)' (\vec{\alpha}^*_i - \vec{\hat{\alpha}}^*_i),$$

where $\hat{L}$ is defined in Equation (5.1) and $\vec{\hat{\alpha}}^*_i$ is the expected value of $\vec{\alpha}^*_i$ given the information up to and including year t_0. For H_j, $j = 1, \ldots, m$, the first-order approximation $h_j(\vec{\alpha}^*_{t_j})$ is given by

$$H_j \approx h_j(\vec{\alpha}^*_{t_j}) = \hat{H}_j + \left(\frac{\partial H_j}{\partial \vec{\alpha}^*_{t_j}}\right)' (\vec{\alpha}^*_{t_j} - \vec{\hat{\alpha}}^*_{t_j}), \tag{5.3}$$

where $\hat{H}_j$ is defined in Equation (5.1). For brevity, we suppress the arguments of l and h_j in the rest of this paper. Also, unless otherwise specified, the partial derivatives $\partial L/\partial \vec{\alpha}^*_i$ and $\partial H_j/\partial \vec{\alpha}^*_{t_j}$ are evaluated at $\vec{\alpha}^*_i = \vec{\hat{\alpha}}^*_i$ and $\vec{\alpha}^*_{t_j} = \vec{\hat{\alpha}}^*_{t_j}$, respectively. To facilitate exposition, in Table 8 we summarize the distinctions among L, $\hat{L}$, l, H_j, $\hat{H}_j$ and h_j.

The hedging strategy is obtained by minimizing the variance of $l - \sum_{i=1}^m N_i h_i$ as of the time when the hedge is established. The variance to be minimized can

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

be expressed as

$$\begin{aligned}
&\mathrm{Var}\left(l - \sum_{i=1}^{m} N_i h_i\right) \\
&= \mathrm{Var}\left(\sum_{i=t_0+1}^{t_0+T} \left(\frac{\partial L}{\partial \vec{\alpha}_i^*}\right)' (\vec{\alpha}_i^* - \vec{\hat{\alpha}}_i^*) - \sum_{j=1}^{m} N_j \left(\frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*}\right)' (\vec{\alpha}_{t_j}^* - \vec{\hat{\alpha}}_{t_j}^*)\right) \\
&= \mathrm{Var}\left(\sum_{\substack{i=t_0+1 \\ i \neq t_1, t_2, \ldots, t_m}}^{t_0+T} \left(\frac{\partial L}{\partial \vec{\alpha}_i^*}\right)' (\vec{\alpha}_i^* - \vec{\hat{\alpha}}_i^*) + \sum_{j=1}^{m} \left(\frac{\partial L}{\partial \vec{\alpha}_{t_j}^*} - N_j \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*}\right)' (\vec{\alpha}_{t_j}^* - \vec{\hat{\alpha}}_{t_j}^*)\right).
\end{aligned}$$

It is interesting to note that $\mathrm{Var}(l - \sum_{i=1}^{m} N_i h_i)$ can be written as the sum of three components:

$$\mathrm{Var}\left(l - \sum_{i=1}^{m} N_i h_i\right) = V_1 + V_2 + V_3,$$

where

$$V_1 = \sum_{\substack{i,j=t_0+1 \\ i,j \neq t_1, t_2, \ldots, t_m}}^{t_0+T} \left(\frac{\partial L}{\partial \vec{\alpha}_i^*}\right)' \mathrm{Cov}(\vec{\alpha}_i^*, \vec{\alpha}_j^*) \left(\frac{\partial L}{\partial \vec{\alpha}_j^*}\right)$$

represents the variance contributed from the hidden states that are not related to the hedging instruments,

$$V_2 = \sum_{i=1}^{m} \sum_{j=1}^{m} \left(\frac{\partial L}{\partial \vec{\alpha}_{t_i}^*} - N_i \frac{\partial H_i}{\partial \vec{\alpha}_{t_i}^*}\right)' \mathrm{Cov}(\vec{\alpha}_{t_i}^*, \vec{\alpha}_{t_j}^*) \left(\frac{\partial L}{\partial \vec{\alpha}_{t_j}^*} - N_j \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*}\right)$$

represents the variance contributed from the hidden states that are directly related to the hedging instruments, and

$$V_3 = 2 \sum_{\substack{i=t_0+1 \\ i \neq t_1, t_2, \ldots, t_m}}^{t_0+T} \sum_{j=1}^{m} \left(\frac{\partial L}{\partial \vec{\alpha}_i^*}\right)' \mathrm{Cov}(\vec{\alpha}_i^*, \vec{\alpha}_{t_j}^*) \left(\frac{\partial L}{\partial \vec{\alpha}_{t_j}^*} - N_j \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*}\right)$$

represents the variance contributed from the interaction between the states that are related and unrelated to the hedging instruments. Note that V_1 is free of N_j, $j = 1, \ldots, m$. We can interpret V_1 to mean the risk that cannot be hedged by the collection of m hedging instruments.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

All partial derivatives involved in $\mathrm{Var}(l - \sum_{i=1}^{m} N_i h_i)$ can be computed analytically. First, we rewrite L in terms of $\vec{\alpha}_t^*$, $t = t_0 + 1, \ldots, t_0 + T$, as

$$L = \sum_{u=1}^{T} e^{-ru} \prod_{t=t_0+1}^{t_0+u} \left(1 + \exp(M'_{x_0,t}\vec{\alpha}_t^*)\right)^{-1}, \tag{5.4}$$

where $M_{x_0,t} = (1, x_0 + t - t_0 - 1 - \bar{x})'$. The partial derivative of L with respect to $\vec{\alpha}_i^*$, evaluated at $\vec{\alpha}_i = \hat{\vec{\alpha}}_i$, can then be computed as

$$\frac{\partial L}{\partial \vec{\alpha}_i^*} = M_{x_0,i} \frac{\exp(M'_{x_0,i}\hat{\vec{\alpha}}_i^*)}{1 + \exp(M'_{x_0,i}\hat{\vec{\alpha}}_i^*)} \left(- \sum_{u=i-t_0}^{T} e^{-ru} \prod_{t=t_0+1}^{t_0+u} \left(1 + \exp(M'_{x_0,t}\hat{\vec{\alpha}}_t^*)\right)^{-1} \right).$$

Similarly, we can express H_j in terms of $\vec{\alpha}_{t_j}^*$ as

$$H_j = e^{-r(t_j - t_0)} \left(q^f_{x_j,t_j} - 1 + \left(1 + \exp((M^{(H)}_{x_j,t_j})'\vec{\alpha}_{t_j}^*)\right)^{-1} \right),$$

where $M^{(H)}_{x_j,t_j} = (1, x_j - \bar{x})'$. The partial derivative of H_j with respect to $\vec{\alpha}_{t_j}^*$, evaluated at $\vec{\alpha}_{t_j} = \hat{\vec{\alpha}}_{t_j}$, can then be calculated as

$$\frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*} = -e^{-r(t_j - t_0)} M^{(H)}_{x_j,t_j} \frac{\exp\left((M^{(H)}_{x_j,t_j})'\hat{\vec{\alpha}}_{t_j}^*\right)}{\left(1 + \exp\left((M^{(H)}_{x_j,t_j})'\hat{\vec{\alpha}}_{t_j}^*\right)\right)^2}.$$

The expression for $\mathrm{Var}(l - \sum_{i=1}^{m} N_i h_i)$ also involves $\mathrm{Cov}(\vec{\alpha}_i^*, \vec{\alpha}_j^*)$, the covariance matrix of $\vec{\alpha}_i^*$ and $\vec{\alpha}_j^*$. To compute $\mathrm{Cov}(\vec{\alpha}_i^*, \vec{\alpha}_j^*)$, we first calculate the covariance matrix of $\vec{\alpha}_i$ and $\vec{\alpha}_j$ as

$$\begin{aligned} \Xi_{i,j} &:= \mathrm{Cov}(\vec{\alpha}_i, \vec{\alpha}_j) \\ &= \begin{cases} A^{|i-j|}(Q + AQA' + \cdots A^{j-(t_0+1)} Q (A^{j-(t_0+1)})'), & i > j \\ (Q + AQA' + \cdots A^{i-(t_0+1)} Q (A^{i-(t_0+1)})')(A^{|i-j|})', & i < j \\ Q + AQA' + \cdots A^{j-(t_0+1)} Q (A^{j-(t_0+1)})', & i = j \end{cases}. \end{aligned} \tag{5.5}$$

Then, $\Xi_{i,j}$ is decomposed into four block matrices as follows:

$$\Xi_{i,j} = \begin{pmatrix} \mathrm{Cov}(\vec{\alpha}_i^*, \vec{\alpha}_j^*) & \mathrm{Cov}(\vec{\alpha}_i^*, \vec{\alpha}_j^{**}) \\ \mathrm{Cov}(\vec{\alpha}_i^{**}, \vec{\alpha}_j^*) & \mathrm{Cov}(\vec{\alpha}_i^{**}, \vec{\alpha}_j^{**}) \end{pmatrix}.$$

Finally, $\Xi^*_{i,j} = \mathrm{Cov}(\vec{\alpha}_i^*, \vec{\alpha}_j^*)$ can be obtained from the upper-left block matrix in $\Xi_{i,j}$.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

To derive the hedging strategy, we first take partial derivative of $\mathrm{Var}(l - \sum_{j=1}^{m} N_j h_j)$ with respect to N_i for $i = 1, 2, \ldots, m$:

$$\begin{aligned}
&\frac{\partial \mathrm{Var}(l - \sum_{j=1}^{m} N_j^* h_j)}{\partial N_i} \\
&= \frac{\partial V_2}{\partial N_i} + \frac{\partial V_3}{\partial N_i} \\
&= -2\sum_{j=1}^{m} \left(\frac{\partial H_i}{\partial \vec{\alpha}_{t_i}^*}\right)' \Xi_{t_i,t_j}^* \left(\frac{\partial L}{\partial \vec{\alpha}_{t_j}^*} - N_j \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*}\right) - 2 \sum_{\substack{j=t_0+1 \\ j \neq t_1, t_2, \ldots, t_m}}^{t_0+T} \left(\frac{\partial H_i}{\partial \vec{\alpha}_{t_i}^*}\right)' \Xi_{t_i,j}^* \frac{\partial L}{\partial \vec{\alpha}_j^*} \\
&= 2\sum_{j=1}^{m} N_j \left(\frac{\partial H_i}{\partial \vec{\alpha}_{t_i}^*}\right)' \Xi_{t_i,t_j}^* \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*} - 2 \sum_{j=t_0+1}^{t_0+T} \left(\frac{\partial H_i}{\partial \vec{\alpha}_{t_i}^*}\right)' \Xi_{t_i,j}^* \frac{\partial L}{\partial \vec{\alpha}_j^*}.
\end{aligned}$$

Then, the optimal hedging strategy is obtained by setting the partial derivatives to zero; that is,

$$\sum_{j=1}^{m} N_j \left(\frac{\partial H_1}{\partial \vec{\alpha}_{t_i}^*}\right)' \Xi_{t_i,t_j}^* \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*} = \sum_{j=t_0+1}^{t_0+T} \left(\frac{\partial H_1}{\partial \vec{\alpha}_{t_i}^*}\right)' \Xi_{t_i,j}^* \frac{\partial L}{\partial \vec{\alpha}_j^*}, \quad i = 1, \ldots, m.$$

This system of linear equations can be written in matrix form as

$$\begin{aligned}
&\begin{pmatrix}
\left(\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*}\right)' \Xi_{t_1,t_1}^* \frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*} & \cdots & \left(\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*}\right)' \Xi_{t_1,t_m}^* \frac{\partial H_m}{\partial \vec{\alpha}_{t_m}^*} \\
\vdots & \ddots & \vdots \\
\left(\frac{\partial H_m}{\partial \vec{\alpha}_{t_m}^*}\right)' \Xi_{t_m,t_1}^* \frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*} & \cdots & \left(\frac{\partial H_m}{\partial \vec{\alpha}_{t_m}^*}\right)' \Xi_{t_m,t_m}^* \frac{\partial H_m}{\partial \vec{\alpha}_{t_m}^*}
\end{pmatrix}
\begin{pmatrix} N_1 \\ \vdots \\ N_m \end{pmatrix} \\
&= \begin{pmatrix}
\sum_{j=t_0+1}^{t_0+T} \left(\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*}\right)' \Xi_{t_1,j}^* \frac{\partial L}{\partial \vec{\alpha}_j^*} \\
\vdots \\
\sum_{j=t_0+1}^{t_0+T} \left(\frac{\partial H_m}{\partial \vec{\alpha}_{t_m}^*}\right)' \Xi_{t_m,j}^* \frac{\partial L}{\partial \vec{\alpha}_j^*}
\end{pmatrix},
\end{aligned} \tag{5.6}$$

As to be explained in Section 5.5, the square matrix on the left-hand side of the equation above is always invertible. Hence, the values of $N_1, \ldots, N_m$ can be

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

obtained readily as follows:

$$\begin{pmatrix} N_1 \\ \vdots \\ N_m \end{pmatrix} = \begin{pmatrix} \left(\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}}\right)' \Xi^*_{t_1,t_1} \frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}} & \cdots & \left(\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}}\right)' \Xi^*_{t_1,t_m} \frac{\partial H_m}{\partial \vec{\alpha}^*_{t_m}} \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial H_m}{\partial \vec{\alpha}^*_{t_m}}\right)' \Xi^*_{t_m,t_1} \frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}} & \cdots & \left(\frac{\partial H_m}{\partial \vec{\alpha}^*_{t_m}}\right)' \Xi^*_{t_m,t_m} \frac{\partial H_m}{\partial \vec{\alpha}^*_{t_m}} \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \sum_{j=t_0+1}^{t_0+T} \left(\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}}\right)' \Xi^*_{t_1,j} \frac{\partial L}{\partial \vec{\alpha}^*_j} \\ \vdots \\ \sum_{j=t_0+1}^{t_0+T} \left(\frac{\partial H_m}{\partial \vec{\alpha}^*_{t_m}}\right)' \Xi^*_{t_m,j} \frac{\partial L}{\partial \vec{\alpha}^*_j} \end{pmatrix}. \tag{5.7}$$

It is noteworthy that the formula above contains $\Xi^*_{s,t}$ for $s, t = t_1, \ldots, t_m$. As such, the resulting hedging strategy incorporates the static and dynamic correlations between the hidden states in the observation equation.

