© The Journal of Risk and Insurance, 2013, Vol. 80, No. 3, 775–797
DOI: 10.1111/j.1539-6975.2013.12002.x

Mortality Modeling With Non-Gaussian Innovations and Applications to the Valuation of Longevity Swaps

Chou-Wen Wang
Hong-Chih Huang
I-Chien Liu

Abstract

This article provides an iterative fitting algorithm to generate maximum likelihood estimates under the Cox regression model and employs non-Gaussian distributions—the jump diffusion (JD), variance gamma (VG), and normal inverse Gaussian (NIG) distributions—to model the error terms of the Renshaw and Haberman (2006) (RH) model. In terms of mean absolute percentage error, the RH model with non-Gaussian innovations provides better mortality projections, using 1900–2009 mortality data from England and Wales, France, and Italy. Finally, the lower hedge costs of longevity swaps according to the RH model with non-Gaussian innovations are not only based on the lower swap curves implied by the best prediction model, but also in terms of the fatter tails of the unexpected losses it generates.

Introduction

Longevity represents an increasingly important risk for defined benefit pension plans and annuity providers, because life expectancy is dramatically increasing in developed countries. In 2007, exposures to improved life expectancy amounted to

---

Chou-Wen Wang is a Professor in the Department of Risk Management and Insurance, National Kaohsiung First University of Science and Technology, Kaohsiung, Taiwan; Fellow of Risk and Insurance Research Center, College of Commerce, National Chengchi University, Taipei, Taiwan. Chou-Wen Wang can be contacted via e-mail: chouwenwang@gmail.com.
Hong-Chih Huang is a Professor in the Department of Risk Management and Insurance; Fellow of Risk and Insurance Research Center, College of Commerce, National Chengchi University, Taipei, Taiwan. Hong-Chih Huang can be contacted via e-mail: jerry2@nccu.edu.tw.
I-Chien Liu is a Ph.D. in the Department of Risk Management and Insurance, National Chengchi University, Taipei, Taiwan. I-Chien Liu can be contacted via e-mail: horn8802@gmail.com.

$400 billion for pension funds and insurance companies in the United Kingdom and United States (see Loeys, Panigirtzoglou, and Ribeiro, 2007). Stochastic mortality models quantify mortality and longevity risks, which makes mortality risk management possible and provides the foundation for pricing and reserving. Among all stochastic mortality models, the Lee-Carter (LC) model, proposed in 1992, is one of the most popular choices because of its ease of implementation and acceptable prediction errors in empirical studies. Various modifications of the LC model have been extended by Brouhns, Denuit, and Vermunt (2002), Renshaw and Haberman (2003, 2006), Cairns, Blake, and Dowd (2006), Li and Chan (2007), Biffis, Denuit, and Devolder (2010), and Hainaut (2012) to attain a broader interpretation. Cairns, Blake, and Dowd (2006) propose a two-factor stochastic mortality model, the CBD model, in which a first factor affects mortality at all ages, whereas a second factor affects mortality at older ages much more than at younger ages. Modeling the number of deaths with the Poisson model, Cairns et al. (2009) classify and compare eight stochastic mortality models, including an extension of the CBD model, with mortality data from England and Wales and the United States. They find that an extension of the CBD model that incorporates the cohort effect fits the English and Welsh data best, whereas for the U.S. data, the Renshaw and Haberman (2006) (RH) model, which also allows for a cohort effect, provides the best fit (Cairns et al., 2009). In addition to the cohort effect, short-term catastrophic mortality events, such as the influenza pandemic in 1918 and the Tsunami in December 2004, may lead to much higher mortality rates. Using empirical data from 1900 to 1984, we find that the residuals in the RH model for England and Wales, France, and Italy exhibit leptokurticity. It is crucial to address such mortality jumps in age–period–cohort mortality models. The main goal of this study is to incorporate non-Gaussian innovations into the RH model.

To take heavy-tailed distributions into account in stochastic mortality models, Milidonis, Lin, and Cox (2011) use a Markov regime-switching model to analyze the 1901–2005 U.S. population mortality data and price mortality securities. In contrast, Biffis (2005) employs affine jump diffusions to model asset prices and mortality dynamics and thus addresses the risk analysis and market valuation of life insurance contracts. For Italian mortality data, Luciano and Vigna (2005) demonstrate that a diffusion process with a jump component (JD) provides a better fit than does a diffusion component in stochastic mortality processes. Cox, Lin, and Wang (2006) employ the JD process to model age-adjusted mortality rates for the United States and United Kingdom and to evaluate the first pure mortality security: the Swiss Re Vita bond. In addition, Lin and Cox (2008) combine a Brownian motion and a discrete Markov chain with a log-normal jump size distribution to price mortality-based securities in an incomplete market framework. Incorporating a jump process into the LC model, Chen and Cox (2009) forecast mortality rates and analyze mortality securitization. That is, these contributions use diffusion processes with jump components, one of the finite-activity Lévy processes, to describe the dynamics of mortality rates.

Hainaut and Devolder (2008) were the first to apply α-stable subordinators—which are strictly positive Lévy processes—to model mortality rates. Giacometti, Ortobelli, and Bertocchi (2009) employ the normal inverse Gaussian (NIG) distribution to model both the error distributions of the LC model, observing that the NIG distributional assumption for the residuals of the LC model is better than the Gaussian one for certain age groups. Wang, Huang, and Liu (2011) fit the LC model with heavy-tailed distributions to mortality rates from 1900 to 1999 and demonstrate that for applications of the LC model, the heavy-tailed distributions appear to be the most appropriate choices for modeling long-term mortality data. However, as proposed by Pitacco (2004), various disadvantages arise in connection with the LC model. To improve the LC model, it is possible to model the number of deaths as a Poisson model, as commonly employed in the literature on mortality modeling (e.g., Wilmoth, 1993; Brouhns, Denuit, and Vermunt, 2002; Renshaw and Haberman, 2006; Cairns et al., 2009; Haberman and Renshaw, 2009). However, with the Poisson error structure, the intensity at age x and time t is determined by the death rate at age x and time t, which is broadly described by stochastic mortality models. Consequently, instead of using a Poisson model with a deterministic intensity function, an alternative means of fitting the number of deaths is to specify a doubly stochastic Poisson process, or Cox process (Cox, 1955), to capture the stochastic intensity. Biffis, Denuit, and Devolder (2010) first implement a doubly stochastic setup in the LC model, introducing a class of equivalent probability measures for pricing life insurance liabilities and mortality-indexed securities. Following the double stochastic setup proposed by Biffis, Denuit, and Devolder (2010), the second goal of this article is to provide an iterative fitting algorithm for estimating the Cox regression model in which mortality rates adhere to the RH model with non-Gaussian innovations.

We use three mortality data sets—England and Wales, France, and Italy—from 1900 to 2009 as the observed data. We first fit the model to the mortality rates from 1900 to 1984 using the normal, JD, variance gamma (VG), and NIG distributions, and then we forecast the development of the mortality curve for the subsequent 25 years. According to the Jarque–Bera statistical test, the assumption of normality must be rejected for the logarithm of mortality rates. Finally, according to the mean absolute percentage errors (MAPEs) of the mortality projections, our empirical results indicate that the RH model with non-Gaussian innovations is the most appropriate choice for modeling long-term mortality data. In addition, as an application for England and Wales, we provide the fair values of longevity swaps and their value at risk (VaR) and conditional tail expectations (CTE). According to the RH model with non-Gaussian innovations, the swap premiums are lower, but the VaR and CTE are higher, which means that the lower price of the hedge is not only based on the lower swap curves implied by the best prediction model, but also in terms of the fatter tails of the unexpected losses it generates.

