SOUTH AFRICAN ACTUARIAL JOURNAL
SAAJ **22** (2022) 79–118
https://dx.doi.org/10.4314/saaj.v22i1.4

An economic scenario generator for embedded derivatives in South Africa

By Alexis Levendis and Eben Maré

Submission date 28 May 2021
Acceptance date 19 October 2022

ABSTRACT
It is well known that interest rate risk is a dominating factor when pricing long-dated contingent claims. The Heston stochastic volatility model fails to capture this risk as the model assumes a constant interest rate throughout the life of the claim. To overcome this, the risk-free interest rate can be modelled by a Hull-White short rate process and can be combined with the Heston stochastic volatility model to form the so-called Heston-Hull-White model. The Heston-Hull-White model allows for correlation between the equity and interest rate processes, a component that is important when pricing long-dated contingent claims. In this paper, we apply the Heston-Hull-White model to price Guaranteed Minimum Maturity Benefits (GMMBs) and Guaranteed Minimum Death Benefits (GMDBs) offered in the life insurance industry in South Africa. We propose a further extension by including stochastic mortality rates based on either a continuous-time Cox-Ingersoll-Ross short rate process or a discrete-time AR(1)-ARCH(1) model. Our findings suggest that stochastic interest rates are the dominating factor when reserving for GMMB and GMDB products. Furthermore, a delta-hedging strategy can help reduce the variability of embedded derivative liabilities.

KEYWORDS
Heston-Hull-White, stochastic volatility, stochastic interest rates, stochastic mortality, GMMB, GMDB, pricing, hedging

CONTACT DETAILS
Alexis Levendis, Department of Actuarial Science, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa; Momentum Metropolitan Holdings Limited; Email: alexilevendis@gmail.com
Eben Maré, Department of Mathematics and Applied Mathematics, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa; Email: eben.mare@up.ac.za

1. INTRODUCTION

1.1 Life insurers often sell policies with embedded guarantees. Pricing these guarantees is challenging due to the financial and insurance risks involved. Two popular products sold by life insurers include the Guaranteed Minimum Maturity Benefit (GMMB) and the Guaranteed Minimum Death Benefit (GMDB). The former provides the policyholder with the greater of the guaranteed amount at maturity or the fund value in which the initial premium was invested, whereas the latter provides the policyholder with the greater of the guaranteed amount or the fund value at the time of death, should death occur before the maturity of the contract. To price these products, a model is required that can capture financial risks, including equity and interest rate volatility, and insurance risks such as mortality.

1.2 The Black-Scholes (1973) model is unsuitable for modelling long-dated guarantees as it does not account for stochastic interest rates or stochastic volatility. Furthermore, it has been shown that the model is inconsistent with the stylised facts of financial returns (see Cont, 2001). Fortunately, the Black-Scholes (1973) model has been extended to incorporate stochastic volatility.

1.3 The Heston (1993) model incorporates stochastic volatility, and, hence, can capture the equity skew observed in financial markets. However, as shown by Kammeyer & Kienitz (2012), stochastic interest rates are a bigger risk driver for long-dated contingent claims than stochastic volatility. Consequently, a model is required that can incorporate stochastic volatility as well as stochastic interest rates.

1.4 The Hull-White (1990) model is a stochastic process that describes the short-term instantaneous interest rate and is popular among practitioners. The model has the benefit in that it can fit the initial term structure of interest rates and has an analytical solution for zero-coupon bonds. Given the model's analytical tractability, researchers in the field of quantitative finance started to investigate its use with the Heston (1993) stochastic volatility model, leading to what is now known as the Heston-Hull-White model.

1.5 A particular challenge faced when pricing GMMB and GMDB products is their long-dated nature. As discussed by Maze (2014), the further a call or put option is from maturity, the more time a stochastic interest rate has to reach its long-run mean and affect the option price. Therefore, it is important for practitioners to include stochastic interest rates when reserving for embedded derivatives. The Heston-Hull-White model will allow for this. Apart from financial risks, insurance risks such as mortality must also be incorporated.

1.6 Traditional actuarial modelling assumes deterministic mortality rates based on actuarial life tables. In reality, mortality rates are stochastic and it may be necessary to price contingent claims based on this assumption. In this paper, we extend the Heston-Hull-White model to include stochastic mortality rates based on either a continuous-time Cox-Ingersoll-Ross

(1985) process or discrete-time AR(1)-ARCH(1)[1] model when pricing products contingent on either survival or death. We call this extension the Heston-Hull-White-Mortality model.