5.4. Delta and delta–nuga hedging methods as special cases

In the generalized state-space hedging approach, we treat L and H_j as explicit functions of the state vectors $\vec{\alpha}^*_t$ for $t > t_0$. The analytical minimization of variance is made possible by approximating L and H_j with a first-order Taylor's expansion around all state vectors involved. In contrast, in the traditional delta and delta–nuga hedging methods, the hedging strategies are derived on the basis of the best estimates L and H_j, which depend exclusively on $\vec{\alpha}^*_{t_0} = (\kappa_1(t_0), \kappa_2(t_0))'$.

In what follows, we show that the traditional delta and delta–nuga hedging methods are indeed special cases of the proposed generalized state-space-hedging method when the following linear relations hold for $t > t_0 + 1$:

$$\begin{cases} \vec{\alpha}^*_t = \vec{\alpha}^*_{t_0+1} + (t - (t_0+1)) \times \vec{\alpha}^{**}_{t_0+1} \\ \vec{\alpha}^{**}_t = \vec{\alpha}^{**}_{t_0+1} \end{cases}. \tag{5.8}$$

5.4.1. *Connections with the traditional delta hedging method.* Suppose that the linear relations specified by Equation (5.8) hold. Assume further that $m = 2$ hedging instruments are used and that the vector of drift terms is constant; i.e., $\vec{\alpha}^{**}_t = \vec{\tilde{\alpha}}^{**}_{t_0}$ for $t \geq t_0 + 1$. Then, $\Xi^*_{i,j} = Q^*$ for $i, j = t_0 + 1, \ldots, (t_0 + T)$, where

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

Q^* is defined in Equation (3.3). It follows that Equation (5.6) can be reduced to

$$\begin{pmatrix} \left(\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}}\right)' Q^* \frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}} & \left(\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}}\right)' Q^* \frac{\partial H_2}{\partial \vec{\alpha}^*_{t_2}} \\ \left(\frac{\partial H_2}{\partial \vec{\alpha}^*_{t_2}}\right)' Q^* \frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}} & \left(\frac{\partial H_2}{\partial \vec{\alpha}^*_{t_2}}\right)' Q^* \frac{\partial H_2}{\partial \vec{\alpha}^*_{t_2}} \end{pmatrix} \begin{pmatrix} N_1 \\ N_2 \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=t_0+1}^{t_0+T} \left(\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}}\right)' Q^* \frac{\partial L}{\partial \vec{\alpha}^*_j} \\ \sum_{j=t_0+1}^{t_0+T} \left(\frac{\partial H_2}{\partial \vec{\alpha}^*_{t_2}}\right)' Q^* \frac{\partial L}{\partial \vec{\alpha}^*_j} \end{pmatrix}. \tag{5.9}$$

Also, we have

$$\begin{aligned} \sum_{i=t_0+1}^{t_0+T} \frac{\partial L}{\partial \vec{\alpha}^*_i} &= -\sum_{i=t_0+1}^{t_0+T} M_{x_0,i} \hat{q}_{x_0+i-t_0-1,i} \left(\sum_{u=i-t_0}^{T} e^{-ru} \prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0+t-t_0-1,t} \right) \\ &= -\sum_{u=1}^{T} e^{-ru} \prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0+t-t_0-1,t} \left(\sum_{t=t_0+1}^{t_0+u} M_{x_0,t} \hat{q}_{x_0+t-t_0-1,t} \right) \\ &= \frac{\partial \hat{L}}{\partial \vec{\alpha}^*_{t_0}} \end{aligned} \tag{5.10}$$

and

$$\frac{\partial H_j}{\partial \vec{\alpha}^*_{t_j}} = -e^{-r(t_j-t_0)} M^{(H)}_{x_j,t_j} \hat{q}_{x_j,t_j} \times \hat{p}_{x_j,t_j} = \frac{\partial \hat{H}_j}{\partial \vec{\alpha}^*_{t_0}} \tag{5.11}$$

for $j = 1, 2$.

Substituting Equations (5.10) and (5.11) into Equation (5.9), we immediately obtain

$$\begin{pmatrix} \left(\frac{\partial \hat{H}_1}{\partial \vec{\alpha}^*_{t_0}}\right)' \\ \left(\frac{\partial \hat{H}_2}{\partial \vec{\alpha}^*_{t_0}}\right)' \end{pmatrix} Q^* \begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \vec{\alpha}^*_{t_0}} & \frac{\partial \hat{H}_2}{\partial \vec{\alpha}^*_{t_0}} \end{pmatrix} \begin{pmatrix} N_1 \\ N_2 \end{pmatrix} = \begin{pmatrix} \left(\frac{\partial \hat{H}_1}{\partial \vec{\alpha}^*_{t_0}}\right)' \\ \left(\frac{\partial \hat{H}_2}{\partial \vec{\alpha}^*_{t_0}}\right)' \end{pmatrix} Q^* \frac{\partial \hat{L}}{\partial \vec{\alpha}^*_{t_0}},$$

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

or equivalently,

$$\begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \frac{\partial \hat{H}_1}{\partial \kappa_2(t_0)} \\ \frac{\partial \hat{H}_2}{\partial \kappa_1(t_0)} & \frac{\partial \hat{H}_2}{\partial \kappa_2(t_0)} \end{pmatrix} Q^* \begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \frac{\partial \hat{H}_2}{\partial \kappa_1(t_0)} \\ \frac{\partial \hat{H}_1}{\partial \kappa_2(t_0)} & \frac{\partial \hat{H}_2}{\partial \kappa_2(t_0)} \end{pmatrix} \begin{pmatrix} N_1 \\ N_2 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \frac{\partial \hat{H}_1}{\partial \kappa_2(t_0)} \\ \frac{\partial \hat{H}_2}{\partial \kappa_1(t_0)} & \frac{\partial \hat{H}_2}{\partial \kappa_2(t_0)} \end{pmatrix} Q^* \begin{pmatrix} \frac{\partial \hat{L}}{\partial \kappa_1(t_0)} \\ \frac{\partial \hat{L}}{\partial \kappa_2(t_0)} \end{pmatrix}.$$

By definition, Q^* is positive definite and hence invertible. If the matrix

$$\begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \frac{\partial \hat{H}_1}{\partial \kappa_2(t_0)} \\ \frac{\partial \hat{H}_2}{\partial \kappa_1(t_0)} & \frac{\partial \hat{H}_2}{\partial \kappa_2(t_0)} \end{pmatrix}$$

is also invertible, then we have

$$\begin{pmatrix} N_1 \\ N_2 \end{pmatrix} = \begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \frac{\partial \hat{H}_2}{\partial \kappa_1(t_0)} \\ \frac{\partial \hat{H}_1}{\partial \kappa_2(t_0)} & \frac{\partial \hat{H}_2}{\partial \kappa_2(t_0)} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial \hat{L}}{\partial \kappa_1(t_0)} \\ \frac{\partial \hat{L}}{\partial \kappa_2(t_0)} \end{pmatrix},$$

which yields the same values of N_1 and N_2 as those implied by the traditional delta hedging method. As Q^* is being canceled out in the derivation, the strategy developed by the delta hedging method does not incorporate any information concerning the variances and covariances of the hidden states.

5.4.2. *Connections with the traditional delta–nuga hedging method.* Suppose that the linear relations specified by Equation (5.8) hold and that $m = 4$ hedging

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

instruments are used. Equation (5.6) then becomes

$$\begin{pmatrix} \left(\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}}\right)' \Xi^*_{t_1,t_1} \frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}} & \cdots & \left(\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}}\right)' \Xi^*_{t_1,t_m} \frac{\partial H_k}{\partial \vec{\alpha}^*_{t_m}} \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial H_k}{\partial \vec{\alpha}^*_{t_m}}\right)' \Xi^*_{t_m,t_1} \frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}} & \cdots & \left(\frac{\partial H_k}{\partial \vec{\alpha}^*_{t_m}}\right)' \Xi^*_{t_m,t_m} \frac{\partial H_k}{\partial \vec{\alpha}^*_{t_m}} \end{pmatrix} \begin{pmatrix} N_1 \\ \vdots \\ N_4 \end{pmatrix}$$
$$= \begin{pmatrix} \sum_{j=t_0+1}^{t_0+T} \left(\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}}\right)' \Xi^*_{t_1,j} \frac{\partial L}{\partial \vec{\alpha}^*_j} \\ \vdots \\ \sum_{j=t_0+1}^{t_0+T} \left(\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_4}}\right)' \Xi^*_{t_4,j} \frac{\partial L}{\partial \vec{\alpha}^*_j} \end{pmatrix}. \tag{5.12}$$

It can be shown easily that under the assumptions made, the covariance matrix $\Xi_{i,j}$ can be written as

$$\Xi_{i,j} = A^{i-(t_0+1)} Q (A^{j-(t_0+1)})'$$

for $i, j = t_0 + 1, \ldots, (t_0 + T)$. Consequently, we have

$$\begin{aligned} \left(\frac{\partial H_i}{\partial \vec{\alpha}^*_{t_i}}\right)' \Xi^*_{t_i,t_j} \frac{\partial H_j}{\partial \vec{\alpha}^*_{t_j}} &= \left(\frac{\partial H_i}{\partial \kappa_1(t_i)} \ \frac{\partial H_i}{\partial \kappa_2(t_i)} \ 0 \ 0\right) \Xi_{t_i,t_j} \left(\frac{\partial H_j}{\partial \kappa_1(t_j)} \ \frac{\partial H_j}{\partial \kappa_1(t_j)} \ 0 \ 0\right)' \\ &= \left(\frac{\partial H_i}{\partial \vec{\alpha}_{t_i}}\right)' \Xi_{t_i,t_j} \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}} \\ &= \left(\frac{\partial H_i}{\partial \vec{\alpha}_{t_i}}\right)' A^{t_i-t_0} P (A^{t_j-t_0})' \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}}, \end{aligned}$$

for $i, j = 1, 2, \ldots, 4$, where $P = A^{-1} Q (A^{-1})'$. Also, we have

$$\begin{aligned} (A^{t_j-t_0})' \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ t_j - t_0 & 0 & 1 & 0 \\ 0 & t_j - t_0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial H_j}{\partial \kappa_1(t_j)} \\ \frac{\partial H_j}{\partial \kappa_2(t_j)} \\ 0 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ x_j - \bar{x} \\ t_j - t_0 \\ (t_j - t_0)(x_j - \bar{x}) \end{pmatrix} \left(-e^{-r(t_j-t_0)} \hat{p}_{x_j,t_j} \hat{q}_{x_j,t_j}\right) \\ &= \begin{pmatrix} \frac{\partial \hat{H}_j}{\partial \kappa_1(t_0)} & \frac{\partial \hat{H}_j}{\partial \kappa_2(t_0)} & \frac{\partial \hat{H}_j}{\partial C_1(t_0)} & \frac{\partial \hat{H}_j}{\partial C_2(t_0)} \end{pmatrix}' = \frac{\partial \hat{H}_j}{\partial \vec{\alpha}_{t_0}}, \end{aligned}$$

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

for $j = 1, 2, 3, 4$. Hence, the left-hand side of Equation (5.12) can be reduced to

$$\begin{pmatrix} \left(\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}}\right)' A^{t_1-t_0} \\ \vdots \\ \left(\frac{\partial H_4}{\partial \vec{\alpha}_{t_4}}\right)' A^{t_4-t_0} \end{pmatrix} P \left(\left(\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}}\right)' A^{t_1-t_0} \quad \cdots \quad \left(\frac{\partial H_4}{\partial \vec{\alpha}_{t_4}}\right)' A^{t_4-t_0} \right)$$
$$= \begin{pmatrix} \left(\frac{\partial \hat{H}_1}{\partial \vec{\alpha}_{t_0}}\right)' \\ \vdots \\ \left(\frac{\partial \hat{H}_4}{\partial \vec{\alpha}_{t_0}}\right)' \end{pmatrix} P \left(\left(\frac{\partial \hat{H}_1}{\partial \vec{\alpha}_{t_0}}\right)' \quad \cdots \quad \left(\frac{\partial \hat{H}_4}{\partial \vec{\alpha}_{t_0}}\right)' \right).$$

Furthermore, from Equation (5.4), we obtain

$$\sum_{i=t_0+1}^{t_0+T} (i - t_0) \frac{\partial L}{\partial \vec{\alpha}_i^*}$$
$$= \sum_{i=t_0+1}^{t_0+T} (i - t_0) M_{x_0,i} \hat{q}_{x_0+i-t_0-1,i} \left(- \sum_{u=i-t_0}^{T} e^{-ru} \prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0+t-t_0-1,t} \right)$$
$$= \sum_{u=1}^{T} e^{-ru} (-1) \prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0+t-t_0-1,t} \left(\sum_{t=t_0+1}^{t_0+u} (t - t_0) M_{x_0,t} \hat{q}_{x_0+t-t_0-1,t} \right),$$

which gives

$$\sum_{i=t_0+1}^{t_0+T} (A^{i-t_0})' \frac{\partial L}{\partial \vec{\alpha}_i} = \frac{\partial \hat{L}}{\partial \vec{\alpha}_{t_0}}.$$

Therefore, the right-hand side of Equation (5.12) can be reduced to

$$\begin{pmatrix} \left(\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}}\right)' A^{t_1-t_0} \\ \vdots \\ \left(\frac{\partial H_4}{\partial \vec{\alpha}_{t_4}}\right)' A^{t_4-t_0} \end{pmatrix} P \sum_{j=t_0+1}^{t_0+T} (A^{j-t_0})' \frac{\partial L}{\partial \vec{\alpha}_j} = \begin{pmatrix} \left(\frac{\partial \hat{H}_1}{\partial \vec{\alpha}_{t_0}}\right)' \\ \vdots \\ \left(\frac{\partial \hat{H}_4}{\partial \vec{\alpha}_{t_0}}\right)' \end{pmatrix} P \frac{\partial \hat{L}}{\partial \vec{\alpha}_{t_0}}.$$