The remainder of this article is organized as follows. In the second section, we provide an iterative fitting algorithm to generate the maximum likelihood estimates of the Cox regression model in which the residuals of the RH model, the mortality indices, and the cohort effects adhere to heavy-tailed distributions. Third section empirically tests the goodness of fit of stochastic mortality models with JD, VG, and NIG distributions; it also offers mortality projections. In the fourth section, we employ the RH model with non-Gaussian innovations to price a longevity swap and calculate its VaR and CTE using England and Wales mortality data. The last section draws some conclusions about our findings.

STOCHASTIC MORTALITY MODELS WITH COX ERROR STRUCTURES

In this section, we first review the RH model, in which the mortality index and cohort effect follow ARIMA models with normal innovations. However, according to the mortality data, the residuals exhibit non-Gaussian distribution. Consequently, we assume that the number of deaths follows a Cox process and that the death rates adhere to the RH model in which the residuals, the mortality indices, and the cohort effects follow three non-Gaussian distributions: JD, VG, and NIG. We also develop an iterative process for calibrating the corresponding parameters of the Cox process with leptokurtic intensity.

RH Model

We analyze changes in mortality as a function of both age x and time t. For mortality forecasting, the cohort-based extension to the LC model proposed by Renshaw and Haberman (2006) is as follows:

$$\ln(m_{x,t}) = \alpha_x + \beta_x k_t + \eta_x \gamma_{t-x} + e_{x,t}, \tag{1}$$

where $m_{x,t}$ is the death rate for age x in calendar year t, defined as running from time t to time $t + 1$; α_x describes the average pattern of mortality over an age group; k_t explains the time trend of the general mortality level; β_x represents age-specific patterns of mortality change, indicating the sensitivity of the logarithm of the force of mortality at age x to variations in k_t; γ_{t-x} is a cohort effect; η_x controls age-specific cohort contributions to the mortality projection; and $e_{x,t}$ represents the error term, which is normally distributed with mean 0 and variance σ_e^2. This structure is designed to capture age–period–cohort effects.

To forecast future mortality dynamics, the mortality index k_t follows a one-dimensional random walk with drift (Lee and Carter, 1992), as follows:

$$k_t - k_{t-1} = \mu + \varepsilon_t, \tag{2}$$

where μ is a drift term and ε_t is a sequence of independent and identically zero-mean Gaussian random variables. Let the year of birth be equal to $c = t - x$. Following the model setup of Renshaw and Haberman (2006) and Cairns et al. (2010), we model the cohort factor γ_c as an ARIMA(1,1,0) process that is independent of k_t:

$$\Delta\gamma_c = \mu_\gamma + \alpha_\gamma(\Delta\gamma_{c-1} - \mu_\gamma) + \sigma_\gamma z_c, \tag{3}$$

where z_c is a sequence of independent and identically standard normal random variables.

Normality Test for the RH Model

According to Table 1 in Dowd et al.'s (2010) article, the residuals of the RH model exhibit leptokurticity. In this subsection, we therefore apply the JB statistic (Jarque and Bera, 1980) to test empirically the normality of the three mortality data sets from

TABLE 1
The Jarque–Bera Test

	England and Wales	France	Italy
Residuals of the RH model	388.341 [<0.001]	1572.824 [<0.001]	15984.422 [<0.001]
First difference in mortality indices	1.161 [0.484]	0.767 [0.500]	2.180 [0.233]
Residuals of cohort effects	370.387 [<0.001]	41.441 [<0.001]	323.478 [<0.001]

Note: The *p*-values of the Jarque–Bera test are in brackets.

England and Wales, France, and Italy for subjects aged 60–89 years during the period 1900–1984. The mortality data came from the Human Mortality Database website.[1] Table 1 contains the results of the JB test for the residuals of the RH model, the first difference of the three countries' mortality indices, and the corresponding cohort effects from 1900 to 1984. The JB statistic rejects the assumption of normality for the residuals of the RH model and the cohort effects. Therefore, we use the heavy-tailed distributions—JD, VG and NIG—to model the non-Gaussian nature of the error terms of the RH model.

Heavy-Tailed Distributions

We model the error terms of the RH model, $e_{x,t}$, ε_t, and z_{t-x}, using the three heavy-tailed distributions: JD, VG, and NIG. In the subsequent subsection, we take $e_{x,t}$ as an example to describe the properties of these heavy-tailed distributions; analogous results are obtained for ε_t, and z_{t-x}. If $e_{x,t}$ adheres to a JD distribution, then

$$e_{x,t} = -\lambda_N \mu_Y + \sigma z + \sum_{i=1}^{N} Y_i, \tag{4}$$

where N is the Poisson distribution with intensity λ_N, z is a standard normal random variable, and each Y_i, independent of z and N, is a normal distribution with mean μ_Y and variance δ_Y^2. The setup in Equation (4) satisfies $E(e_{x,t}) = 0$ and $V(e_{x,t}) = \sigma^2 + \lambda_N(\mu_Y^2 + \delta_Y^2)$. The probability density function (pdf) of $e_{x,t}$ is of the form:

$$f_{e_{x,t}}^{\mathrm{JD}}(y|\sigma, \lambda_N, \mu_Y, \delta_Y) = \sum_{n=0}^{\infty} \frac{\lambda_N^n \, \mathrm{e}^{-\lambda_N}}{n!} \Phi(y|(n-\lambda_N)\mu_Y, \sigma^2 + n\delta_Y^2), \tag{5}$$

where $\Phi(y|\tilde{\mu}, \tilde{\sigma}^2)$ is the normal pdf evaluated at y with mean $\tilde{\mu}$ and variance $\tilde{\sigma}^2$. The moment-generating function of the JD distribution is

[1]See http://www.mortality.org/.

$$M_{e_{x,t}}^{\mathrm{JD}}(u) = E(\exp(ue_{x,t})) = \exp[-u\lambda_N\mu_Y + 0.5u^2\sigma^2 + \lambda_N(\mathrm{e}^{u\mu_Y+0.5u^2\delta_Y^2} - 1)]. \tag{6}$$

When the residuals follow an NIG distribution, the pdf of $e_{x,t}$ instead is of the form:

$$f_{e_{x,t}}^{\mathrm{NIG}}(y|\alpha,\beta,\delta,\theta) = \frac{\alpha\delta}{\pi}\exp\left(\delta\sqrt{\alpha^2-\beta^2} + \beta(y-\theta)\right)\frac{K_1\left(\alpha\sqrt{\delta^2+(y-\theta)^2}\right)}{\sqrt{\delta^2+(y-\theta)^2}}, \tag{7}$$

where K_γ is the modified Bessel function of the second kind with index γ, δ is the scale parameter, θ is the shift parameter, and α and β determine the shape of the NIG distribution. To ensure $E(e_{x,t}) = 0$, we have $\theta = -\beta\delta/\sqrt{\alpha^2-\beta^2}$. The parameters must fulfill two constraints: $\delta \geq 0$ and $\alpha > |\beta|$. The moment-generating function of the NIG distribution is

$$M_{e_{x,t}}^{\mathrm{NIG}}(u) = E(\exp(ue_{x,t})) = \exp\left(-\frac{u\beta\delta}{\sqrt{\alpha^2-\beta^2}} + \delta\left(\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+u)^2}\right)\right). \tag{8}$$

The NIG distribution is one of the most promising distributions for asset returns proposed in the prior literature, with several attractive theoretical properties and analytical tractability. It therefore has been used repeatedly for financial applications as the unconditional return distribution (Eberlein and Keller, 1995; Prause, 1997; Rydberg, 1997; Bølviken and Benth, 2000; Lillestøl, 2000) and for stochastic mortality modeling (Giacometti, Ortobelli, and Bertocchi, 2009; Wang, Huang, and Liu, 2011).