1.7 The remainder of this paper is structured as follows: Section 2 reviews relevant literature on guaranteed maturity benefits, the Heston-Hull-White model, and stochastic mortality; Section 3 introduces the mathematical theory of the Heston-Hull-White model; Section 4 reviews the continuous-time Cox-Ingersoll-Ross (1985) short rate process and its extension proposed by Brigo & Mercurio (2001), and the discrete time AR(1)-ARCH(1) model that will be used for mortality rates; Section 5 shows the formulas that will be used to price the GMMB and GMDB products; Section 6 introduces the data that will be used to calibrate the Heston-Hull-White-Mortality model and price the GMMB and GMDB products; Section 7 shows our pricing and hedging results, and Section 8 concludes the findings.

2. LITERATURE REVIEW

2.1 In 2007, ING, a large Dutch banking group, presented a problem from the financial industry where the goal was to derive a closed-form or semi-closed form solution for a call option under the hybrid Heston-Hull-White model. To understand the complexity of this challenge, Oosterlee (2007) writes:

> Quite a few mathematicians took up this ING challenge, and during the week three subgroups were formed, each approaching the problem from a different side. A particular challenge here was that some Dutch professors in financial mathematics in earlier attempts were not able to come up with a closed form option pricing solution for this particular model.

From this challenge, three interesting approaches were suggested to solve the model, see Muskulus (2007), in 't Hout et al. (2007), and Fang & Janssens (2007).

2.2 The Heston-Hull-White model has become popular since the so-called ING challenge. Kammeyer & Kienitz (2009) considered an interest rate process that is independent of the equity and volatility processes, and showed the resulting characteristic function and calibration of the model based on the Carr & Madan (1999) fast Fourier transform (FFT) method. Still, the challenge remained to incorporate a non-zero correlation structure between the equity, volatility and interest rate processes.

2.3 Grzelak & Oosterlee (2011) made a breakthrough by deriving approximations for specific terms in the Heston-Hull-White model that allow for a non-zero correlation structure between the equity, volatility, and interest rate processes. The authors conclude that the approximations yield accurate prices for European options. This was a big contribution to the field of quantitative finance and the model has since been applied to price long-dated contingent claims.

1 First-order autoregressive and autoregressive conditional heteroskedastic model

2.4 Maze (2014) applied the Heston-Hull-White model to price long-dated European call options and compared the results to the standard Black-Scholes (1973) model. The author showed that call option prices under the Heston-Hull-White model increase when compared to the Black-Scholes (1973) model as the time to maturity increases. This is an important finding, especially considering the long-dated nature of embedded derivatives such as GMMB and GMDB products.

2.5 Patel (2019) implemented the Heston-Hull-White model to price long-dated European call options and compared the deterministic and stochastic approximations for the Heston-Hull-White model proposed by Grzelak & Oosterlee (2011). The author concludes that the deterministic approximation is a feasible way to calibrate the Heston-Hull-White model, whereas the stochastic approximation is inefficient.

2.6 As mentioned earlier, traditional actuarial modelling assumes deterministic mortality rates. In reality, mortality rates tend to behave randomly. Cairns et al. (2008) discuss various stochastic mortality models including the discrete-time Lee-Carter (1992) model and continuous-time short rate models. Mortality modelling is similar to interest rate modelling, hence, it is relatively simple to extend the theory of interest rates to mortality.

2.7 A weakness of many short rate models is that they cannot fit the initial term structure of interest rates. Examples include the Vasicek (1977) and Cox-Ingersoll-Ross (1985) models. The Hull-White (1990) model overcomes this issue, but, due to its Gaussian distribution, can produce negative values. Brigo & Mercurio (2001) extended the Cox-Ingersoll-Ross (1985) model to include a deterministic shift that allows the model to fit the initial term structure of interest rates whilst maintaining its analytical tractability for zero-coupon bonds. The model is termed the CIR++ model.

2.8 Truter (2012) applied the CIR++ model to price defaultable zero-coupon bonds. The author showed that the hazard rate, or instantaneous probability of default, can be modelled as a CIR++ process. From there, an analytical solution exists to calculate survival probabilities (the probability that a counterpart does not default). Simply replacing the hazard rate with the force of mortality (instantaneous rate of mortality), the CIR++ model can be used to calculate survival probabilities in a life insurance context.