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

Finally, we can rewrite Equation (5.12) as

$$\begin{pmatrix} \left(\frac{\partial \hat{H}_1}{\partial \vec{\alpha}_{t_0}}\right)' \\ \vdots \\ \left(\frac{\partial \hat{H}_4}{\partial \vec{\alpha}_{t_0}}\right)' \end{pmatrix} P \begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \vec{\alpha}_{t_0}} & \cdots & \frac{\partial \hat{H}_4}{\partial \vec{\alpha}_{t_0}} \end{pmatrix} \begin{pmatrix} N_1 \\ \vdots \\ N_4 \end{pmatrix} = \begin{pmatrix} \left(\frac{\partial \hat{H}_1}{\partial \vec{\alpha}_{t_0}}\right)' \\ \vdots \\ \left(\frac{\partial \hat{H}_4}{\partial \vec{\alpha}_{t_0}}\right)' \end{pmatrix} P \frac{\partial \hat{L}}{\partial \vec{\alpha}_{t_0}},$$

or equivalently

$$\begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \cdots & \frac{\partial \hat{H}_1}{\partial C_2(t_0)} \\ \vdots & \ddots & \vdots \\ \frac{\partial \hat{H}_4}{\partial \kappa_1(t_0)} & \cdots & \frac{\partial \hat{H}_4}{\partial C_2(t_0)} \end{pmatrix} P \begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \cdots & \frac{\partial \hat{H}_4}{\partial \kappa_1(t_0)} \\ \vdots & \ddots & \vdots \\ \frac{\partial \hat{H}_1}{\partial C_2(t_0)} & \cdots & \frac{\partial \hat{H}_4}{\partial C_2(t_0)} \end{pmatrix} \begin{pmatrix} N_1 \\ \vdots \\ N_4 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \cdots & \frac{\partial \hat{H}_1}{\partial C_2(t_0)} \\ \vdots & \ddots & \vdots \\ \frac{\partial \hat{H}_4}{\partial \kappa_1(t_0)} & \cdots & \frac{\partial \hat{H}_4}{\partial C_2(t_0)} \end{pmatrix} P \begin{pmatrix} \frac{\partial \hat{L}}{\partial \kappa_1(t_0)} \\ \vdots \\ \frac{\partial \hat{L}}{\partial C_2(t_0)} \end{pmatrix}.$$

By definition, Q is positive definite and hence $P = A^{-1}Q(A^{-1})'$ is invertible. If the other matrices in the equation above are invertible, then we immediately obtain

$$\begin{pmatrix} N_1 \\ \vdots \\ N_4 \end{pmatrix} = \begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \cdots & \frac{\partial \hat{H}_4}{\partial \kappa_1(t_0)} \\ \vdots & \ddots & \vdots \\ \frac{\partial \hat{H}_1}{\partial C_2(t_0)} & \cdots & \frac{\partial \hat{H}_4}{\partial C_2(t_0)} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial \hat{L}}{\partial \kappa_1(t_0)} \\ \vdots \\ \frac{\partial \hat{L}}{\partial C_2(t_0)} \end{pmatrix},$$

which yields N_1, N_2, N_3 and N_4 that are exactly the same as those implied by the traditional delta–nuga hedging method. The hedging strategy is free of Q^* and therefore incorporates no information concerning the variances and covariances of the hidden states.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

5.4.3. *Distinctions between the proposed method and the traditional methods.* Although the delta and delta–nuga methods can be viewed as special cases of the generalized state-space method, they are fundamentally different from the generalized state-space method in several ways.

First, under the generalized state-space method, the hedging strategy is derived by optimizing a specific objective function: $\mathrm{Var}(L - \sum_{j=1}^{m} N_j H_j)$. In contrast, the delta and delta–nuga methods rely merely on sensitivity matching and involves no optimization.

Second, through $\Xi^*_{s,t}$, the generalized state-space method incorporates both static and dynamic correlations between different hidden states. Such correlations are not taken into account in the delta and delta–nuga methods.

Third, instead of assuming the linear relation specified in Equation (5.8), the generalized state-space method recognizes that L and H_j are functions of $\vec{\alpha}^*_{t_0+1}, \ldots, \vec{\alpha}^*_{t_0+T}$ and $\vec{\alpha}^*_{t_j}$, respectively. Even if the same model is assumed, the generalized state-space method would still be different (in terms of both the derivation and the resulting notional amounts) from the delta and delta–nuga methods.

Finally, in the generalized state-space method, the drifts themselves are state variables. However, in the delta–nuga hedging method, the drift term in the time-varying dynamic factor is regarded as a constant parameter which is recalibrated from time to time.

5.5. Comments on the hedging methods

5.5.1. *Sub-optimality of the traditional methods.* In the previous sub-section, we have shown that the traditional delta and delta–nuga hedging methods are special cases of the generalized state-space hedging method when the linear relations specified in Equation (5.8) hold. Equivalently speaking, if the linear relations do not hold, then the notional amounts $N_1, \ldots, N_m$ computed using delta or delta–nuga hedging methods do not minimize

$$\mathrm{Var}\left(L - \sum_{j=1}^{m} N_j H_j\right),$$

the variance of the present values of all cash flows associated with the hedged position. Therefore, in general, the hedging strategies derived from the traditional methods are sub-optimal relative to those derived from the generalized state-space approach. The degree of sub-optimality is quantified in Section 6 where a numerical illustration is presented.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

5.5.2. *The singularity problem.* Recall that in the traditional delta hedging method, the solution to N_1 and N_2 exists only if the following matrix is invertible:

$$\begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \frac{\partial \hat{H}_2}{\partial \kappa_1(t_0)} \\ \frac{\partial \hat{H}_1}{\partial \kappa_2(t_0)} & \frac{\partial \hat{H}_2}{\partial \kappa_2(t_0)} \end{pmatrix} = \begin{pmatrix} -e^{-r(t_1-t_0)}\hat{p}_{x_1,t_1}\hat{q}_{x_1,t_1} & -e^{-r(t_2-t_0)}\hat{p}_{x_2,t_2}\hat{q}_{x_2,t_2} \\ -e^{-r(t_1-t_0)}(x_1-\bar{x})\hat{p}_{x_1,t_1}\hat{q}_{x_1,t_1} & -e^{-r(t_2-t_0)}(x_2-\bar{x})\hat{p}_{x_2,t_2}\hat{q}_{x_2,t_2} \end{pmatrix}.$$

If the references ages of both q-forwards are the same (i.e., $x_1 = x_2$), then the second row of the matrix would become perfectly linearly dependent on the first row and hence the square matrix is not invertible. Note that this problem exists even though the reference years t_1 and t_2 of the q-forwards are different.

The same problem also applies to the traditional delta–nuga hedging method, under which the solution to $N_1, \ldots, N_4$ exists only if the square matrix in Equation (5.2) is invertible. The first three rows of the square matrix in Equation (5.2) are respectively given by

$$\begin{aligned} \text{Row 1:} \quad & \frac{\partial \hat{H}_j}{\partial \kappa_1(t_0)} = -e^{-r(t_j-t_0)}\hat{p}_{x_j,t_j}\hat{q}_{x_j,t_j}, && j = 1,2,3,4; \\ \text{Row 2:} \quad & \frac{\partial \hat{H}_j}{\partial \kappa_2(t_0)} = -e^{-r(t_j-t_0)}(x_j-\bar{x})\hat{p}_{x_j,t_j}\hat{q}_{x_j,t_j}, && j = 1,2,3,4; \\ \text{Row 3:} \quad & \frac{\partial \hat{H}_j}{\partial C_1(t_0)} = -e^{-r(t_j-t_0)}(t_j-t_0)\hat{p}_{x_j,t_j}\hat{q}_{x_j,t_j}, && j = 1,2,3,4. \end{aligned}$$

It is clear that when the references ages of the q-forwards are the same (i.e., $x_1 = x_2 = x_3 = x_4$), the first and second rows in the matrix are perfectly linearly dependent on each other, which means the inverse of the matrix does not exist. On top of that, the problem of singularity will also occur if the q-forwards are linked to the same cohort; that is, $t_j - x_j = c$ for $j = 1, \ldots, 4$, where c denotes the year of birth to which the q-forwards are linked. This is because in this case, the difference between rows 3 and 2 becomes

$$-e^{-r(t_j-t_0)}(c-t_0+\bar{x})\hat{p}_{x_j,t_j}\hat{q}_{x_j,t_j}, \quad j = 1,2,3,4,$$

which is perfectly linearly dependent on row 1. This fact suggests that although it seems natural to choose q-forwards that are linked to the same cohort (as the one associated with the annuity liability), such a choice is not desirable if the delta–nuga hedging method is used.

The singularity problem does not apply to the generalized state-space-hedging methodology, provided that the state vectors beyond year t_0 are not deterministically related as specified in Equation (5.8). To explain why, let us

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

have a closer scrutiny of the square matrix in Equation (5.6). Because

$$\mathrm{Cov}(h_i, h_j) = \left(\frac{\partial H_i}{\partial \vec{\alpha}^*_{t_i}}\right)' \Xi^*_{t_i,t_j} \frac{\partial H_j}{\partial \vec{\alpha}^*_{t_j}}$$

for $i, j = 1, 2, \ldots, m$, the square matrix

$$\Sigma_h := \begin{pmatrix} \left(\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}}\right)' \Xi^*_{t_1,t_1} \frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}} & \cdots & \left(\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}}\right)' \Xi^*_{t_1,t_m} \frac{\partial H_m}{\partial \vec{\alpha}^*_{t_m}} \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial H_m}{\partial \vec{\alpha}^*_{t_m}}\right)' \Xi^*_{t_m,t_1} \frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}} & \cdots & \left(\frac{\partial H_m}{\partial \vec{\alpha}^*_{t_m}}\right)' \Xi^*_{t_m,t_m} \frac{\partial H_m}{\partial \vec{\alpha}^*_{t_m}} \end{pmatrix}$$

can be viewed as the covariance matrix of $\vec{h} = (h_1, \ldots, h_m)'$. By the spectral theorem, the square matrix can be written as

$$\Sigma_h = U \Lambda U',$$

where U is an orthogonal matrix and Λ is a diagonal matrix containing the eigenvalues of Σ_h. It immediately follows that

$$U' \Sigma_h U = \Lambda.$$

We can view $U' \Sigma_h U$ as the covariance matrix of the random vector $U' \vec{h}$, of which each element is a linear combination of the elements in $\vec{h}$. From Equation (5.3) and the fact that $t_1 < \ldots < t_m$, we can infer that there does not exist a linear combination of $h_1, \ldots, h_m$ such that the linear combination is non-random. Therefore, all diagonal elements in Λ must be straightly positive. With straightly positive eigenvalues, the invertibility of Σ_h is guaranteed.

5.5.3. *The hedging instrument selection problem.* In using the traditional delta- or delta–nuga hedging method, the number of hedging instruments must be either 2 (for delta) or 4 (for delta–nuga). In contrast, the generalized state-space-hedging method can be implemented with any number of hedging instruments, as the solution to Equation (5.6) exists for any $m \geq 1$. In this sense, the generalized state-space hedging method may be considered as more adaptable to different stages of market development.

As the market grows, more hedging instruments will become available. For reasons such as transaction costs, a hedger may wish to use only a subset of (rather than all) instruments that are available in the market. One possible criterion for choosing the subset of instruments is the resulting variance of the present values of all cash flows involved. To formulate this method mathematically, we first define a subset $\mathbb{M}$ which contains m element(s) out of a set of k elements:

$$\mathbb{M} = \{(a_1, a_2, \ldots, a_m) : a_i \in \{1, 2, \ldots, k\}, a_1 < a_2 < \ldots < a_m\},$$

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

where m and k represent the number of instruments that the hedger wishes to use and the total number of instruments available in the market, respectively. The collection of m instrument(s) selected can written as

$$\arg\min_{S_m \in \mathbb{M}} \left\{ \text{Var} \left(\sum_{\substack{i=t_0+1 \\ i \neq t_j, j \in S_m}}^{t_0+T} \left(\frac{\partial L}{\partial \vec{\alpha}_i^*} \right)' (\vec{\alpha}_i^* - \hat{\vec{\alpha}}_i^*) + \sum_{j \in S_m} \left(\frac{\partial L}{\partial \vec{\alpha}_{t_j}^*} - N_j \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*} \right)' (\vec{\alpha}_{t_j}^* - \hat{\vec{\alpha}}_{t_j}^*) \right) \right\}.$$

6. Illustrating the hedging methods

6.1. Assumptions

In this illustration, the liability being hedged is a 30-year life annuity immediate (i.e., $T = 30$). The annuity is sold to a person aged $x_0 = 70$ at the end of year $t_0 = 2010$, and a longevity hedge is established at the same time as the annuity is sold. We assume that the annuitant is subject to exactly the same mortality as Canadian males.

The hedging instruments under consideration are q-forwards. It is assumed that at the time when the hedge is established, there are exactly $k = 4$ q-forwards available in the market. The reference ages and years of these q-forwards are summarized below:

j	Reference Age x_j	Reference Year t_j
1	76	2017
2	85	2026
3	92	2033
4	100	2040

The first three q-forwards are associated with the same cohort (with year-of-birth 1941), whereas the last q-forward is associated with the cohort born one year earlier. We intentionally avoid having all four q-forwards linked to the same cohort, because otherwise the traditional delta–nuga hedging method would fail due to the singularity problem. Note that the maturities of the q-forwards are approximately $T/4$, $T/2$, $3T/4$ and T, respectively. We assume that the q-forwards are also linked to the mortality of Canadian males.