When the residuals follow a VG distribution, the pdf of $e_{x,t}$ is of the form:

$$f_{e_{x,t}}^{\mathrm{VG}}(y|\alpha,\beta,\gamma,\theta) = \frac{(\alpha^2-\beta^2)^\gamma|y-\theta|^{\gamma-0.5}K_{\gamma-0.5}(\alpha|y-\theta|)}{\sqrt{\pi}(2\alpha)^{\gamma-0.5}\Gamma(\gamma)}\exp(\beta(y-\theta)). \tag{9}$$

Similarly, θ is the shift parameter that satisfied $E(e_{x,t}) = 0$, and γ, α, and β determine the shape of the VG distribution. The parameters must fulfill the following constraints: $\gamma \geq 0$ and $\alpha > |\beta|$. The moment-generating function of the VG distribution is given by

$$M_{e_{x,t}}^{\mathrm{VG}}(u) = E(\exp(ue_{x,t})) = \exp\left(-\frac{2u\beta\gamma}{\alpha^2-\beta^2}\right)\left(\frac{\alpha^2-\beta^2}{\alpha^2-(\beta+u)^2}\right)^\gamma. \tag{10}$$

Note that when $\alpha = (G+M)/2$, $\beta = (G-M)/2$, and $\gamma = C$, we obtain the VG distribution, which is a special case of the CGMY distribution defined by Carr et al. (2002).

A Cox Process With Leptokurtic Intensity

We assume that $D_{x,t}$, or the number of deaths at age x during year t, adheres to a Cox process, also known as a doubly stochastic Poisson process. That is, $D_{x,t} \sim \text{Cox}(\lambda_{x,t})$, where $\lambda_{x,t} = E_{x,t}m_{x,t}$ is a nonnegative stochastic intensity process, and $E_{x,t}$ is the exposure to risk at age x during year t. When death rates adhere to the RH model, $\lambda_{x,t}$ can be modeled as

$$\lambda_{x,t} = E_{x,t}m_{x,t} = E_{x,t}\,\exp(\alpha_x + \beta_x k_t + \eta_x \gamma_{t-x} + e_{x,t}), \tag{11}$$

where $e_{x,t}$ is assumed to be an age- and period-homogeneous heavy-tailed distribution that captures leptokurticity. Let $d_{x,t}$ be the corresponding number of deaths actually observed. Conditional on $e_{x,t} = y$, the number of deaths $D_{x,t}$ becomes a Poisson distribution with intensity $E_{xt}\,\exp(\alpha_x + \beta_x k_t + \eta_x \gamma_{t-x} + y)$. As a result, the log-likelihood function based on the Cox regression model is defined as

$$\text{LLF} = \sum_{x,t} \int_{-\infty}^{\infty} \log f(D_{x,t} = d_{x,t} | e_{x,t} = y) f_{e_{x,t}}(y)\,\mathrm{d}y, \tag{12}$$

where

$$\begin{aligned} &\log f(D_{x,t} = d_{x,t}|e_{x,t} = y) \\ &\qquad = d_{x,t}\,\log(E_{x,t}\,\exp(\alpha_x + \beta_x k_t + \eta_x\gamma_{t-x} + y)) - E_{x,t}\,\exp(\alpha_x + \beta_x k_t \\ &\qquad\quad + \eta_x\gamma_{t-x} + y) - \log(d_{x,t}!). \end{aligned} \tag{13}$$

Thus, to find the maximum likelihood estimates of the parameters of the RH model, we can maximize Equation (12) with respect to α_x, β_x, k_t, η_x, γ_{t-x}, and the error term distribution parameters. The closed-form solution of the log-likelihood function in Equation (12) is derived as follows:

$$\text{LLF} = \sum_{x.t} [d_{x,t}(\alpha_x + \beta_x k_t + \eta_x\gamma_{t-x}) - (E_{x,t}\,\exp(\alpha_x + \beta_x k_t + \eta_x\gamma_{t-x}))M_{e_{x,t}}(1)] + C, \tag{14}$$

where $M_{e_{x,t}}(u)$ is the moment-generating function of $e_{x,t}$, and C represents a constant term equal to $\sum_{x.t}[d_{x,t}\,\log\,E_{x,t} - \log(d_{x,t}!)]$. The proof of Equation (14) is in Appendix A. Note that when $e_{x,t}$ is ignored while modeling the number of deaths (i.e., $M_{e_{x,t}}(1) = 1$), the log-likelihood function defined in Equation (14) is precisely the same as that proposed by Wilmoth (1993), Brouhns, Denuit, and Vermunt (2002), and Cairns et al. (2009).

Similar to the two-step procedures of Lee and Carter (1992), Brouhns, Denuit, and Vermunt (2002), and Renshaw and Haberman (2006), we first calibrate the parameters α_x, β_x, k_t, η_x, and γ_{t-x} with an updating scheme. Then, we estimate μ, μ_γ, α_γ, σ_γ, and the distribution parameters of the residuals of the mortality indices and cohort effects. In the first step, there are six sets of parameters, namely, the α_x, β_x, k_t, η_x, and γ_{t-x} parameters, as well as the $e_{x,t}$ distribution parameters. Following Brouhns, Denuit, and Vermunt (2002) and Renshaw and Haberman (2006), we use the following updating scheme: Let n_x be the total number of ages. Starting with $\eta_x = 1/n_x$ and $\gamma_{t-x} = 0$ and then obtaining α_x, β_x, and k_t from the approximation method of the LC model, we calibrate the corresponding $e_{x,t}$ distribution parameters by maximizing the sample log-likelihood function:

$$\text{LLF}_{e_{x,t}} = \sum_{x,t} \log(f(e_{xt}|\alpha_x, \eta_x, \gamma_{t-x}, \beta_x, k_t)). \tag{15}$$

Then, with the $e_{x,t}$ distribution parameters estimated from Equation (15), we employ an iterating method to estimate the corresponding parameters of the RH model according to the elementary Newton method (Goodman, 1979; Brouhns, Denuit, and Vermunt, 2002; Renshaw and Haberman, 2006).

Following the estimating procedure of Renshaw and Haberman (2006), the parameters are estimated by iteration. In each iteration step, we update a single set of parameters; the other parameters are fixed at their current estimates using the following updating scheme:

$$\text{update}(\theta) = u(\theta) = \theta - \frac{\partial LLF/\partial\theta}{\partial^2 LLF/\partial\theta^2}. \tag{16}$$

Consequently, the updating scheme is as follows:

$$u(\alpha_x) = \alpha_x + \frac{\sum_t [d_{x,t} - E_{x,t}\exp(\alpha_x + \beta_x k_t + \eta_x\gamma_{t-x})M_{e_{x,t}}(1)]}{\sum_t [E_{x,t}\exp(\alpha_x + \beta_x k_t + \eta_x\gamma_{t-x})M_{e_{x,t}}(1)]}, \tag{17}$$

$$u(\gamma_z) = \gamma_z + \frac{\sum_{\substack{x,t \\ z=t-x}} [d_{x,t}\eta_x - E_{x,t}\eta_x\exp(\alpha_x + \beta_x k_t + \eta_x\gamma_z)M_{e_{x,t}}(1)]}{\sum_{\substack{x,t \\ z=t-x}} [E_{x,t}\eta_x^2\exp(\alpha_x + \beta_x k_t + \eta_x\gamma_z)M_{e_{x,t}}(1)]}, \tag{18}$$

$$u(\eta_x) = \eta_x + \frac{\sum_t [d_{x,t}\gamma_{t-x} - E_{x,t}\gamma_{t-x}\exp(\alpha_x + \beta_x k_t + \eta_x\gamma_{t-x})M_{e_{x,t}}(1)]}{\sum_t [E_{x,t}\gamma_{t-x}^2\exp(\alpha_x + \beta_x k_t + \eta_x\gamma_{t-x})M_{e_{x,t}}(1)]}, \tag{19}$$

$$u(k_t) = k_t + \frac{\sum_x [d_{x,t}\beta_x - E_{x,t}\beta_x\exp(\alpha_x + \beta_x k_t + \eta_x\gamma_{t-x})M_{e_{x,t}}(1)]}{\sum_x [E_{x,t}\beta_x^2\exp(\alpha_x + \beta_x k_t + \eta_x\gamma_{t-x})M_{e_{x,t}}(1)]}, \tag{20}$$

and

$$u(\beta_x) = \beta_x + \frac{\sum_t [d_{x,t}k_t - E_{x,t}k_t\exp(\alpha_x + \beta_x k_t + \eta_x\gamma_{t-x})M_{e_{x,t}}(1)]}{\sum_t [E_{x,t}k_t^2\exp(\alpha_x + \beta_x k_t + \eta_x\gamma_{t-x})M_{e_{x,t}}(1)]}. \tag{21}$$