2.9 Literature on the use of hybrid models to price embedded derivatives or variable annuities is limited. Wang (2011) applied the Heston-Hull-White model to price the Guaranteed Minimum Withdrawal Benefit (GMWB) product and discussed two semi-analytical techniques that can be used to calibrate the model to market data. The author considered the FFT technique of Carr & Madan (1999) and the Fourier-Cosine (COS) technique introduced by Fang & Oosterlee (2008). The author concludes that the Heston-Hull-White model is an appropriate model to use when pricing long-dated claims.

2.10 Ignatieva et al. (2016) presented a framework for pricing guaranteed maturity benefits with a regime-switching and stochastic mortality model. The authors considered a continuous-time two-factor affine mortality model and assume that the unsystematic mortality risk can be diversified away, hence, the instantaneous mortality intensity is the same under the physical and risk-neutral measures.

2.11 Ballotta et al. (2020) proposed a Lévy-based hybrid model that allows for dependence between interest rates and equity prices. Furthermore, the authors divide insurance risk into two components: mortality risk and surrender risk. The authors conclude that surrender risk is particularly important in the valuation of variable annuities due to the large number of policyholders who terminate their contracts prematurely.

2.12 Veilleux (2016) tested the impact of stochastic volatility, stochastic interest rates, and stochastic mortality on the hedge efficiency of the Guaranteed Lifetime Withdrawal Benefit (GLWB). The author considered various models including a regime-switching lognormal model to capture stochastic volatility, the Hull-White (1990) model for stochastic interest rates, and the Lee-Carter (1992) model for stochastic mortality. The author concludes that stochastic interest rates and stochastic mortality have a significant impact on the hedge efficiency of GLWB benefits.

2.13 In South Africa, the Actuarial Society of South Africa's advisory practice note, APN 110: Allowance for Embedded Investment Derivatives,[2] gives guidance to actuaries on how to reserve for embedded derivatives including GMMB and GMDB products. The advisory practice note is rather open-ended and leaves the actuary with many modelling choices. An extract from APN 110 reads:

> No specific investment return projection model is prescribed. The actuary may use any market-consistent stochastic investment return projection model that he/she deems appropriate for purposes of quantifying reserves required to meet the potential cost of embedded investment derivatives.

The only minimum prescription set by APN 110 is that the model must be "market-consistent", meaning that it must match the prices of tradable assets as closely as possible.

2.14 Limited research has been conducted in South Africa comparing different modelling approaches for embedded derivatives. Ngugi et al. (2015) applied a Variance-Gamma model to price GMMB and GMDB products written on the FTSE/JSE All Share Index (ALSI) and compared this to a regime-switching model (see Hardy (2003) for details). The authors conclude that the Variance-Gamma model is more aligned with the stylised facts of financial returns (see Cont, 2001). Furthermore, the authors suggest further research on the topic by

2 https://www.actuarialsociety.org.za/download/apn-110-allowance-for-embedded-investment-derivatives/

including stochastic interest rates and stochastic mortality. This suggestion is what motivated our research paper.

2.15 We are not aware of literature pertaining to the Heston-Hull-White model used to price embedded derivatives in the presence of stochastic mortality. Therefore, our contribution to the literature is by pricing GMMB and GMDB products with an extension of the Heston-Hull-White model that we term the Heston-Hull-White-Mortality model. We will apply the model to price and hedge GMMB and GMDB products written on the FTSE/JSE TOP40 index in South Africa.

2.16 In the next section, we introduce the mathematical theory of the Heston-Hull-White model.

3. THE HESTON-HULL-WHITE MODEL

3.1 Let $\mathbb{Q}$ denote the risk-neutral measure. Under the $\mathbb{Q}$-measure, the Heston-Hull-White model is given by the following system of stochastic differential equations (SDEs):

$$\begin{cases} dS_t &= r_t S_t dt + \sqrt{v_t} S_t dW_t^s, & S_0 > 0, \\ dv_t &= \kappa\left(\overline{v} - v_t\right) dt + \sigma\sqrt{v_t} dW_t^v, & v_0 > 0, \\ dr_t &= \lambda\left(\theta_t - r_t\right) dt + \eta dW_t^r, & r_0 \in \mathbb{R}, \end{cases}$$

where S_t is the underlying asset price a time t, v_t is the instantaneous variance of the underlying asset at time t driven by a Cox-Ingersoll-Ross (1985) process, and r_t is the instantaneous short term interest rate at time t driven by a Hull-White (1990) process.