The procedure which we use to assess hedge effectiveness can be summarized as follows:

i. Preparation
   - Compute the partial derivatives involved.
   - Assuming that the hedger wishes to use m q-forwards, select m out of the $k = 4$ available q-forwards on the basis of variance minimization.
   - Calculate the optimal notional amounts.

ii. Simulation
   - Simulate 10,000 mortality scenarios from the CBD/LLCBD model that is fitted to the data for Canadian males over a calibration window of 1941–2010 and an age range of 50–89.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

- For each simulated mortality scenario, calculate the realized values of L and H_j, $j = 1, \ldots, m$. An interest rate of $r = 0.01$ is used for discounting purposes.

iii. Evaluation

- On the basis of the 10,000 realizations of L and H_j, $j = 1, \ldots, m$, compute the value HE.

The objectives of this illustration are threefold: (1) to compare the hedge effectiveness produced by the traditional and newly proposed methods; (2) to assess how the negligence of stochastic drifts may affect the resulting hedge effectiveness; (3) to demonstrate the interaction among the model assumption, the hedging method and the duration of the liability being hedged. To better achieve these objectives, we divide the empirical results into the following four groups, depending on the hedging method and mortality model from the hedging strategy is derived:

	The mortality model on which the hedging strategy is based	Hedging method
Group 1	The original CBD model	Traditional delta hedging
Group 2	The LLCBD model	Traditional delta and delta–nuga hedging
Group 3	The original CBD model	Generalized state-space hedging
Group 4	The LLCBD model	Generalized state-space hedging

6.2. Result I: A comparison of different hedging methods

In this sub-section, we focus on comparing the hedge effectiveness produced by different hedging methods. The relevant results are displayed in Tables 9 and 10.

Let us first study Table 9, which compares the results in Groups 1 and 3. For these two groups, the hedging strategies are based on the same model (the original CBD model) but are derived using different hedging techniques. Here, we intend to focus on the difference in hedging techniques, so we assume that the actual (simulation) mortality model is also the original CBD model. We report the results when $m = 2$ hedging instruments are used, because the results in Group 1 are derived from the traditional delta hedging method which requires exactly two hedging instruments. For Group 1, the best hedge effectiveness obtained is greater than 92%, but different combinations of q-forwards lead to highly different values of HE. In the worst case when the third and fourth q-forwards are used, the value of HE is even negative, which means the seller of the annuity is subject to even more longevity risk when the hedge is in place. For Group 3, the values of HE are much less sensitive to the choice of the two q-forwards. Also, given the same choice of q-forwards, the hedge effectiveness in Group 3

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 9

THE HEDGE EFFECTIVENESS AND NOTIONAL AMOUNTS FOR ALL POSSIBLE COMBINATIONS OF $m = 2$ q-FORWARDS, GROUPS 1 AND 3. THE "–" SIGN INDICATES THAT THE CORRESPONDING q-FORWARD IS NOT USED. THE SIMULATION MODEL IS THE ORIGINAL CBD MODEL.

Group 1 Assumed Model: The Original CBD Model Method: Traditional Delta Hedging					Group 3 Assumed Model: The Original CBD Model Method: Generalized State-Space Hedging				
HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
0.9230	98.0364	47.4528	–	–	0.9358	71.5625	47.9503	–	–
0.8997	137.3202	–	17.0804	–	0.9135	107.8487	–	18.1090	–
0.8183	154.1561	–	–	7.5688	0.8322	132.5633	–	–	7.0199
−0.3565	–	165.8758	−42.6257	–	0.9053	–	44.2286	8.4076	–
0.0177	–	130.3489	–	−13.2221	0.8924	–	51.7275	–	2.4632
−5.2683	–	–	156.3945	−61.7343	0.7463	–	–	24.6273	−0.2105

TABLE 10

THE HEDGE EFFECTIVENESS AND NOTIONAL AMOUNTS FOR ALL POSSIBLE COMBINATIONS OF $m = 2$ AND $m = 4$ q-FORWARDS, GROUPS 2 AND 4. THE "–" SIGN INDICATES THAT THE CORRESPONDING q-FORWARD IS NOT USED. THE SIMULATION MODEL IS THE LLCBD MODEL.

Group 2 Assumed Model: The LLCBD Model Method: Traditional Delta/Delta–Nuga Hedging					Group 4 Assumed Model: The LLCBD Model Method: Generalized State-Space Hedging				
HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
Number of q-Forwards Used: $m = 2$									
0.7209	110.4838	68.2613	–	–	0.8409	−25.5044	76.0989	–	–
0.8701	158.4984	–	27.8511	–	0.8957	97.7825	–	24.9453	–
0.1140	179.0761	–	–	13.2179	0.6569	92.8728	–	–	7.4087
−3.5174	–	225.3339	−64.0867	–	0.9579	–	41.0866	13.4239	–
−5.7736	–	178.2121	–	−21.2905	0.9386	–	56.9217	–	3.3286
−32.7169	–	–	242.3726	−101.8099	0.8764	–	–	30.2868	−2.9065
Number of q-Forwards Used: $m = 4$									
0.8754	157.7860	1.0128	27.4379	−0.0000	0.9737	66.7273	38.9987	13.2875	0.8685

is always higher than that in Group 1, indicating that hedgers can make better use of the hedging instruments if they use the generalized state-space hedging method. This advantage may be attributed to the fact that the generalized state-space-hedging method incorporates information about the static and dynamic correlations between the hidden states, but the delta hedging method does not.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

We then move on to studying Table 10, which compares the results from Groups 2 and 4. As before, the hedging strategies for these two groups are based on the same model (the LLCBD model) but different hedging techniques. For Group 4, all results are derived from the generalized state-space hedging method, whereas for Group 2, the results for $m = 2$ and $m = 4$ are respectively obtained by using the traditional delta and delta–nuga hedging methods. The simulation model used here is the LLCBD model, consistent with the model on which the hedging strategies are based. Compared to the values of HE in Group 2, the values of HE in Group 4 are consistently higher and are more robust relative to the choice of hedging instruments. It is also noteworthy that some values of HE in Group 2 are negative, as the traditional hedging methods do not guarantee a reduction in variance. This problem is discussed further in Section 6.5.1.

The results presented in this sub-section demonstrate the previously made argument concerning the sub-optimality of the traditional hedging methods. Other things equal, the generalized state-space hedging method yields a better hedge effectiveness, regardless of whether the original CBD model or the LLCBD model is assumed.

In this sub-section, the values of HE for Groups 1 and 3 are calculated using scenarios generated from the original CBD model, while those for Groups 2 and 4 are computed using scenarios simulated from the LLCBD model. For the readers' information, the estimated values of $\mathrm{Var}(L)$ under the CBD and LLCBD models are 0.0913 and 0.0715, respectively. The reason that $\mathrm{Var}(L)$ calculated using the CBD scenarios is larger can be understood by revisiting Figure 6, which shows that the LLCBD fan charts are initially narrower than the corresponding CBD fan charts, but become wider about 20 years from the forecast origin as they widen at faster rates. The reason behind can also be visualized in Figure 14, which compares the standard deviations of the cohort death probabilities of the annuitants (in logit scale) under the two models. The standard deviations under the CBD model are higher for the first half of the annuity's maximum duration, but the opposite is true for the second half. Because cash flows in distant future are less influential (due to the effects of discounting and survivorship), $\mathrm{Var}(L)$ depends more heavily on the uncertainty surrounding the earlier cohort death probabilities and is consequently higher under the CBD model.

6.3. Result II: The impact of model mis-specification

We have previously argued that compared to the original CBD model, the LLCBD model (with stochastic drifts) may more realistically represent the true underlying mortality dynamics. In this sub-section, we examine how much hedge effectiveness may be lost if the true underlying mortality dynamics are driven by a model with stochastic drifts while a model with constant drifts is assumed in deriving the hedging strategies. To achieve this goal, we compare the results in

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

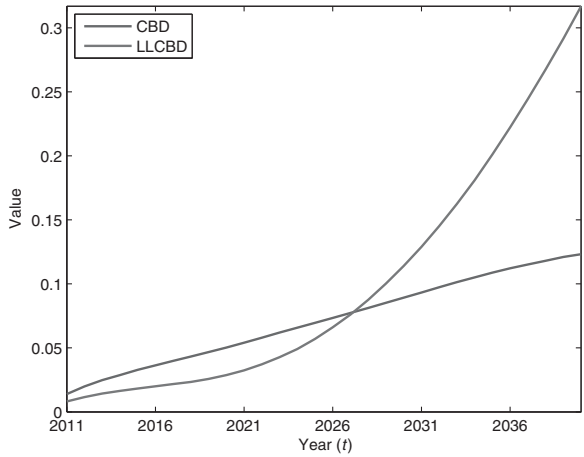

FIGURE 14: The standard deviations of the annuitants' cohort death probabilities in logit scale (i.e., $\ln(q_{t-1941,t}/(1-q_{t-1941,t}))$ for $t = 2011, \ldots, 2041$), estimated using the CBD and LLCBD models.

TABLE 11

THE HEDGE EFFECTIVENESS AND THE CORRESPONDING NOTIONAL AMOUNTS WHEN $m = 1, 2, 3, 4$ q-FORWARDS ARE USED, GROUPS 3 AND 4. THE "–" SIGN INDICATES THAT THE CORRESPONDING q-FORWARD IS NOT USED. THE SIMULATION MODEL IS THE LLCBD MODEL.

Group 3 Assumed Model: The Original CBD Model Method: Generalized State-Space Hedging					Group 4 Assumed Model: The LLCBD Model Method: Generalized State-Space Hedging				
HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
Number of q-Forwards Used: $m = 1$									
0.8000	–	60.0893	–	–	0.8575	–	–	22.3178	–
Number of q-Forwards Used: $m = 2$									
0.6496	71.5625	47.9503	–	–	0.9579	–	41.0866	13.4239	–
Number of q-Forwards Used: $m = 3$									
0.8557	75.7070	29.6824	9.3110	–	0.9732	64.1560	36.5098	16.1386	–
Number of q-Forwards Used: $m = 4$									
0.8728	76.1325	29.7784	7.9015	0.7337	0.9737	66.7273	38.9987	13.2875	0.8685

Groups 3 and 4 (see Table 11) when the true (simulation) model is the LLCBD model.

All results in Groups 3 and 4 are based on the generalized state-space-hedging method and mortality scenarios simulated from the LLCBD model. The differences between the results in these two groups are because of the

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

difference in the model from which the notional amounts are derived. Since the generalized state-space hedging method permits us to use any number of hedging instruments, we consider both the situation when all available q-forwards are used and the situation when only a subset of the available q-forwards are included. When less than four q-forwards are used, the q-forwards are chosen using the method described in Section 5.5.3. For Group 3, HE does not necessarily increase with m, because the model on which the hedging strategies are based is inconsistent with the simulation model.

Compared to the corresponding values in Group 3, the values of HE in Group 4 are consistently higher. This observation suggests that the negligence of stochastic drifts may result in a material loss of hedge effectiveness, if the true underlying model is one with stochastic drifts.

6.4. Result III: The interaction among different factors

In this sub-section, we demonstrate the interaction among the model assumption, the hedging method and the duration of the liability being hedged. To achieve this goal, we now consider annuity liabilities with durations ranging from $T = 25$ to $T = 40$ years. For each annuity liability, the longevity hedge is composed of $m = 2$ q-forwards, which are linked to the same cohort of individuals as the annuity liability and have maturities of (approximately) $T/2$ and $T/4$. All other previously made assumptions, including the assumptions about the values of x_0, t_0 and r, remain unchanged. The simulation model used is still the LLCBD model. The results are presented in Figure 15.

We first compare the trends of HE for Groups 2 and 4. As expected, the trend for Group 4, which is based on the generalized state-space hedging method, is always higher than that for Group 2, which is based on the traditional delta-hedging method. A more interesting observation is that the gap between the two trends is roughly constant. This observation indicates that the benefit from using the more general hedging method is somewhat fixed, with little dependence on the duration of the liability being hedged.

Next, we compare the trends of HE for Groups 3 and 4. The gap between the two trends widens rapidly as the duration of the liability being hedged increases. From this observation, we can infer that when the true model is one with stochastic drifts, the benefit of incorporating stochastic drifts into the hedging strategy is more remarkable if the liability being hedged is longer dated. This conclusion is reasonable because the assumption about the drifts, which determine the gradients of future mortality trends, should have more long-run than short-run effects.

Finally, we compare the trends of HE for Groups 1 and 4. The gap between the HE trends for these two groups is approximately the sum of the gap between the HE trends for Groups 2 and 4 and the gap between the HE trends for Groups 3 and 4. It reflects the overall benefit of using both the generalized state-space hedging method and the assumption of stochastic drifts in deriving the hedging strategies.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

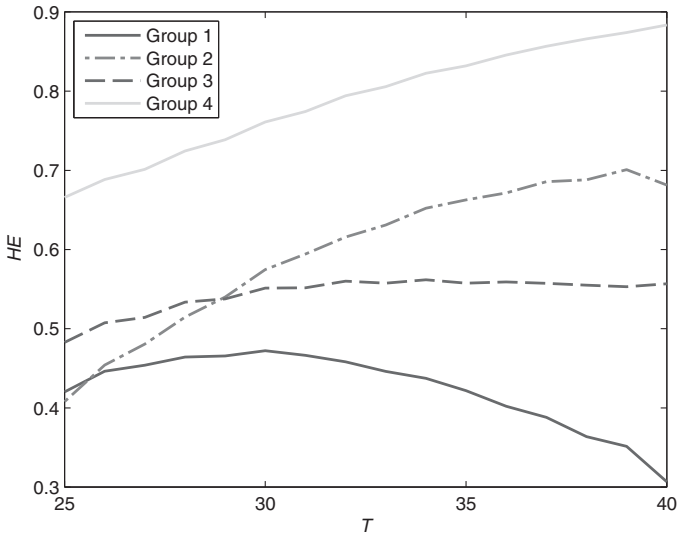

FIGURE 15: The relationship between the hedge effectiveness (HE) and the duration (T) of the liability being hedged, Groups 1, 2, 3 and 4. The simulation model is the LLCBD model.