We repeat the updating cycle (Equations (15) to (21)) and stop when the change in the log-likelihood function in Equation (14) is relatively small.[2] Model identification can be conveniently achieved with parameter constraints: $\sum_t k_t = 0$, $\sum_x \beta_x = 1$, $\sum_x \eta_x = 1$, and $\sum_t \gamma_{t-x} = 0$.[3]

After obtaining the mortality indices and cohort effects, we can calculate the parameters of Equations (2) and (3) by maximizing the log-likelihood function, as follows:

$$\sum_t \log(f(\varepsilon_t)) \quad \text{and} \quad \sum_{s=t-x} \log(f(z_s)), \tag{22}$$

where $f(\varepsilon_t)$ and $f(z_s)$ are the probability density functions of ε_t and z_s, respectively.

EMPIRICAL ANALYSIS

In this section, we investigate the goodness-of-fit distributions for the number of deaths, the first differences of the mortality indices, and the cohort effects. Using the mortality data from 1900 to 1984,[4] we fit the residuals of the RH model to four distributions: normal, JD, VG, and NIG. We then fit the mortality indices and cohort effects from the best fitting model according to the Bayesian information criterion (BIC) to the same four distributions. Finally, we project the subsequent 25-year mortality rates (1985–2009).

In-Sample Goodness of Fit

Using mortality data from 1900 to 1984, we first investigate the goodness-of-fit distributions of the number of deaths for England and Wales, France, and Italy. Table 2 presents the LLF, Akaike information criterion (AIC), and BIC statistics[5] for the number of deaths at which the residuals of the RH model adhere to the normal, JD, VG, and NIG models. All three criteria indicate that the normal distribution is the worst fitting model for the number of deaths. They also indicate that the VG model is consistently the best model for the number of deaths in the three mortality data sets.

---

[2]We adopt 10^{-7} as the default value.

[3]Similar to Brouhns, Denuit, and Vermunt (2002), after updating the k_t parameters, we impose a centering constraint $\sum_t k_t = 0$ by removing $\sum_t k_t$ from k_t. After updating the β_x parameters, a scaling constraint $\sum_x \beta_x = 1$ must be imposed by dividing the estimates for β_x by $\sum_x \beta_x$ and multiplying the estimates for k_t by the same number. Following the analogical procedure, the constraints of η_x and γ_{t-x} are also achieved.

[4]To account mortality jumps—which might be caused by 1918 influenza pandemic, wars, or natural catastrophes such as tsunamis, Lin and Cox (2008) and Chen and Cox (2009) employ U.S. mortality data starting from 1900 to analyze mortality securitization. Along this line, we also use mortality data from 1900 to 2009 for in-sample goodness of fit and mortality projection.

[5]AIC $= -\text{LLF} + \text{NP}$ and BIC $= -\text{LLF} + 0.5 \times \text{NP} \times \log(\text{NS})$, where NP is the effective number of parameters being estimated and NS is the number of observations.

TABLE 2
Goodness-of-Fit Measures for the Number of Deaths

	England and Wales			France			Italy		
Model	LLF	AIC	BIC	LLF	AIC	BIC	LLF	AIC	BIC
Normal	−32,911	33,197	34,033	−30,594	30,880	31,716	−45,472	45,758	46,593
JD	−32,889	33,178	34,022	−30,470	30,759	31,604	−44,040	44,329	45,173
VG	***−32,744***	***33,032***	***33,874***	***−30,317***	***30,605***	***31,446***	***−43,164***	***43,452***	***44,294***
NIG	−32,894	33,182	34,024	−30,476	30,764	31,605	−44,623	44,911	45,753

Note: The bold italic values represent the best models for the number of deaths.

TABLE 3
Goodness-of-Fit Tests for the First Difference in Mortality Indices

	England and Wales			France			Italy		
Model	LLF	AIC	BIC	LLF	AIC	BIC	LLF	AIC	BIC
Normal	−166.73	168.73	***171.16***	−175.85	***177.85***	***180.28***	−182.00	***184.00***	***186.43***
JD	***−163.96***	168.96	175.03	***−175.44***	180.44	186.51	***−181.02***	186.02	192.10
VG	−164.66	***168.66***	173.53	−175.60	179.60	184.46	−181.11	185.11	189.98
NIG	−165.39	169.39	174.25	−175.59	179.59	184.45	−181.12	185.12	189.98

Note: The bold italic values represent the best models for the first difference in mortality indices.

Therefore, we use the mortality indices and cohort effects obtained from the VG model to investigate the pattern of the error terms of the time and cohort effects.

The test results for the first difference in mortality indices are in Table 3, which contains the LLF, AIC, and BIC statistics for the normal, JD, VG, and NIG distributions. The Gaussian model is the worst according to the LLF criterion, which also indicates that the best fit for the three mortality data sets derives from the JD model. The best in-sample goodness of fit for mortality indices changes for the normal distribution of all mortality data, because the BIC introduces a penalty term for the effective number of parameters.

In Table 4 we present the LLF, AIC, and BIC statistics for the normal, JD, VG, and NIG distributions for cohort effects. The LLF, AIC, and BIC statistics consistently indicate that the best fit for Italy derives from the JD model, but for England and Wales and France, it derives from the VG model. All three criteria indicate that the normal distribution is the worst fitting model for the cohort effects. Consequently, with mortality data from three countries over the period 1900–1984, in-sample model selection criteria indicate a preference for modeling the RH model with non-Gaussian innovations.

TABLE 4
Goodness-of-Fit Tests for the Residuals of Cohort Effects

	England and Wales			France			Italy		
Model	LLF	AIC	BIC	LLF	AIC	BIC	LLF	AIC	BIC
Normal	−97.09	100.09	104.17	−108.57	111.57	115.65	−130.57	133.57	137.65
JD	−84.99	90.99	99.14	−98.89	104.89	113.05	***−109.45***	***115.45***	***123.60***
VG	***−83.92***	***88.92***	***95.71***	***−98.73***	***103.73***	***110.53***	−116.86	121.86	128.65
NIG	−84.07	89.07	95.86	−99.41	104.41	111.21	−114.48	119.48	126.27

Note: The bold italic values represent the best models for the residuals of cohort effects.

Mortality Projection

To assess out-of-sample performance, we apply the parameters estimated from the time period 1900–1984 to obtain 25-year mortality projections, calculating the MAPE as follows:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right|, \tag{23}$$

where A_i is the logarithm of the historical mortality rate, F_i is the natural logarithm of the forecast mortality rate, and n is the number of observations.

By applying the calibrated parameters of the RH model to the VG innovations (the best model according to BIC), we reveal the impact of the different distributions on the mortality projection for MAPE from 1985 to 2009 (Table 5). A lower value indicates better predictive power for the distribution. For comparison, we also provide the

TABLE 5
MAPE of Logarithm of Mortality Projection in 1985–2009 (Unit: %)

Model	England and Wales	France	Italy
Original RH–normal	4.6383	4.4374	6.0001
Original RH–JD	4.6515	4.4539	5.9488
Original RH–VG	4.7143	4.4348	5.9399
Original RH–NIG	4.6529	4.4550	5.9218
VG–normal	4.5255	***4.4066***	5.8332
VG–JD	4.5382	4.4222	5.7841
VG–VG	***4.5235***	4.4379	5.7752
VG–NIG	4.5392	4.4236	***5.7586***

Note: Original RH–normal is the same as M2 of Cairns et al. (2009). The X–Y model corresponds to an X error term in the RH model and to Y distributions for the time and cohort effects. The bold italic values represent the best mortality projections for the mortality data, in terms of the MAPE criterion.

mortality projection of the original RH model with four forecasting distributions—normal, JD, VG, and NIG (the original RH–normal model corresponds to the M2 model of Cairns et al., 2009). The VG–VG model[6] is the best mortality projection for the mortality data of England and Wales. The VG–normal model provides the best one for the mortality data from France, and the VG–NIG model is the best one for Italy. As a result, in terms of the MAPE criterion, the RH model with non-Gaussian innovations provides better mortality projection than that obtained from the original RH model with normal innovations.

APPLICATION: THE VALUATION OF LONGEVITY SWAPS

In this section, we first price a longevity swap. Using the mortality data of England and Wales from 1900 to 2009, we then refit the RH model to attain the fair swap premium of the longevity swap for both the original RH model (M2) and the best projection model. Finally, we provide the VaR and CTE of the longevity swaps.

Pricing Longevity Swaps

The traditional method of transferring longevity risk in a pension plan or an annuity book is to sell the liability through an insurance or reinsurance contract, known as pension buyouts. These tactics have attracted increasing attention since 2006, especially in the United Kingdom. However, such transactions involve the transfer of all risks, including longevity and investment risk. To transfer longevity risk only to capital markets, Blake and Burrows (2001) first advocate the use of longevity bonds, whose coupon payments depend on the proportion of the population surviving to particular ages. The EIB/BNP longevity bond was the first securitization instrument designed to transfer longevity risk but ultimately was withdrawn. The lack of success in issuing longevity bonds led to new securitization instruments, such as longevity swaps,[7] which first reached the public domain with a transaction between JPM and Canada Life in July 2008. As Blake et al. (2012) show, 16 publicly announced longevity swaps were executed between 2007 and 2012 in the United Kingdom. In this context, the valuation of longevity swaps represents an important research topic for developing capital market solutions for longevity risk.

Longevity swaps have been widely explored in the prior literature (Dawson, 2002; Lin and Cox, 2005; Dowd et al., 2006; Dawson et al., 2010; Biffis et al., 2011; Wang and Yang, Forthcoming). Dowd et al. (2006) introduce the mechanism for transferring longevity risk; this instrument involves exchanging actual pension payments for a series of preagreed fixed payments. On each payment date, the fixed-rate payer (e.g., pension plan) receives from the hedge supplier a random mortality-dependent payment and, in return, makes a fixed payment to the hedge supplier. Dowd et al. (2006) demonstrate that the hedge is almost perfect when the reference index is based on the survivor experience of the insurer's annuity book. If the expected reference

[6]A VG–NIG model corresponds to a VG error term in the RH model and to NIG distributions for the time and cohort effects.

[7]For the recent development of longevity-linked securities, see Blake et al. (2012) and references therein.

indices and insurers' own survivor experiences are highly correlated, the longevity swap can still hedge the insurer against a considerable amount of the aggregate longevity risk it faces. In this article, following the vanilla longevity swap structure analyzed by Dowd et al. (2006) and Dawson et al. (2010), we discuss a T-year bespoke longevity swap linked to a benchmark cohort of a given initial age for the England and Wales mortality data.[8]

For a given time horizon T, we consider a filtered probability space $(\Omega, \mathbb{F}, \{\mathcal{F}_t\}_{t=0}^T, P)$ on which the death time is modeled as a stopping time τ with respect to $\mathbb{F} = \{\mathcal{F}_t\}_{t=0}^T$. As mentioned by Biffis, Denuit, and Devolder (2010) and Hainaut (2012), $\mathbb{F}$ is the enlarged filtration $\mathbb{H} \vee \mathbb{G}$ where $\mathbb{H} = \{*_t\}_{t=0}^T$ is the filtration related to risk factors and $\mathbb{G} = \{\mathcal{G}_t\}_{t=0}^T$ is such that $\mathbb{F}$ is the minimal enlargement of $\mathbb{H}$ ensuring that τ is an $\mathbb{F}$-stopping time. Conditional on the path followed by the mortality rates, the t-year survival probability that a 65-year-old person in calendar year 2009 + t reaches age 65 + t is of the form:

$$S(t) = P(\tau > t | *_T) = \exp\left(- \int_0^t m_{65+s,2009+s} \, \mathrm{d}s \right). \tag{24}$$

We assume that the mortality rates are constant within certain age and time windows but may vary from one window to the next. Specifically, given any integer age x and calendar year t, we presume that

$$m_{x+\xi,t+\tau} = m_{x,t} \quad \text{for } 0 \leq \xi, \tau < 1. \tag{25}$$

Thus,

$$S(t) = \exp\left(- \sum_{h=0}^{t-1} m_{65+h,2009+h} \right). \tag{26}$$

To transfer longevity risk, on each of the payment dates t, the fixed-rate payer pays the notional principal multiplied by a prespecified fixed proportion $(1 + \pi)H(t)$ to the floating-rate payer and receives the notional principal multiplied by $S(t)$, where $H(t)$ is anticipated by using the best estimate of the underlying mortality model, and π is the swap premium that would be set so that the initial value of the swap is zero for each party.

The distribution function of $S(t)$ under the real-world (physical) probability measure P is

$$F_t(y) = \mathrm{Prob}_P(S(t) \leq y). \tag{27}$$

[8]To bear no basis risk, the variable payments in bespoke longevity swaps are designed to match precisely the mortality experience of each individual hedger.

Wang (2000) proposes a distortion operator to change the probability measure from the real-world probability measure P to an equivalent martingale measure Q, with the following transformation:[9]

$$\tilde{F}_t(y) = \Phi(\Phi^{-1}(F_t(y)) + \lambda(t)), \tag{28}$$

where $\lambda(t)$ may be interpreted as the market price of longevity risk associated with the survival probability $S(t)$,[10] and Φ is the standard normal distribution function. Therefore, as shown by Denuit, Devolder, and Goderniaux (2007), the expectation value of $S(t)$ associated with $\lambda(t)$ under the equivalent martingale measure Q is defined as

$$E_Q[S(t)] = \int_0^1 (1 - \tilde{F}_t(y))\ \mathrm{d}y = \int_0^1 (1 - \Phi(\Phi^{-1}(F_t(y)) + \lambda(t)))\ \mathrm{d}y. \tag{29}$$

Let M be the total annuities issued to an initial population that consists of persons aged 65 years who also are alive in 2009. Under the equivalent martingale measure Q, from the point of view of the hedger with a pay-fixed longevity swap, the fair value at issue year 2010, denoted by LS_0, can be calculated as

$$\mathrm{LS}_0 = E_Q\left[\sum_{t=1}^{T} \exp\left(-\int_0^t r(u)\ \mathrm{d}u\right) M(S(t) - (1+\pi)H(t))\right], \tag{30}$$

where $r(t)$ is the risk-free rate. We also consider the term structure of the interest rate in our valuation framework. Let $B(t, T)$ denote the price of a zero-coupon bond issued at time t that pays \$1 at time T, $t \leq T$. With the assumption that mortality rates and financial risk are independent, the fair value of a pay-fixed longevity swap takes the form:

$$\mathrm{LS}_0 = M \sum_{t=1}^{T} \mathrm{B}(0,t) E_Q[S(t)] - M(1+\pi) \sum_{t=1}^{T} \mathrm{B}(0,t) H(t). \tag{31}$$

The fair swap premium π, which is set when the initial value of the swap equals zero, is given by

[9]The Wang transform represents only one possible choice among several incomplete market pricing methods. For example, Biffis et al. (2010) provide the equivalent changes of measures that preserve the structure of the LC model and the tractability of the doubly stochastic setup. The specification of both a real-world and an equivalent martingale measure raises the issue of whether the doubly stochastic setting applies under the two measures. For more details, please refer to the Proposition 3.2 in Biffis et al. (2010).