6.5. Further issues

6.5.1. *Why delta and delta–nuga methods may perform unsatisfactorily sometimes?* In Section 6.2, we reveal that the delta and delta–nuga hedging methods may sometimes lead to a low or even negative hedge effectiveness. We may attribute this problem to the linearity assumption on which these methods are based.

To illustrate, we construct a scenario analysis that is based on the following assumptions:

- Mortality model on which the hedging strategy is based: the original CBD model.
- Simulation model: the original CBD model.
- Hedging instruments: the second and third q-forward, $(x_2, t_2) = (85, 2026)$ and $(x_3, t_3) = (92, 2033)$.

Under these assumptions, the values of HE produced by the delta hedging method and the generalized state-space hedging method are -0.3565 and 0.9053, respectively (see Table 9).

Six mortality scenarios — built on hypothetical sample paths of $\kappa_1(t)$ and $\kappa_2(t)$ — are used to analyze why the two methods perform so differently (see Figure 16). Scenarios (i) and (ii) are the most extreme "linear" scenarios within the 95% prediction intervals, while Scenarios (iii) to (vi) are extreme "nonlinear" scenarios within the 95% prediction intervals.[6] For each scenario and hedging method, we calculate the realized values of $L - N_2 H_2 - N_3 H_3 - \hat{L}$ (the

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

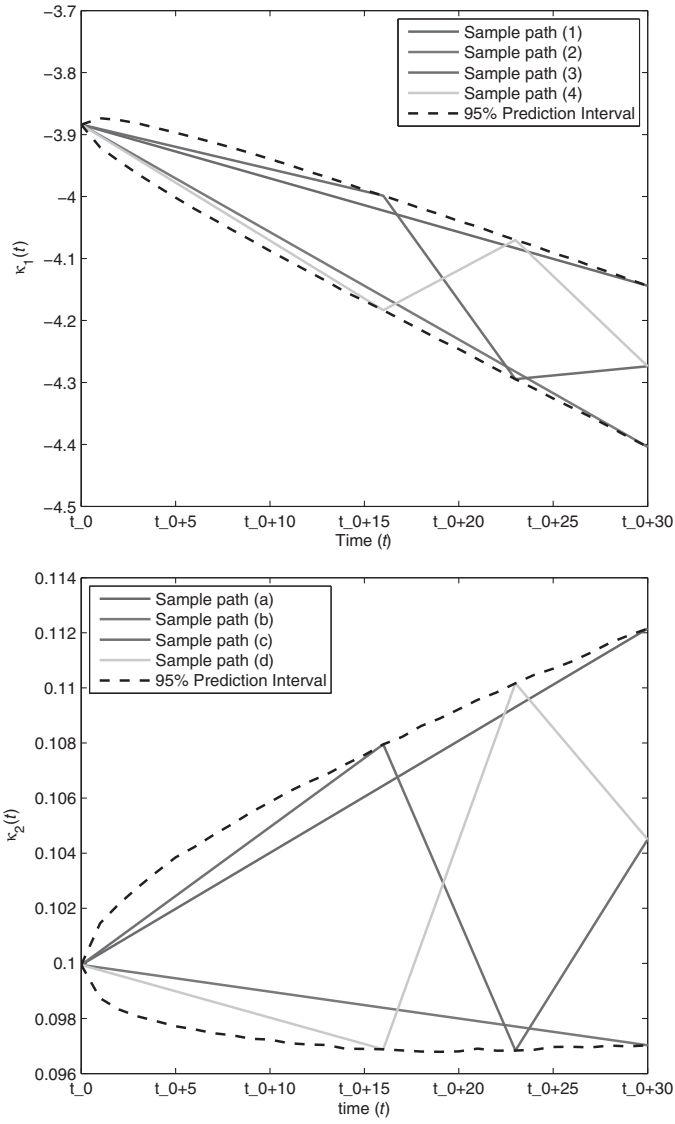

FIGURE 16: Six extreme mortality scenarios: Scenarios (i) to (vi) are formed by (1) × (a), (2) × (b), (3) × (c), (3) × (d), (4) × (c), (4) × (d), respectively. The dotted lines represent the 95% prediction intervals.

hedged position) and $L-\hat{L}$ (the unhedged position). The results are tabulated in Table 12.

For Scenarios (i) and (ii), the hedge derived using the delta hedging method yields values of $L - N_2H_2 - N_3H_3 - \hat{L}$ that are much smaller than $L - \hat{L}$ in magnitude, suggesting that the hedge can withstand extreme mortality scenarios that are "linear". However, for remaining four scenarios, all of which are "non-

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 12

THE VALUES OF $L - N_2H_2 - N_3H_3 - \hat{L}$ AND $L - \hat{L}$ UNDER THE SIX HYPOTHETICAL EXTREME MORTALITY SCENARIOS DESCRIBED IN FIGURE 16.

Scenario	Sample Paths	N_2H_2	N_3H_3	$L - N_2H_2 - N_3H_3 - \hat{L}$	$L - \hat{L}$
			Delta Hedge		
(i)	(1) × (a)	−1.3804	0.9932	−0.1379	−0.5251
(ii)	(2) × (b)	1.2271	−0.8361	0.1658	0.5568
(iii)	(3) × (c)	−1.9107	−0.9401	2.5764	−0.2743
(iv)	(3) × (d)	−0.0769	0.1562	−0.1848	−0.1055
(v)	(4) × (c)	0.0511	−0.1431	0.1883	0.0963
(vi)	(4) × (d)	1.6191	1.1475	−2.5397	0.2270
			Generalized State-Space Hedge		
(i)	(1) × (a)	−0.3681	−0.1959	0.0389	−0.5251
(ii)	(2) × (b)	0.3272	0.1649	0.0647	0.5568
(iii)	(3) × (c)	−0.5095	0.1854	0.0497	−0.2743
(iv)	(3) × (d)	−0.0205	−0.0308	−0.0542	−0.1055
(v)	(4) × (c)	0.0136	0.0282	0.0545	0.0963
(vi)	(4) × (d)	0.4317	−0.2263	0.0216	0.2270

linear", the hedge derived using the delta hedging method performed very badly (even worse than not hedging), indicating that the hedge is vulnerable to nonlinear changes in the dynamic factors. We also observe that $L - N_2H_2 - N_3H_3 - \hat{L}$ and $L - \hat{L}$ may take the same or different signs, which means that the poor performance is sometimes because the hedge has gone in the wrong direction and sometimes because the hedge has gone in the right direction but too far.

For all six scenarios, the generalized state-space hedging method performs satisfactorily, yielding values of $L - N_2H_2 - N_3H_3 - \hat{L}$ that are much smaller than $L - \hat{L}$ in magnitude.

6.5.2. *The best achievable hedge effectiveness.* Given the 10,000 mortality scenarios simulated from the CBD/LLCBD model, we can iteratively search for the combination of $N_1, \ldots, N_m$ what would maximize HE. The resulting value of HE may be regarded as the "best achievable" hedge effectiveness (given the 10,000 scenarios) and can be used as a benchmark to assess the quality of the proposed hedging strategy.

We consider hedge portfolios with $m = 2, 4$ q-forwards, chosen from the collection of four q-forwards described in Section 6.1. Table 13 displays, for each portfolio, the best achievable value of HE and the corresponding values of $N_1, \ldots, N_m$ on the basis of the mortality scenarios simulated from (a) the original CBD model and (b) the LLCBD model. The values reported in Table 13 (a) are very close to those shown in Table 9 (Group 3), whereas the

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 13

THE "BEST ACHIEVABLE" HEDGES GIVEN THE 10,000 MORTALITY SCENARIOS SIMULATED FROM (A) THE ORIGINAL CBD MODEL AND (B) THE LLCBD MODEL.

HE	N_1	N_2	N_3	N_4
(a) Simulation Model: The Original CBD Model				
0.9358	−70.9784	−47.8004	–	–
0.9136	−107.4581	–	−17.8760	–
0.8322	−132.2430	–	–	−6.9878
0.9054	–	−44.4610	−8.0472	–
0.8926	–	−52.0698	–	−2.2432
0.7465	–	–	−24.3666	0.2981
(b) Simulation Model: The LLCBD Model				
0.8410	−22.8565	75.3767	–	–
0.8967	93.0628	–	24.1189	–
0.6609	79.1050	–	–	6.8454
0.9586	–	41.4408	12.7478	–
0.9400	–	57.0345	–	3.0335
0.8769	–	–	29.1619	−2.6613
0.9749	62.2898	39.3229	13.0727	0.6231

values reported in Table 13 (b) are very close to those shown in Table 10 (Group 4). The results suggest that the performance of the generalized state-space hedging strategy is nearly as good as the best achievable one.

6.5.3. *Poisson risk.* The baseline results assume no Poisson risk (a.k.a. sampling risk). We now incorporate Poisson uncertainty into the hedging results by treating the cohort of annuitants as a random survivorship group.

Let $l_{x,t}$ be the number of annuitants who survive to age x at the beginning of year t, and $d_{x,t}$ be the number of annuitants who die in year t (between age x and $x+1$). For a fixed initial number of annuitants l_{x_0,t_0+1}, we have the following:

$$d_{x_0+t-t_0-1,t} \sim \text{Poisson}(l_{x_0+t-t_0-1,t} \times \tilde{q}_{x_0+t-t_0-1,t})$$

and

$$l_{x_0+t-t_0,t+1} = l_{x_0+t-t_0-1,t} - d_{x_0+t-t_0-1,t},$$

where $t = t_0+1, \ldots, t_0+T$ and $\tilde{q}_{x,t}$ is the underlying unobserved probability of death in year t (between age x and $x+1$).

In generating the hedging results, we simulate, for each of the 10,000 simulated mortality scenarios from the LLCBD model, one realization of $\{l_{x_0+t-t_0+1,t+1};\, t = t_0, t_0+1, \ldots, t_0+T-1\}$. The per contract value of the liability

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

being hedged in each mortality scenario is then computed as follows:

$$L = \sum_{u=1}^{T} e^{-ru} \frac{l_{x_0+u,t_0+1+u}}{l_{x_0,t_0+1}}.$$

The calculation of $N_1, \ldots, N_m$ remains the same as when Poisson uncertainty is not taken into account, because the q-forward hedge aims to mitigate only systematic longevity risk. There is also no change to the calculation of H_j, $j = 1, \ldots, m$, because the q-forwards are assumed to be linked to the smoothed death probabilities that are free of Poisson uncertainty.

We calculate the hedge effectiveness for annuity liabilities with different initial numbers of annuitants: $l_{x_0,t_0+1} = 10^3, 10^4, 10^5, 10^6, 10^7$. The results (see Table 14) are based on hedge portfolios with $m = 1, 2, 3, 4$ q-forwards, built by applying the generalized state-space method to the original CBD model (Group 3) and the LLCBD model (Group 4). When there are only 1,000 annuitants initially, Poisson risk erodes hedge effectiveness by about 30 to 40 percentage points. The erosion in hedge effectiveness is much milder (less than 10 percentage points) when the initial number of annuitants is 10,000, and becomes negligible when the initial number of annuitant grows to 100,000. When the initial number of annuitants is 10^7, the resulting values of HE are identical (to four significant figures) to those when Poisson risk is assumed to be absent. Our results are in line with those produced by Li and Hardy (2011) and Cairns *et al.* (2014).

6.5.4. *Sensitivity to the covariance between state vectors.* Recall that Q represents the covariance matrix of the innovation vectors, thereby governing the static and dynamic correlations between the hidden states. The hedging strategy derived from the generalized state-space method depends on Q, but those derived from the delta- and delta–nuga methods do not. Hence, it is warranted to examine how the performance of different hedging strategies may change when Q is altered.

We consider the following five hypothetical situations:

a. all elements are scaled up or down;
b. elements related to $\kappa_1(t)$ (i.e., $Q_{1,i}$ and $Q_{i,1}$ for $i = 1, 2, 3, 4$) are scaled up or down;
c. elements related to $\kappa_2(t)$ (i.e., $Q_{2,i}$ and $Q_{i,2}$ for $i = 1, 2, 3, 4$) are scaled up or down;
d. elements related to $C_1(t)$ (i.e., $Q_{3,i}$ and $Q_{i,3}$ for $i = 1, 2, 3, 4$) are scaled up or down;
e. elements related to $C_2(t)$ (i.e., $Q_{4,i}$ and $Q_{i,4}$ for $i = 1, 2, 3, 4$) are scaled up or down.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 14

THE CALCULATED VALUES OF HE WHEN POISSON RISK IS ABSENT AND PRESENT. THE HEDGE PORTFOLIOS ARE COMPOSED OF $m = 1, 2, 3, 4$ q-FORWARDS AND ARE BUILT BY APPLYING THE GENERALIZED STATE-SPACE METHOD TO THE ORIGINAL CBD MODEL (GROUP 3) AND THE LLCBD MODEL (GROUP 4).