[10]For simplicity, let $\lambda(t)$ be constant in the numerical examples.

TABLE 6
Goodness-of-Fit Measures for the Number of Deaths

	England and Wales		
Model	LLF	AIC	BIC
Period: 1900–2009			
Normal	−42042	42378	43403
JD	−41894	42233	43267
VG	***−41795***	***42133***	***43165***
NIG	−41819	42157	43188
Period: 1960–2009			
Normal	−13351	13567	14140
JD	−13338	13557	14139
VG	***−13333***	***13551***	***14131***
NIG	−13336	13554	14133

Note: The bold italic values represent the best models for the number of deaths.

$$\pi = \frac{\sum_{t=1}^{T} \mathrm{B}(0,t) E_Q[S(t)]}{\sum_{t=1}^{T} \mathrm{B}(0,t) H(t)} - 1. \tag{32}$$

The analytical computation of $E_Q[S(t)]$ is difficult to implement. We explain briefly the Monte Carlo algorithm to compute the expected value of the t-year survival probability under the equivalent martingale measure Q in Appendix B.

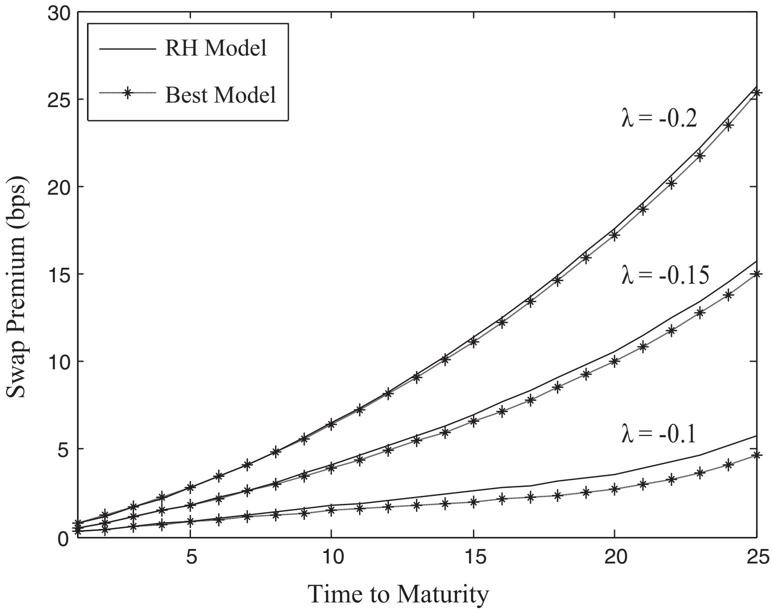
Numerical Analysis

To simulate the mortality rates, we first refit the RH model with four distributions—normal, JD, VG, and NIG—to the mortality data of England and Wales over the two periods 1900–2009 and 1960–2009 in Table 6. Similar to the results based on the 1900–1984 period, the best model for England and Wales is still the VG model. Consequently, applying the calibrated parameters over the period 1900–2009, we use the best prediction models presented in Table 5 to simulate mortality rates, which is the VG–VG model for England and Wales.

In this section, we provide a numerical example of the longevity swaps based on a cohort of 65-year-old persons in calendar year 2009. The initial term structure is obtained from the U.S. Department of the Treasury.[11] We also assume that $M = 1$. Figure 1 depicts the swap premium curve by varying the level of the risk-adjustment

[11]See http://www.treasury.gov/resource-center/data-chart-center/interest-rates/pages/TextView.aspx?data=yieldYear&year=2009. The 1-year, 2-year, 3-year, 5-year, 7-year, 10-year, 20-year, and 30-year yield rates are 0.47, 1.14, 1.7, 2.69, 3.39, 3.85, 4.58, and 4.63 percent on December 31, 2009, respectively. We use the linear interpolation to obtain other yield rates.

FIGURE 1
Swap Premium Curves for Distinct Level of Risk-Adjusted Parameter λ



parameter λ. The fair swap premium is higher for a longer duration swap, because long-duration contracts are usually more expensive for covering longevity risk. In addition, the fair swap premiums of the RH model (the M2 model of Cairns et al., 2009) are higher than those of the best prediction model.

Table 7 reveals the fair swap premiums with time to maturity equal to 25 years when λ is −0.1, −0.15, and −0.2, with parallel shifts upward of 0, 2, and 4 percent in the yield curve. From Table 7, we see that the lower the λ and the interest rates are, the higher is the fair swap premium. Similarly, the fair swap premiums of the RH model are higher than those of the best prediction model, even when the yield curve moves up in parallel.

TABLE 7
Swap Premiums for Different Interest Rates (Units: bps)

Yield Rates	Model	$\lambda = -0.1$	$\lambda = -0.15$	$\lambda = -0.2$
Original yield curve	RH	5.75	15.77	25.74
	Best	4.63	15.01	25.31
Parallel shift up of 2%	RH	4.92	13.49	22.01
	Best	3.95	12.82	21.63
Parallel shift up of 4%	RH	4.22	11.53	18.80
	Best	3.38	10.95	18.47

Note: Time to maturity is 25 years.

TABLE 8
The MTM Values of Longevity Swaps

		λ		
Model	Yield Rates	−0.1	−0.15	−0.2
RH	Original yield curve	0	0.0134	0.0267
	Parallel shift up of 2%	−0.0009	0.0086	0.0181
	Parallel shift up of 4%	−0.0014	0.0055	0.0123
Best	Original yield curve	0	0.0139	0.0276
	Parallel shift up of 2%	−0.0008	0.0091	0.0189
	Parallel shift up of 4%	−0.0012	0.0060	0.0130

Note: Assume that λ is −0.1 and maturation time is 25 years in the baseline case.

As market conditions change (e.g., mortality patterns, a parallel shift in yield curve), the marking-to-market (MTM) procedure could mean that the longevity swap switches status in the hedger's balance sheet between that of an asset and that of a liability. Assume that λ is −0.1 and the maturation time is 25 years, as in our baseline case. The initial swap premiums are 5.75 and 4.63 bps for the RH and best prediction models in the baseline case, respectively. In Table 8, applying Equation (31), we report the impacts of market condition changes (a parallel shift in yield curve and different risk-adjustment parameters λ) on the MTM profits or losses of the longevity swaps. When the yield curve moves up in parallel, *ceteris paribus*, the fair value of the longevity swap decreases, which means that a parallel shift up in the yield curve leads to a loss for the hedger. In addition, a lower level of the risk-adjustment parameter results in a higher expected value of survival probability (higher mortality improvement), which in turn leads to a higher value of the longevity swap. Note that, as shown in Table 8, the risk-adjustment parameter has a larger impact than the parallel shift up in the yield curve on the fair value of the longevity swap. Consequently, as life expectancy increases dramatically in developed countries, it is reasonable to find the recent surge in transactions in longevity swaps.