	Group 3					Group 4				
Without Poisson Risk										
	HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
$m = 1$	0.8000	–	60.0893	–	–	0.8575	–	–	22.3178	–
$m = 2$	0.6496	71.5625	47.9503	–	–	0.9579	–	41.0866	13.4239	–
$m = 3$	0.8557	75.7070	29.6824	9.3110	–	0.9732	64.1560	36.5098	16.1386	–
$m = 4$	0.8728	76.1325	29.7784	7.9015	0.7337	0.9737	66.7273	38.9987	13.2875	0.8685
With Poisson Risk, $l_{x_0,t_0+1} = 1,000$										
	Group 3					Group 4				
	HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
$m = 1$	0.4271	–	60.0893	–	–	0.4487	–	–	22.3178	–
$m = 2$	0.3451	71.5625	47.9503	–	–	0.5062	–	41.0866	13.4239	–
$m = 3$	0.4528	75.7070	29.6824	9.3110	–	0.5143	64.1560	36.5098	16.1386	–
$m = 4$	0.4618	76.1325	29.7784	7.9015	0.7337	0.5151	66.7273	38.9987	13.2875	0.8685
With Poisson Risk, $l_{x_0,t_0+1} = 10,000$										
	Group 3					Group 4				
	HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
$m = 1$	0.7358	–	60.0893	–	–	0.7829	–	–	22.3178	–
$m = 2$	0.5986	71.5625	47.9503	–	–	0.8776	–	41.0866	13.4239	–
$m = 3$	0.7876	75.7070	29.6824	9.3110	–	0.8921	64.1560	36.5098	16.1386	–
$m = 4$	0.8027	76.1325	29.7784	7.9015	0.7337	0.8926	66.7273	38.9987	13.2875	0.8685
With Poisson Risk, $l_{x_0,t_0+1} = 100,000$										
	Group 3					Group 4				
	HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
$m = 1$	0.7922	–	60.0893	–	–	0.8503	–	–	22.3178	–
$m = 2$	0.6432	71.5625	47.9503	–	–	0.9499	–	41.0866	13.4239	–
$m = 3$	0.8475	75.7070	29.6824	9.3110	–	0.9648	64.1560	36.5098	16.1386	–
$m = 4$	0.8644	76.1325	29.7784	7.9015	0.7337	0.9653	66.7273	38.9987	13.2875	0.8685
With Poisson Risk, $l_{x_0,t_0+1} = 1,000,000$										
	Group 3					Group 4				
	HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
$m = 1$	0.7994	–	60.0893	–	–	0.8567	–	–	22.3178	–
$m = 2$	0.6491	71.5625	47.9503	–	–	0.9570	–	41.0866	13.4239	–
$m = 3$	0.8550	75.7070	29.6824	9.3110	–	0.9723	64.1560	36.5098	16.1386	–
$m = 4$	0.8720	76.1325	29.7784	7.9015	0.7337	0.9727	66.7273	38.9987	13.2875	0.8685
With Poisson Risk, $l_{x_0,t_0+1} = 10,000,000$										
	Group 3					Group 4				
	HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
$m = 1$	0.8000	–	60.0893	–	–	0.8575	–	–	22.3178	–
$m = 2$	0.6496	71.5625	47.9503	–	–	0.9578	–	41.0866	13.4239	–
$m = 3$	0.8557	75.7070	29.6824	9.3110	–	0.9731	64.1560	36.5098	16.1386	–
$m = 4$	0.8727	76.1325	29.7784	7.9015	0.7337	0.9736	66.7273	38.9987	13.2875	0.8685

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 15

THE VALUES OF HE, N_1, N_2, N_3 AND N_4 WHEN Q IS ALTERED IN DIFFERENT MANNERS.

	Group 2 Assumed Model: The LLCBD Model Method: Traditional Delta–Nuga Hedging					Group 4 Assumed Model: The LLCBD Model Method: Generalized State-Space Hedging				
Multiplier	HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
(a) All Elements are Scaled Up or Down										
0.2	0.8903	157.7860	1.0128	27.4379	−0.0000	0.9803	66.7273	38.9987	13.2875	0.8685
0.5	0.8849	157.7860	1.0128	27.4379	−0.0000	0.9779	66.7273	38.9987	13.2875	0.8685
2	0.8544	157.7860	1.0128	27.4379	−0.0000	0.9641	66.7273	38.9987	13.2875	0.8685
5	0.7788	157.7860	1.0128	27.4379	−0.0000	0.9287	66.7273	38.9987	13.2875	0.8685
(b) Elements Related to $\kappa_1(t)$ (i.e., $Q_{1,i}$ and $Q_{i,1}$ for $i = 1, 2, 3, 4$) are Scaled Up or Down										
0.2	0.9583	157.7860	1.0128	27.4379	−0.0000	0.9867	87.1269	41.1404	13.2163	1.1126
0.5	0.9401	157.7860	1.0128	27.4379	−0.0000	0.9838	78.9923	40.7965	13.2949	1.0401
2	0.3661	157.7860	1.0128	27.4379	−0.0000	0.9002	50.0976	33.4547	12.4363	0.1343
5	0.4538	157.7860	1.0128	27.4379	−0.0000	0.8918	62.6005	32.7750	11.8342	−2.9903
(c) Elements Related to $\kappa_2(t)$ (i.e., $Q_{2,i}$ and $Q_{i,2}$ for $i = 1, 2, 3, 4$) are Scaled Up or Down										
0.2	0.9306	157.7860	1.0128	27.4379	−0.0000	0.9821	67.7555	40.5623	13.2294	0.9485
0.5	0.9151	157.7860	1.0128	27.4379	−0.0000	0.9801	67.4009	39.9414	13.2894	0.9192
2	0.7206	157.7860	1.0128	27.4379	−0.0000	0.9450	65.5549	37.4718	13.1057	0.6844
5	0.6440	157.7860	1.0128	27.4379	−0.0000	0.9133	78.7940	37.4970	12.9966	0.4223
(d) Elements Related to $C_1(t)$ (i.e., $Q_{3,i}$ and $Q_{i,3}$ for $i = 1, 2, 3, 4$) are Scaled Up or Down										
0.2	0.8134	157.7860	1.0128	27.4379	−0.0000	0.9673	48.8085	33.5628	12.5653	1.0407
0.5	0.7978	157.7860	1.0128	27.4379	−0.0000	0.9745	21.5914	30.1378	12.2326	0.8986
2	0.9412	157.7860	1.0128	27.4379	−0.0000	0.9781	82.4635	41.9468	13.3386	0.9666
5	0.8541	157.7860	1.0128	27.4379	−0.0000	0.8821	85.7639	43.2876	12.6638	1.1792
(e) Elements Related to $C_2(t)$ (i.e., $Q_{3,i}$ and $Q_{i,3}$ for $i = 1, 2, 3, 4$) are Scaled Up or Down										
0.2	0.9161	157.7860	1.0128	27.4379	−0.0000	0.9784	84.0152	40.8037	13.4928	1.0814
0.5	0.8608	157.7860	1.0128	27.4379	−0.0000	0.9684	79.1166	40.3986	13.6192	0.7173
2	0.8781	157.7860	1.0128	27.4379	−0.0000	0.9480	53.2548	39.4064	13.2301	1.0908
5	0.1753	157.7860	1.0128	27.4379	−0.0000	0.5724	74.0987	41.8060	12.9891	1.1607

Four scaling factors, 0.2, 0.5, 2 and 5, are considered. The simulation model used is the LLCBD model (with the altered covariance matrix). The test results for hedges with $m = 4$ q-forwards are presented in Table 15.

As expected, when the delta–nuga hedging method is used, altering Q has no impact on the values of $N_1, \ldots, N_4$. However, the value of HE changes considerably when Q is scaled up or down. The value of HE is the lowest (0.1753) when the elements related to $C_2(t)$ are scaled up by a factor of 5.

When the generalized state-space hedging method is used, the values of $N_1, \ldots, N_4$ are adaptive to the modifications made to Q. As the notional amounts are "corrected" accordingly,[7] the resulting value of HE is much more robust relative to changes in Q. For every situation under consideration, the value of HE produced by the generalized state-space hedging method is higher than that produced by the delta–nuga hedging method.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 16

THE VALUES OF HE AND $N_1, \ldots, N_m$ PRODUCED BY THE GENERALIZED STATE-SPACE METHOD WHEN DIFFERENT NUMBERS OF q-FORWARDS ARE USED.

HE	N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8	N_9	N_{10}
0.8425	83.1379	–	–	–	–	–	–	–	–	–
0.8517	9.6612	78.6630	–	–	–	–	–	–	–	–
0.8583	19.5117	4.3579	67.1848	–	–	–	–	–	–	–
0.8640	26.2820	4.5859	4.9870	57.6373	–	–	–	–	–	–
0.8685	30.9677	4.7438	5.2967	5.6566	49.0497	–	–	–	–	–
0.8717	34.1813	4.8520	5.5091	5.9726	6.2574	40.9674	–	–	–	–
0.8736	36.3330	4.9245	5.6513	6.1842	6.5404	6.7341	33.1843	–	–	–
0.8749	37.7136	4.9710	5.7425	6.3200	6.7220	6.9642	7.0601	25.6171	–	–
0.8754	38.5350	4.9987	5.7968	6.4008	6.8301	7.1011	7.2282	7.2236	18.2404	–
0.8758	38.9536	5.0128	5.8245	6.4420	6.8851	7.1709	7.3139	7.3269	7.2210	11.0539

6.5.5. *Sensitivity to the number of* q-*forwards used.* As previously mentioned, the generalized state-space hedging method can be implemented with any number of q-forwards (m), because the solution to Equation (5.6) exists for any $m \geq 1$. We now examine how the performance of the generalized state-space hedging method may change as the number of q-forwards used increases.

We consider the following collection of q-forwards:

j	Reference Age x_j	Reference Year t_j	j	Reference Age x_j	Reference Year t_j
1	80	2025	6	80	2030
2	80	2026	7	80	2031
3	80	2027	8	80	2032
4	80	2028	9	80	2033
5	80	2029	10	80	2034

Note that it is not possible to use this collection of q-forwards to build delta or delta–nuga hedges, because they are all linked to the same reference age.[8]

When $m = 1$, the hedge contains q-forward $j = 1$; when $m = 2$, the hedge contains q-forwards $j = 1, 2$; and so on. The hedging results for $m = 1, \ldots, 10$ are reported in Table 16. As m increases, the value of HE increases. However, the rate of increase in HE reduces with m, suggesting that the marginal benefit of adding an additional q-forward becomes small when the hedge portfolio already contains a large number of q-forwards.

When m becomes very large, inverting Σ_h may become difficult as it may be close to singular due to its large dimension. Users of this method should be mindful of this potential problem, although having a very large m is not likely.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 17

RESULT I (A COMPARISON OF DIFFERENT HEDGING METHODS) BASED ON THE ALTERNATIVE SET OF q-FORWARDS. FOR GROUPS 1 AND 3, THE SIMULATION MODEL USED IS THE ORIGINAL CBD MODEL. FOR GROUPS 2 AND 4, THE SIMULATION MODEL USED IS THE LLCBD MODEL.

Group 1 Assumed model: The Original CBD Model Method: Traditional Delta Hedging					Group 3 Assumed Model: The Original CBD Model Method: Generalized State-Space Hedging				
HE	N_1	N_2	N_3	N_4	*HE*	N_1	N_2	N_3	N_4
0.8510	152.5614	43.2934	–	–	0.9135	112.2606	35.4131	–	–
−1.6532	−378.3325	–	562.1118	–	0.7646	112.5175	–	137.0038	–
−1.2703	−168.3086	–	–	367.6104	0.7597	152.6432	–	–	97.0379
0.7050	–	30.8523	161.5324	–	0.9106	–	30.8712	70.1558	–
0.4764	–	22.7090	–	174.7847	0.9107	–	32.3435	–	51.5237
−2.8836	–	–	−450.4640	662.2052	0.7358	–	–	120.3346	48.6849
Group 2 Assumed Model: The LLCBD Model Method: Traditional Delta/Delta–Nuga Hedging					Group 4 Assumed Model: The LLCBD Model Method: Generalized State-Space Hedging				
HE	N_1	N_2	N_3	N_4	*HE*	N_1	N_2	N_3	N_4
0.6880	178.7526	59.9580	–	–	0.7905	11.7794	57.6941	–	–
−9.1519	−469.6742	–	806.0370	–	0.7086	−238.5952	–	184.9745	–
−12.2549	−213.1537	–	–	575.5422	0.7694	−124.0271	–	–	118.4957
0.0151	–	43.4293	222.2012	–	0.8300	–	45.7134	51.2775	–
−1.6542	–	32.6105	–	262.5109	0.8929	–	37.2561	–	58.1600
−18.8856	–	–	−669.7701	1053.7844	0.7267	–	–	−44.6941	136.6734
0.5892	349.2159	66.6085	−399.0373	221.0896	0.9132	−20.9026	37.8663	−60.2340	96.9143

6.5.6. *Sensitivity to choice of* q-*forwards.* Finally, we examine the sensitivity of the hedging results to the choice of q-forwards. The following set of q-forwards is considered:

j	Reference Age x_j	Reference Year t_j
1	70	2020
2	88	2025
3	75	2030
4	77	2035

In contrast to the q-forwards used in the baseline results, the q-forwards considered here are linked to distinct birth cohorts.

Table 17 shows Result I (a comparison of different hedging methods) on the basis of the alternative set of q-forwards. The conclusion obtained in Section 6.2 remains unchanged: The generalized state-space method consistently yields a better hedge effectiveness compared to the delta and delta–nuga methods, regardless of whether the original CBD model or the LLCBD model is assumed in the derivation.

Table 18 displays Result II (the impact of model mis-specification) on the basis of the alternative set of q-forwards. The conclusion drawn in Section 6.3

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

TABLE 18

RESULT II (THE IMPACT OF MODEL MIS-SPECIFICATION) DERIVED USING THE ALTERNATIVE SET OF q-FORWARDS. THE SIMULATION MODEL USED IS THE LLCBD MODEL.

Group 3 Assumed Model: The Original CBD Model Method: Generalized State-Space Hedging					Group 4 Assumed Model: The LLCBD Model Method: Generalized State-Space Hedging				
HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
Number of q-Forwards Used: $m = 1$									
0.7413	–	42.9262	–	–	0.7899	–	57.7622	–	–
Number of q-Forwards Used: $m = 2$									
0.8825	–	32.3435	–	51.5237	0.8929	–	37.2561	–	58.1600
Number of q-Forwards Used: $m = 3$									
0.7906	86.7842	28.9276	–	39.8774	0.9123	–	39.0704	−72.4252	101.2808
Number of q-Forwards Used: $m = 4$									
0.7840	82.7462	28.4464	10.8472	34.4610	0.9132	−20.9026	37.8663	−60.2340	96.9143

still stands: Neglecting stochastic drifts may lead to a material loss of hedge effectiveness, if the true underlying model is one with stochastic drifts.