From the point of view of the hedger, the unexpected loss at time t is of the form:

$$L(t) = M((1+\pi)H(t) - S(t)), \quad t = 1, \ldots, T. \tag{33}$$

The present value of the total unexpected loss, denoted as PVL, is given by

$$\text{PVL} = \sum_{t=1}^{T} B(0,t)L(t). \tag{34}$$

Figure 2 depicts the pdf of PVL for the RH model and the best prediction model of England and Wales mortality data; it also marks the areas for the other three subplots in the upper left-hand panel. We find that the pdf of PVL for the best prediction model possesses leptokurticity. In addition, Table 9 presents the VaR and CTE of the PVL with maturation times of up to 25 years. It is clear that, compared to the RH model, the

FIGURE 2
Probability Density Functions of Present Value of the Losses ($\lambda = -0.1$, $T = 25$)

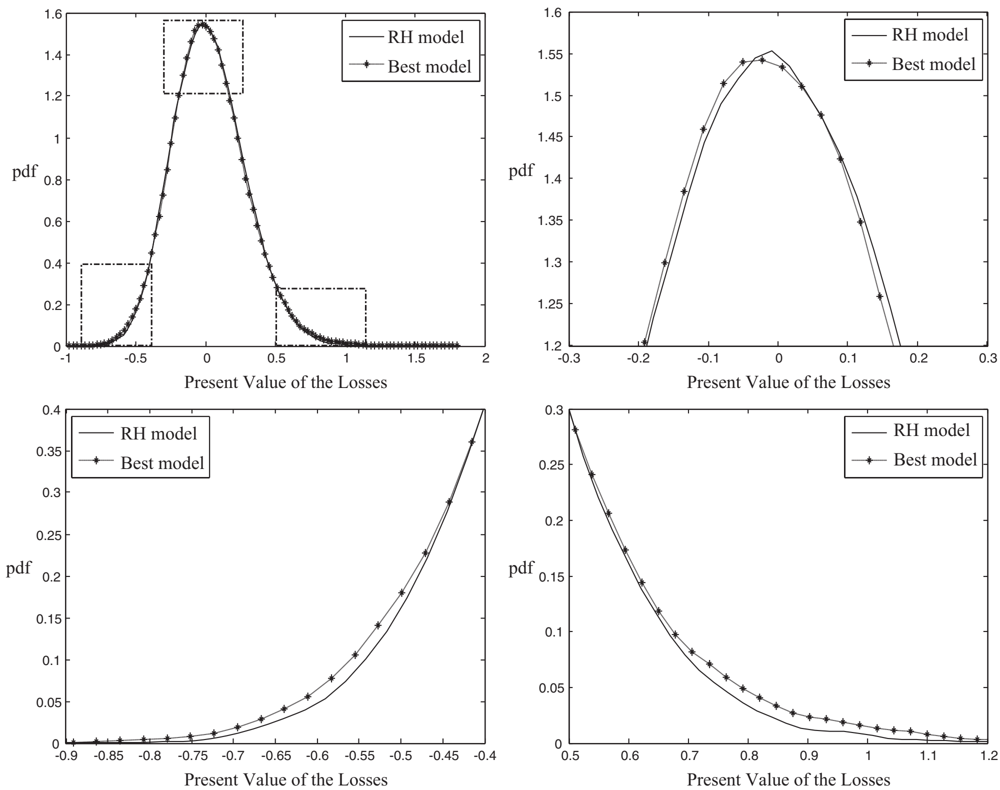


TABLE 9
The VaR and CTE of the Losses for Different Maturation Times ($\lambda = -0.1$)

Time to Maturity	Model	VaR95	VaR99	CTE95	CTE99
10	RH	0.0685	0.0999	0.0878	0.1163
	Best	0.0715	0.1126	0.0970	0.1377
15	RH	0.1679	0.2427	0.2146	0.2859
	Best	0.1744	0.2708	0.2344	0.3298
20	RH	0.3094	0.4477	0.3959	0.5266
	Best	0.3225	0.4942	0.4279	0.5981
25	RH	0.4761	0.6861	0.6063	0.8021
	Best	0.4939	0.7459	0.6504	0.8980

best prediction model has higher VaR and CTE. Because shorter-duration contracts cover less longevity risk, the VaR and CTE values are smaller for shorter duration longevity swaps. The differences of RH and the best prediction model are larger for longer durations. Therefore, the loss distribution of longevity swaps is centralized and heavy tailed, which leads to lower price of the hedge but fatter tails of the unexpected losses. It is critical to have a good mortality model to calculate accurate loss distributions.

CONCLUSIONS AND SUGGESTIONS

Many researchers have examined mortality rates and explored various models. Some studies have demonstrated that improvements in the LC model occur when the model is adjusted by fitting the Poisson regression model to the number of deaths and considering an age–period–cohort extension of the LC model. Under the Poisson error structure though, intensity consists of the death rate, which is commonly modeled by stochastic mortality models. In addition, empirical results demonstrate that mortality rate improvements exhibit jump properties. We therefore attempt to provide an iterative fitting algorithm for estimating the Cox regression model, under which death rates adhere to the RH model with three heavy-tailed distributions—JD, VG, and NIG.

Using three mortality data sets from England and Wales, France, and Italy, we find consistent support for the non-Gaussian residuals of the RH model. Specifically, when we calibrate the parameters of the RH model, the VG model provides the best fit for the three countries according to the BIC criterion. For mortality projection from the three mortality data sets, we find that the normal distribution provides weak mortality projection performance, whereas the non-Gaussian distributions provide good mortality projections. In the longevity swap application, we demonstrate that the swap curves of the original RH model are higher than those of the RH model with non-Gaussian innovations. In addition, the VaR and CTE of the original RH model are lower than those of the RH model with non-Gaussian innovations. The lower hedge cost according to the RH model with non-Gaussian innovations is not only due to the lower swap curves, but also in terms of the fatter tails of the unexpected losses it generates. As a result, choosing an appropriate leptokurtic model is critical to mortality projection and securitization of longevity risk.

APPENDIX A

THE PROOF OF THE LOG-LIKELIHOOD FUNCTION

When the death rates follow the RH model, the explicit solution of the log-likelihood function in Equation (12) can be rewritten as follows:

$$\begin{aligned}\text{LLF} &= \sum_{x,t} \int_{-\infty}^{\infty} \log f\left(D_{x,t} = d_{x,t} | e_{x,t} = y\right) f_{e_{x,t}}(y)\ \mathrm{d}y \\ &= \sum_{x.t} \Big(\int_{-\infty}^{\infty} d_{x,t} \Big(\log\ E_{x,t} + \alpha_x + \beta_x k_t + \eta_x \gamma_{t-x} + y\Big) f_{e_{x,t}}(y)\ \mathrm{d}y \\ &\quad - \int_{-\infty}^{\infty} E_{x,t}\ \exp(\alpha_x + \beta_x k_t + \eta_x \gamma_{t-x})\ \exp(y) f_{e_{x,t}}(y)\ \mathrm{d}y - \log(d_{x,t}!). \end{aligned} \tag{A1}$$

Because $E(e_{x,t}) = 0$, we have

$$\begin{aligned}\text{LLF} &= \sum_{x,t} \big[d_{x,t}(\log E_{x,t} + \alpha_x + \beta_x k_t + \eta_x \gamma_{t-x}) \\ &\quad - E_{x,t} \exp(\alpha_x + \beta_x k_t + \eta_x \gamma_{t-x}) \int_{-\infty}^{\infty} \exp(y) f_{e_{x,t}}(y)\, dy - \log(d_{x,t}!)\Big] \\ &= \sum_{x.t} \big[d_{x,t}(\alpha_x + \beta_x k_t + \eta_x \gamma_{t-x}) - (E_{x,t} \exp(\alpha_x + \beta_x k_t + \eta_x \gamma_{t-x})) M_{e_{x,t}}(1)\big] \\ &\quad + \sum_{x.t} \big[d_{x,t} \log E_{x,t} - \log(d_{x,t}!)\big]. \end{aligned} \quad \text{(A2)}$$

This completes the proof of Equation (14).

APPENDIX B

EXPECTED VALUE OF SURVIVAL PROBABILITY UNDER Q

The procedure of computing the expected value of the t-year survival probability under the equivalent martingale measure Q is as follows:

Step 1: After calibrating the parameters of the RH model, we use a Monte Carlo simulation with N iterations to generate the futures mortality rates and the survival probabilities under the real-world probability measure P. According to the N simulated values of t-year survival probabilities, we can construct the corresponding empirical cumulative distribution function (cdf) $F_t(\cdot)$ and its inverse cdf $F_t^{-1}(\cdot)$ under P.