7. CONCLUDING REMARKS

Longevity risk comprises of diffusion risk and drift risk. Although both sources of risk are significant, the latter is often ignored in the existing stochastic mortality models. In this paper, we introduce the LLCBD model which captures drift risk by allowing the drifts themselves to follow a random walk. Written in a state-space form, the LLCBD model contains four hidden states, $\kappa_1(t)$, $\kappa_2(t)$, $C_1(t)$ and $C_2(t)$, all of which have demographic intuitions behind. All hidden states and parameters in the model can be estimated in one single stage by using the EM algorithm and the Kalman filter.

As an illustration, we estimate the LLCBD model to the historical mortality data of Canadian males. The adequacy of the model's fit is confirmed by Harvey's (1990) test that is based on the model's vector of prediction errors. In comparison to the original CBD model, the LLCBD model provides a better goodness-of-fit in terms of AIC, and yields more accurate short- and long-term forecasts in terms of ME and MSE. We also find that the LLCBD model generates forecasts that are more consistent with the observed trends in the recent past and are more robust relative to changes in the length of the calibration window. Because the LLCBD model incorporates additionally the uncertainty

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

associated with the drifts, it results in wider long-term prediction intervals that reflect the possibilities of future trend changes.

Another contribution of this paper is the generalized state-space hedging method, from which one can construct an index-based longevity hedge to mitigate both diffusion and drift risks. As explained in Section 5.5, the proposed hedging method can ameliorate the problems of sub-optimality and singularity that the traditional delta- and delta–nuga hedging methods are subject to. The proposed hedging method does not impose any requirement on the number of hedging instruments used. It also works for any combination of hedging instruments, provided that the payoffs from the hedging instruments are not perfectly correlated with one another.

As an example, we apply the generalized state-space hedging method to a hypothetical hedging scenario. The results of this application point to three conclusions. First, the proposed hedging method performs better than the traditional delta and delta–nuga hedging methods, no matter which of the two models under consideration is assumed. Second, ignoring stochastic drifts in the derivation of the hedging strategies would lead to a material reduction in hedge effectiveness, if the true underlying model is one with stochastic drifts. Third, the negative impact of ignoring stochastic drifts is particularly significant if the duration of the liability being hedged is long.

To focus on the issue concerning drift risk, in this paper, we choose to build our model on the simplest version of the CBD model, which does not take cohort effects into account. We acknowledge that cohort effects are significant in certain populations, and that it is not trivial to incorporate cohort effects in a state-space representation in which the vector of hidden states evolve over time rather than year of birth. In future research, it would be interesting to investigate how the LLCBD model can be further extended to incorporate cohort effects. This goal may possibly be accomplished by using the parsimonious approach introduced by Marvos *et al.* (2014), whereby cohort effects are captured through the dependence among residuals.

To ease exposition, we have also assumed that the q-forwards and the annuity liability are associated with exactly the same population. In reality, the q-forwards are more likely to be linked to a broad-based mortality index rather than the hedger's own population, giving rise to population basis risk that may reduce hedge effectiveness. Future research warrants a study on how our contributions may be applied to situations when population basis risk exists. Such a study would encompass an extension of the LLCBD model to a version which, similar to the models contributed by Li and Lee (2005), Cairns *et al.* (2011b), Dowd *et al.* (2011) and Li *et al.* (2015b), models the mortality of two populations in a coherent manner. The study would also include an adjustment of the generalized state-space hedging method to capture the imperfect correlations between the mortality improvements of the two populations involved.

The longevity hedging strategy presented in this paper is static, as no adjustment is made to the hedge after inception. Static hedging strategies generally require longer dated instruments, but the majority of capital market investors

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

prefer to invest in shorter dated ones. It would therefore be useful to extend the proposed hedging method from static to dynamic. The accomplishment of this research goal is likely to computationally demanding, as nested simulations are normally required when assessing the effectiveness of a dynamic longevity hedge. To overcome the computational challenge, approximation methods such as that proposed by Cairns (2011) may be required.

Acknowledgements

This work is supported by the Global Risk Institute, the Society of Actuaries Center of Actuarial Excellence Programme and the Natural Sciences and Engineering Research Council of Canada (Discovery Grant RGPIN-356050-2013).

Notes

1. The approximation is exact if deaths are uniformly distributed between two consecutive integer ages.
2. If $v_{x,t}$'s are i.i.d. but *not* normal, then $\mathcal{D}$ would converge to another constant instead.
3. We do not consider Model M3 (the age-period cohort model), because it is simply a special case of Model M2 with $b_1(x)$ and $b_2(x)$ being constants instead of functions of age. Also, Model M4 (the P-splines regression) is excluded, in part because it is based on a regression rather than stochastic processes and in part because it does not yield sample mortality paths which are needed for the analyses in later parts of the paper.
4. The Life and Longevity Markets Association (www.llma.org).
5. In effect, $\hat{L}$ and $\hat{H}_j$ are respectively the values of L and H_j calculated by switching off all random components.
6. The annuity liability and all hedging instruments used are related to ages greater than $\bar{x} = 69.5$, so the projected mortality would be high (low) when both $\kappa_1(t)$ and $\kappa_2(t)$ are high (low). It follows that (1) and (a) are the linear sample paths that would result in the highest projected mortality, whereas (2) and (b) are the linear sample paths that would lead to the lowest projected mortality.
7. Situation (a) is an exception. If all elements of Q are scaled by the same factor, then according to Equation (5.5), all elements in $\Xi^*_{i,j}$, for any i and j, would also be scaled by exactly the same factor. As a result, we can cancel out the scaling factor in Equation (5.7), resulting in no change in the values of $N_1, \ldots, N_4$.
8. When the q-forwards are linked to the same reference age, the delta and delta–nuga methods are subject to the singularity problem (see Section 5.5.2).

References

Ahmadi, S. and Li, J.S.-H. (2014) Coherent mortality forecasting with generalized linear models: A modified time-transformation approach. *Insurance: Mathematics and Economics*, **59**, 194–221.

Booth, H., Maindonald, J. and Smith, L. (2002) Applying Lee-Carter under conditions of variable mortality decline. *Population Studies*, **56**, 325–336.

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

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

Cairns, A.J.G., Blake, D. and Dowd, K. (2006) A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, **73**, 687–718.
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. and Khalaf-Allah, M. (2011a) Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, **48**, 355–367.
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. (2011b) Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin*, **41**, 29–59.
Canadian Institute of Actuaries (2014) Final report on Canadian pensioners' mortality. Available at http://www.cia-ica.ca/docs/default-source/2014/214013e.pdf.
Carter, L.R. (1996) Forecasting U.S. Mortality: A comparison of Box-Jenkins ARIMA and structural time series models. *The Sociological Quarterly*, **37**, 127–144.
Cavanaugh, J. and Shumway, R. (1997) A bootstrap variant of AIC for state-space model selection. *Statistica Sinica*, **7**, 473–496.
Coelho, E. and Nunes, L. (2011) Forecasting mortality in the event of a structural change. *Journal of the Royal Statistical Society Series A*, **174**, 713–736.
Coughlan, G. (2009) Longevity risk transfer: Indices and capital market solutions. In *The Handbook of Insurance Linked Securities* (eds. P.M. Barrieu and L. Albertini), pp. 261–281. London: Wiley.
Coughlan, G., Blake, D., MacMinn, R., Cairns, A.J.G. and Dowd, K. (2013) Longevity risk and hedging solutions. In *Handbook of Insurance* (eds. G. Dionne), pp. 997–1035. New York: Springer.
Coughlan, G.D., Khalaf-Allah, M., Ye, Y., Kumar, S., Cairns, A.J.G., Blake, D. and Dowd, K. (2011) Longevity hedging 101: A framework for longevity basis risk analysis and hedge effectiveness. *North American Actuarial Journal*, **15**, 150–176.
de Jong, P. and Tickle, L. (2006) Extending Lee-Carter mortality forecasting. *Mathematical Population Studies*, **13**, 1–18.
Denuit, M. and Goderniaux, A. (2005) Closing and projecting lifetables using log-linear models. *Bulletin of the Swiss Association of Actuaries*, **1**, 29–49.
Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D. and Khalaf-Allah, M. (2010a) Evaluating the goodness of fit of stochastic mortality models. *Insurance: Mathematics and Economics*, **47**, 255–265.
Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D. and Khalaf-Allah, M. (2010b) 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**, 331–356.
Gallop, A. (2006) Mortality improvements and evolution of life expectancies. Paper presented at the Seminar on Demographic, Economic and Investment Perspectives for Canada, Office of the Superintendent of Financial Institutions Canada.
Girosi, F. and King, G. (2005) A reassessment of Lee-Carter mortality forecasting method. *International Journal of Forecasting*, **21**, 249–260.
Hári, N., de Waegenaere, A., Melenberg, B. and Nijman, T.E. (2008) Estimating the term structure of mortality. *Insurance: Mathematics and Economics*, **42**, 492–504.
Harvey, A.C. (1990) *Forecasting, Structural Time Series Models and the Kalman Filter.* New York: Cambridge University Press.
Human Mortality Database (2014) University of California, Berkeley (USA), and Max Planck Institute of Demographic Research (Germany) Available at www.mortality.org or www.humanmortality.de (data downloaded on 1 April 2014).
Holmes, E.E. (2013) Derivation of an EM algorithm for constrained and unconstrained multivariate autoregressive state-space (MARSS) models. preprint (arXiv:1302.3919).

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

IMHOF, J.P. (1961) Computing the distribution of quadratic forms in normal variables. *Biometrika*, **48**, 419–426.

KALMAN, R.E. (1960) A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, **82**, 35–45.

KALMAN, R.E. and BUCY, R.S. (1961) New results in linear filtering and prediction theory. *Journal of Fluids Engineering*, **83**, 95–108.

KANNISTO, V., LAURISTEN, J., THATCHER, A.R. and VAUPEL, J.W. (1994) Reduction in mortality at advanced ages: Several decades of evidence from 27 countries. *Population Development Review*, **20**, 793–810.

LAMOTTE, L.R. and MCWHORTER, A. (1978) An exact test for the presence of random walk Coefficients in a linear regression model. *Journal of the American Statistical Association*, **364**, 816–820.

LEE, R. and CARTER, L. (1992) Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, **87**, 659–671.

LEE, R. and MILLER, T. (2001) Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, **38**, 537–549.

LI, H., DE WAEGENAERE, A. and MELENBERG, B. (2015a) The choice of sample size for mortality forecasting: A Bayesian learning approach. *Insurance: Mathematics and Economics*, **63**4, 153–168.

LI, J.S.-H., CHAN, W.S. and CHEUNG, S.H. (2011) Structural changes in the Lee-Carter mortality indexes: Detection and implications. *North American Actuarial Journal*, **15**, 13–31.

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, J.S.-H., ZHOU, R. and HARDY, M.R. (2015b) A step-by-step guide to building multi-population stochastic mortality models. *Insurance: Mathematics and Economics*, **63**, 121–134.

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

LI, S.H. and CHAN, W.S. (2005) Outlier analysis and mortality forecasting: The United Kingdom and Scandinavian countries. *Scandinavian Actuarial Journal*, **3**, 187–211.

LUCIANO, E., REGIS, L. and VIGNA, E. (2012) Delta-gamma hedging of mortality and interest rate risk. *Insurance: Mathematics and Economics*, **50**, 402–412.

MAVROS, G., CAIRNS, A.J.G., KLEINOW, T. and STREFTARIS, G. (2015) *A parsimonious approach to stochastic mortality modelling with dependent residuals,* Working Paper, Heriot-Watt University.

MILIDONIS, A., LIN, Y. and COX, S.H. (2011) Mortality regimes and pricing. *North American Actuarial Journal*, **15**, 266–289.

NYBLOM, J. and MÄKELÄINEN, T. (1983) Comparisons of tests for the presence of random walk coefficients in a simple linear model. *Journal of the American Statistical Association*, **384**, 856–864.

OEPPEN, J. and VAUPEL, J.W. (2002) Broken limits to life expectancy. *Science*, **296**, 1029–1031.

O'HARE, C. and LI, Y. (2015) Identifying structural breaks in stochastic mortality models. *Journal of Risk and Uncertainty in Engineering Part B*, doi: 10.1115/1.4029740.

PEDROZA, C. (2006) A Bayesian forecasting model: Predicting U.S. male mortality. *Biostatistics*, **7**, 530–550.

RENSHAW, A. and HABERMAN, S. (2003) Lee-Carter mortality forecasting: A parallel generalized linear modelling approach for England and wales mortality projections. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **52**, 119–137.

SHUMWAY, R.H. and STOFFER, D.S. (2006) *Time Series Analysis and Its Applications: With R Examples*. New York: Springer-Verlag.

Society of Actuaries. (2014) Mortality Improvement Scale MP-2014 Report. Available at https://www.soa.org/Files/Research/Exp-Study/research-2014-mp-report.pdf.

STOFFER, D.S. and WALL, K.D. (1991) Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter. *Journal of the American Statistical Association*, **86**, 1024–1033.

SWEETING, P.J. 2011. A trend-change extension of the Cairns-Blake-Dowd model. *Annals of Actuarial Science*, **5**, 143–162.
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.
VAN BERKUM, F., ANTONIO, K. and VELLEKOOP, M. (2014) The impact of multiple structural changes on mortality predictions. *Scandinavian Actuarial Journal*, doi: 10.1080/03461238.2014.987807.
VAUPEL, J.W. (1997) The remarkable improvements in survival at older ages. *Philosophical transactions of the Royal Society of London, B*, **352**, 1799–1804.
WILMOTH, J.R. ANDREEV, E., JDANOV, D.A. and GLEI, D.A. (2005) Methods protocol for the human mortality database. Available at: www.mortality.org.
ZHOU, K.Q. and LI, J.S.-H. (2014) Dynamic longevity hedging in the presence of population basis risk: A feasibility analysis from technical and economic perspectives. Paper presented at the *10th International Longevity Risk and Capital Markets Solutions Symposium*, Santiago, Chile.
ZHOU, R. and LI, J.S.H. (2013) A cautionary note on pricing longevity index swaps. *Scandinavian Actuarial Journal*, **2013**, 1–23.