Step 2: We know that the probability-integral transform of a random variable is distributed as standard uniform. Consequently, we have, according to Equation (28),

$$\tilde{F}_t(S(t)) = U = \Phi(\Phi^{-1}(F_t(S(t))) + \lambda), \quad \text{(B1)}$$

where U is a standard uniform random variable. Rearranging Equation (B1) and drawing N random numbers from a standard uniform distribution, we can generate N possible values of the t-year survival probabilities under the equivalent martingale measure Q, as follows:

$$S(t) = F_t^{-1}(\Phi(\Phi^{-1}(U) - \lambda)). \quad \text{(B2)}$$

Averaging the N values of the t-year survival probabilities produces the expected value of t-year survival probability under Q. A higher value of N leads to a more precise setup for $F_t(\cdot)$, $F_t^{-1}(\cdot)$, and $E_Q[S(t)]$. We use $N = 100{,}000$.

REFERENCES

Biffis, E., 2005, Affine Processes for Dynamic Mortality and Actuarial Valuations, *Insurance: Mathematics and Economics*, 37: 443-468.

Biffis, E., D. Blake, L. Pitotti, and A. Sun, 2011, The Cost of Counterparty Risk and Collateralization in Longevity Swaps, Pensions Institute Discussion Paper PI-1107, June.

Biffis, E., M. Denuit, and P. Devolder, 2010, Stochastic Mortality Under Measure Changes, *Scandinavian Actuarial Journal*, 4: 284-311.

Blake, D., and W. Burrows, 2001, Survivor Bonds: Helping to Hedge Mortality Risk, *Journal of Risk and Insurance*, 68: 339-348.

Blake, D., A. J. G. Cairns, G. Coughlan, K. Dowd, and R. MacMinn, 2012, The New Life Market, Discussion Paper.

Bølviken, E., and F. E. Benth, 2000, Quantification of Risk in Norwegian Stocks via the Normal Inverse Gaussian Distribution, Proceedings of the AFIA 2000 Colloquium, Tromsø, Norway, pp. 87–98.

Brouhns, N., M. Denuit, and J. K. Vermunt, 2002, A Poisson Log-Bilinear Regression Approach to the Construction of Projected Life Tables, *Insurance: Mathematics and Economics*, 31: 373-393.

Cairns, A. J. G., D. Blake, and K. Dowd, 2006, A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration, *Journal of Risk and Insurance*, 73: 687-718.

Cairns, A. J. G., D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, and M. Khalaf-Allah, 2010, A Framework for Forecasting Mortality Rates with an Application to Six Stochastic Mortality Models, Pensions Institute Discussion Paper PI-0801, March.

Cairns, A. J. G., D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, A. Ong, and I. Balevich, 2009, A Quantitative Comparison of Stochastic Mortality Models Using Data From England and Wales and the United States, *North American Actuarial Journal*, 13: 1-35.

Carr, P., H. Geman, D. Madan, and M. Yor, 2002, The Fine Structure of Asset Returns: An Empirical Investigation, *Journal of Business*, 75: 305-332.

Chen, H., and S. H. Cox, 2009, Modeling Mortality With Jumps: Applications to Mortality Securitization, *Journal of Risk and Insurance*, 76: 727-751.

Cox, D. R., 1955, Some Statistical Methods Connected With Series of Events (With Discussion), *Journal of the Royal Statistical Society, Series B*, 17: 129-164.

Cox, S. H., Y. Lin, and S. S. Wang, 2006, Multivariate Exponential Tilting and Pricing Implications for Mortality Securitization, *Journal of Risk and Insurance*, 73: 719-736.

Dawson, P., 2002, Mortality Swaps, Mimeo, Cass Business School.

Dawson, P., D. Blake, A. J. G. Cairns, and K. Dowd, 2010, Survivor Derivatives: A Consistent Pricing Framework, *Journal of Risk and Insurance*, 77: 579-596.

Denuit, M., P. Devolder, and A. C. Goderniaux, 2007, Securitization of Longevity Risk: Pricing Survivor Bonds With Wang Transform in the Lee-Carter Framework, *Journal of Risk and Insurance*, 74: 87-113.

Dowd, K., D. Blake, A. J. G. Cairns, and P. Dawson, 2006, Survivor Swaps, *Journal of Risk and Insurance*, 73: 1-17.

Dowd, K., A. J. G. Cairns, D. Black, G. D. Coughlan, D. Epstein, and M. Khalaf-Allah, 2010, Evaluating the Goodness of Fit of Stochastic Mortality Models, *Insurance: Mathematics and Economics*, 47: 255-265.

Eberlein, E., and U. Keller, 1995, Hyperbolic Distributions in Finance, *Bernoulli*, 1: 281-299.

Giacometti, R., S. Ortobelli, and M. I. Bertocchi, 2009, Impact of Different Distributional Assumptions in Forecasting Italian Mortality Rates, *Investment Management and Financial Innovations*, 6: 186-193.

Goodman, L. A., 1979, Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories, *Journal of the American Statistical Association*, 74: 537-552.

Haberman, S., and A. E. Renshaw, 2009, On Age-Period-Cohort Parametric Mortality Rate Projections, *Insurance: Mathematics and Economics*, 45: 255-270.

Hainaut, D., 2012, Multidimensional Lee-Carter Model With Switching Mortality Processes, *Insurance: Mathematics and Economics*, 50: 236-246.

Hainaut, D., and P. Devolder, 2008, Mortality Modelling With Lévy Processes, *Insurance: Mathematics and Economics*, 42: 409-418.

Jarque, C. M., and A. K. Bera, 1980, Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals, *Economic Letters*, 6: 255-259.

Lee, R. D., and L. R. Carter, 1992, Modeling and Forecasting U.S. Mortality, *Journal of the American Statistical Association*, 87: 659-675.

Li, S. H., and W. S. Chan, 2007, The Lee-Carter Model for Forecasting Mortality, Revisited, *North American Actuarial Journal*, 11: 68-89.

Lillestøl, J., 2000, Risk Analysis and the NIG Distribution, *Journal of Risk*, 2: 41-56.

Lin, Y., and S. H. Cox, 2005, Securitization of Mortality Risks in Life Annuities, *Journal of Risk and Insurance*, 72: 227-252.

Lin, Y., and S. H. Cox, 2008, Securitization of Catastrophe Mortality Risks, *Insurance: Mathematics and Economics*, 42: 628-637.

Loeys, J., N. Panigirtzoglou, and R. Ribeiro, 2007, *Longevity: A Market in the Making* (London: J.P. Morgan Research Publication).

Luciano, E., and E. Vigna, 2005, Non Mean Reverting Affine Processes for Stochastic Mortality, ICER Applied Mathematics Working Paper No. 4. Available at SSRN: http://ssrn.com/abstract=724706.

Milidonis, A., Y. Lin, and S. H. Cox, 2011, Mortality Regimes and Pricing, *North American Actuarial Journal*, 15: 266-289.

Pitacco, E., 2004, Survival Models in Dynamic Context: A Survey, *Insurance: Mathematics and Economics*, 35: 279-298.

Prause, K., 1997, Modelling Financial Data Using Generalized Hyperbolic Distributions, FDM Preprint 48, University of Freiburg.

Renshaw, A. E., and S. Haberman, 2003, Lee-Carter Mortality Forecasting With Age-Specific Enhancement, *Insurance: Mathematics and Economics*, 33: 255-272.

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

Rydberg, T. H., 1997, The Normal Inverse Gaussian Lévy Process: Simulation and Approximation, *Communications in Statistics: Stochastic Models*, 13: 887-910.

Wang, C. W., H. C. Huang, and I. C. Liu, 2011, A Quantitative Comparison of the Lee-Carter Model Under Different Types of Non-Gaussian Innovations, *Geneva Papers on Risk and Insurance—Issues and Practice*, 36: 675-696.

Wang, C. W., and S. S. Yang, Forthcoming, Pricing Survivor Derivatives with Cohort Mortality Dependence Under the Lee-Carter Framework, *Journal of Risk and Insurance*.

Wang, S. S., 2000, A Class of Distortion Operations for Pricing Financial and Insurance Risks, *Journal of Risk and Insurance*, 67: 15-36.

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