YANXIN LIU
Department of Finance
University of Nebraska-Lincoln, Lincoln, NE, USA
E-Mail: yanxin.liu@unl.edu

JOHNNY SIU-HANG LI (Corresponding author)
Department of Statistics and Actuarial Science
University of Waterloo, Waterloo, ON, Canada
E-Mail: shli@uwaterloo.ca

APPENDIX: ESTIMATION PROCEDURE

In this appendix, we detail the estimation procedure, which is adapted from the work of Holmes (2013). As before, we use $[t_a, t_b]$ and $[x_a, x_b]$ to denote the sample period and sample age range to which the model is fitted, respectively. The vector of observations at time t is $\vec{y}_t = (y_{x_a,t}, \ldots, y_{x_b,t})'$.

Following Holmes (2013), we use an Expectation–Maximization (EM) algorithm to obtain maximum likelihood parameter estimates. Under our model assumptions,

$$\vec{\eta}_t \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, Q) \quad \text{and} \quad \vec{\epsilon}_t \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, \text{I}_{x_b - x_a + 1}\sigma_\epsilon^2).$$

It immediately follows that the log-likelihood function is given by

$$\ln(\mathcal{L}) = -\frac{1}{2\sigma_\epsilon^2}\sum_{t=t_a}^{t_b}(\vec{y}_t - B\vec{\alpha}_t)'(\vec{y}_t - B\vec{\alpha}_t) - \frac{(t_b - t_a + 1)(x_b - x_a + 1)}{2}\ln(\sigma_\epsilon^2)$$

$$-\frac{1}{2}\sum_{t=t_a+1}^{t_b}(\vec{\alpha}_t - A\vec{\alpha}_{t-1})'\,Q^{-1}\,(\vec{\alpha}_t - A\vec{\alpha}_{t-1}) - \frac{1}{2}\sum_{t=t_a+1}^{t_b}\ln|Q| + c_l,$$

where c_l is a constant that is free of the hidden states and parameters. When fitting the special case with constant drifts, $\vec{\alpha}_t$ and Q should be replaced by $\vec{\alpha}_t^*$ and Q^*, respectively.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

The EM algorithm iterates over two steps until convergence. In the first step, which is known as the Expectation step, the expectation of the log-likelihood is computed:

$$\begin{aligned}\Psi &= \mathrm{E}\left[\ln(\mathcal{L})\right] \\ &= -\frac{1}{2\sigma_\epsilon^2}\sum_{t=t_a}^{t_b}\left(\mathrm{E}[\vec{y}_t'\vec{y}_t] - 2\mathrm{E}[\vec{y}_t' B\vec{\alpha}_t] + \mathrm{E}[\vec{\alpha}_t' B' B\vec{\alpha}_t]\right) \\ &\quad - \frac{(t_b - t_a + 1)(x_b - x_a + 1)}{2}\ln(\sigma_\epsilon^2) \\ &\quad - \frac{1}{2}\sum_{t=t_a+1}^{t_b}\left(\mathrm{E}[\vec{\alpha}_t' Q^{-1}\vec{\alpha}_t] - 2\mathrm{E}[\vec{\alpha}_{t-1}' A' Q^{-1}\vec{\alpha}_t] + \mathrm{E}[\vec{\alpha}_{t-1}' A' Q^{-1} A\vec{\alpha}_{t-1}]\right) \\ &\quad - \frac{t_b - t_a}{2}\ln|Q| + c_l.\end{aligned}$$

In the second step, which is known as the Maximization step, parameter estimates are obtained by maximizing the expected log-likelihood Ψ. We now derive the update equations for parameters σ_ϵ^2 and Q in the Maximization step.

Update equation for σ_ϵ^2:
Differentiating Ψ with respect to σ_ϵ^2, we have

$$\begin{aligned}\frac{\partial\Psi}{\partial\sigma_\epsilon^2} &= -\frac{1}{2}\sum_{t=t_a}^{t_b}\frac{-\mathrm{E}\left[\vec{y}_t'\vec{y}_t\right] + 2\mathrm{E}\left[\vec{y}_t' B\vec{\alpha}_t\right] - \mathrm{E}\left[\vec{\alpha}_t' B' B\vec{\alpha}_t\right]}{\sigma_\epsilon^4} \\ &\quad - \frac{(t_b - t_a + 1)(x_b - x_a + 1)}{2\sigma_\epsilon^2}.\end{aligned}$$

Setting the partial derivative to 0, we can obtain the update equation for R as follows:

$$\begin{aligned}\sigma_\epsilon^2 &= \frac{1}{u}\sum_{t=t_a}^{t_b}\left(\mathrm{E}\left[\vec{y}_t'\vec{y}_t\right] - 2\mathrm{E}\left[\vec{y}_t' B\vec{\alpha}_t\right] + \mathrm{E}\left[\vec{\alpha}_t' B' B\vec{\alpha}_t\right]\right) \\ &= \frac{1}{u}\sum_{t=t_a}^{t_b}\mathrm{E}\left[(\vec{y}_t - B\vec{\alpha}_t)'(\vec{y}_t - B\vec{\alpha}_t)\right] \\ &= \frac{1}{u}\sum_{t=t_a}^{t_b}\mathrm{E}[\mathrm{vec}((\vec{y}_t - B\vec{\alpha}_t)'\mathrm{I}_{x_b - x_a + 1}(\vec{y}_t - B\vec{\alpha}_t))] \\ &= \frac{1}{u}\sum_{t=t_a}^{t_b}\mathrm{E}\left[(\vec{y}_t - B\vec{\alpha}_t)' \otimes (\vec{y}_t - B\vec{\alpha}_t)'\right]\mathrm{vec}\left(\mathrm{I}_{x_b - x_a + 1}\right) \\ &= \frac{1}{u}\sum_{t=t_a}^{t_b}\mathrm{E}\left[\mathrm{vec}((\vec{y}_t - B\vec{\alpha}_t)(\vec{y}_t - B\vec{\alpha}_t)')'\right]\mathrm{vec}\left(\mathrm{I}_{x_b - x_a + 1}\right) \\ &= \frac{1}{u}\sum_{t=t_a}^{t_b}\left(\mathrm{vec}(\mathrm{E}[\vec{y}_t\vec{y}_y'])' - 2\mathrm{vec}(\mathrm{E}[\vec{y}_t\vec{\alpha}_t']B')' + \mathrm{vec}(B\mathrm{E}[\vec{\alpha}_t\vec{\alpha}_t']B')'\right)\mathrm{vec}\left(\mathrm{I}_{x_b - x_a + 1}\right),\end{aligned}$$

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

where vec(X) represents the vectorization of X, $\otimes$ is the kronecker product operator and $u = (t_b - t_a + 1)(x_b - x_a + 1)$.

Update equation for Q:
The partial derivative of Ψ with respect to Q can be calculated as

$$\frac{\partial \Psi}{\partial Q} = \frac{1}{2}\sum_{t=t_a+1}^{t_b} Q^{-1}\Big(E[\vec{\alpha}_t\vec{\alpha}_t'] - E[\vec{\alpha}_t\vec{\alpha}_{t-1}']A' - AE[\vec{\alpha}_{t-1}\vec{\alpha}_t'] + AE[\vec{\alpha}_{t-1}\vec{\alpha}_{t-1}']A'\Big)Q^{-1} - \frac{t_b - t_a}{2}Q^{-1}.$$

Setting the partial derivative to zero, we obtain the following update equation for Q:

$$Q = \frac{1}{t_b - t_a}\sum_{t=t_a+1}^{t_b}\Big(E[\vec{\alpha}_t\vec{\alpha}_t'] - E[\vec{\alpha}_t\vec{\alpha}_{t-1}']A' - AE[\vec{\alpha}_{t-1}\vec{\alpha}_t'] + AE[\vec{\alpha}_{t-1}\vec{\alpha}_{t-1}']A'\Big).$$

Calculating the expectations in the update equations:
We now explain how the expectations in the update equations for σ_ϵ^2 and Q can be evaluated. Let $\mathrm{E}_t(\cdot)$, $\mathrm{Var}_t(\cdot)$, $\mathrm{Cov}_t(\cdot,\cdot)$, respectively be the expectation, variance and covariance conditioned on the information up to and including time t. The expectations in the update equations are calculated on the basis of all data, so that, for example, $\mathrm{E}[\vec{\alpha}_t\vec{\alpha}_t']$ is computed as $\mathrm{E}_{t_b}[\vec{\alpha}_t\vec{\alpha}_t']$.

To calculate $\mathrm{E}_{t_b}[\vec{\alpha}_t\vec{\alpha}_{t-1}']$, $\mathrm{E}_{t_b}[\vec{\alpha}_t\vec{\alpha}_t']$, $\mathrm{E}_{t_b}[\vec{\alpha}_{t-1}\vec{\alpha}_t']$ and $\mathrm{E}_{t_b}[\vec{\alpha}_{t-1}\vec{\alpha}_{t-1}']$, we use the Kalman smoother algorithm. The first step in the algorithm is to calculate the expectations of $\vec{\alpha}_t\vec{\alpha}_t'$ and $\vec{\alpha}_t$, conditioned on the information up to an including time t, using the Kalman filter:

$$\begin{cases} \mathrm{E}_t[\vec{\alpha}_t] = \mathrm{E}_{t-1}[\vec{\alpha}_t] + \mathcal{K}_t(y_t - B\mathrm{E}_{t-1}[\vec{\alpha}_t]) \\ \mathrm{Var}_t(\vec{\alpha}_t) = (\mathrm{I}_m - \mathcal{K}_t B)\mathrm{Var}_{t-1}(\vec{\alpha}_t) \\ \mathrm{E}_t[\vec{\alpha}_t\vec{\alpha}_t'] = \mathrm{Var}_t(\vec{\alpha}_t) + \mathrm{E}_t[\vec{\alpha}_t](\mathrm{E}_t[\vec{\alpha}_t])' \end{cases},$$

where $\mathrm{E}_{t-1}[\vec{\alpha}_t] = A\mathrm{E}_{t-1}[\vec{\alpha}_{t-1}]$, $\mathrm{Var}_{t-1}(\vec{\alpha}_t) = A\mathrm{Var}_{t-1}(\vec{\alpha}_{t-1})A' + Q$ and

$$\mathcal{K}_t = \mathrm{Var}_{t-1}(\vec{\alpha}_t)B'\left(B\mathrm{Var}_{t-1}(\vec{\alpha}_t)B' + \mathrm{I}_{x_b - x_a + 1}\cdot\sigma_\epsilon^2\right)^{-1}$$

is the Kalman gain at time t.

The second step of the algorithm utilizes the Kalman smoother,

$$\begin{cases} \mathrm{E}_{t_b}[\vec{\alpha}_{t-1}] = \mathrm{E}_{t-1}[\vec{\alpha}_{t-1}] + J_{t-1}(\mathrm{E}_{t_b}[\vec{\alpha}_t] - \mathrm{E}_{t-1}[\vec{\alpha}_t]) \\ \mathrm{Var}_{t_b}(\vec{\alpha}_{t-1}) = \mathrm{Var}_{t-1}(\vec{\alpha}_{t-1}) + J_{t-1}(\mathrm{Var}_{t_b}(\vec{\alpha}_t) - \mathrm{Var}_{t-1}(\vec{\alpha}_t))J_t' \\ \mathrm{E}_{t_b}[\vec{\alpha}_t\vec{\alpha}_t'] = \mathrm{Var}_{t_b}(\vec{\alpha}_t) + \mathrm{E}_{t_b}[\vec{\alpha}_t](\mathrm{E}_{t_b}[\vec{\alpha}_t])' \end{cases},$$

and the lag-1 covariance smoother,

$$\begin{cases} \mathrm{Cov}_{t_b}(\vec{\alpha}_{t_b}, \vec{\alpha}_{t_b-1}) = (\mathrm{I}_m - \mathcal{K}_{t_b}B)A\mathrm{Var}_{t_b-1}(\vec{\alpha}_{t_b-1}) \\ \mathrm{Cov}_{t_b}(\vec{\alpha}_{t-1}, \vec{\alpha}_{t-2}) = \mathrm{Var}_{t-1}(\vec{\alpha}_{t-1})J_{t-2}' + J_{t-1}(\mathrm{Cov}_{t_b}(\vec{\alpha}_t, \vec{\alpha}_{t-1}) \\ \qquad\qquad - A\mathrm{Var}_{t-1}(\vec{\alpha}_{t-1}))J_{t-2}' \\ \mathrm{E}_{t_b}[\vec{\alpha}_t\vec{\alpha}_{t-1}'] = \mathrm{Cov}_{t_b}(\vec{\alpha}_t, \vec{\alpha}_{t-1}') + \mathrm{E}_{t_b}[\vec{\alpha}_t](\mathrm{E}_{t_b}[\vec{\alpha}_{t-1}])' \end{cases},$$

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press

where $J_{t-1} = \mathrm{Var}_{t-1}(\vec{\alpha}_{t-1})\, A'(\mathrm{Var}_{t-1}(\vec{\alpha}_t))^{-1}$. By applying the two sets of equations above recursively for $t = t_b, t_b - 1, \ldots$, the required expectations can be obtained readily.

The update equation for σ_ϵ^2 contains two additional expectations, namely $\mathrm{E}[\vec{y}_t \vec{\alpha}_t']$ and $\mathrm{E}\left[\vec{y}_t' \vec{y}_t\right]$. These two expectations are computed as follows:

$$\mathrm{E}[\vec{y}_t \vec{y}_t'] = \vec{\tilde{y}}_t \vec{\tilde{y}}_t';$$
$$\mathrm{E}[\vec{y}_t \vec{\alpha}_t'] = \vec{\tilde{y}}_t \mathrm{E}_{t_b}[\vec{\alpha}_t]',$$

where $\vec{\tilde{y}}_t$ represents the realization of $\vec{y}_t$ and $\mathrm{E}_{t_b}[\vec{\alpha}_t]$ can be evaluated by the Kalman smoother.

https://doi.org/10.1017/asb.2016.33 Published online by Cambridge University Press