3.2 The parameters of the variance process are defined as follows: κ denotes the mean reversion speed of the variance, $\overline{v}$ is the long-run mean of the variance, and σ is the volatility of the variance.

3.3 The parameters of the short-term interest rate process are defined as follows: θ_t denotes the time-dependent mean reversion level of the interest rate, λ denotes the mean reversion speed of the interest rate, and η is the volatility of interest rate.

3.4 In the system of SDEs, the Brownian motions, dW_t^s, dW_t^v, and dW_t^r are correlated by $\rho_{s,v}$, $\rho_{s,r}$, and $\rho_{v,r}$ respectively.

3.5 Taking $x_t = \log S_t$, Grzelak & Oosterlee (2011) consider the log-dynamics for the Heston-Hull-White model as it is often easier to work with this form. Applying Itô's lemma, the log-dynamics for the Heston-Hull-White model are given by the following system of SDEs under the $\mathbb{Q}$-measure:

$$\begin{cases} dx_t &= \left(r_t - \frac{1}{2}v_t\right)dt + \sqrt{v_t}\,dW_t^s, \quad x_0 = \log S_0 > 0, \\ dv_t &= \kappa\left(\overline{v} - v_t\right)dt + \sigma\sqrt{v_t}\,dW_t^v, \quad v_0 > 0, \\ dr_t &= \lambda\left(\theta_t - r_t\right)dt + \eta dW_t^r, \qquad r_0 \in \mathbb{R}. \end{cases}$$

3.6 Patel (2019) summarises the work of Grzelak & Oosterlee (2011) and shows that the Heston-Hull-White log-dynamics can be expressed in terms of independent Brownian motions. If $\boldsymbol{B}_t$ is a standard three-dimensional Brownian motion, then

$$d\boldsymbol{X}_t = \mu\left(\boldsymbol{X}_t\right)dt + \sigma\left(\boldsymbol{X}_t\right)d\boldsymbol{B}_t, \tag{1}$$

where

$$\boldsymbol{X}_t = \begin{bmatrix} x_t \\ v_t \\ r_t \end{bmatrix},$$

$$\mu\left(\boldsymbol{X}_t\right) = \begin{bmatrix} r_t - \frac{1}{2}v_t \\ \kappa\left(\overline{v} - v_t\right) \\ \lambda\left(\theta_t - r_t\right) \end{bmatrix},$$

and

$$\sigma\left(\boldsymbol{X}_t\right) = \begin{bmatrix} \sqrt{v_t} & 0 & 0 \\ \rho_{s,v}\sigma\sqrt{v_t} & \sqrt{1-\rho_{s,v}^2}\,\sigma\sqrt{v_t} & 0 \\ \rho_{s,r}\eta & \dfrac{\rho_{v,r} - \rho_{s,r}\rho_{s,v}}{\sqrt{1-\rho_{s,v}^2}}\eta & \sqrt{1-\rho_{s,r}^2 - \left(\dfrac{\rho_{v,r} - \rho_{s,r}\rho_{s,v}}{\sqrt{1-\rho_{s,v}^2}}\right)}\eta \end{bmatrix}.$$

See Patel (2019) for the proof.

3.7 The covariance matrix, $\Sigma\left(\boldsymbol{X}_t\right) = \sigma\left(\boldsymbol{X}_t\right)\sigma(\boldsymbol{X}_t)^\top$, for the Heston-Hull-White model is given by:

$$\Sigma\left(\boldsymbol{X}_t\right) = \begin{bmatrix} v_t & \rho_{s,v}\sigma v_t & \rho_{s,r}\eta\sqrt{v_t} \\ * & \sigma^2 v_t & \rho_{v,r}\eta\sqrt{v_t} \\ * & * & \eta^2 \end{bmatrix}. \tag{2}$$

When the underlying asset and variance processes are correlated to the interest rate process, the Heston-Hull-White model cannot be written in an affine form. The reader is referred to Duffie et al. (2000), Grzelak & Oosterlee (2011), and Patel (2019) for more information on affine models. In short, a model is said to be affine if each element of its drift and covariance can be written as a linear function of the state variables.

3.8 From equation (2), it is clear that the covariance matrix contains elements that are nonlinear functions of the state variables, more specifically $\sqrt{v_t}$. In this paper, we are only interested in the case where the interest rate process is correlated to the spot process, hence, we set $\rho_{v,r} = 0$. To make the model affine, Grzelak & Oosterlee (2011) propose a deterministic approximation by replacing $\sqrt{v_t}$ with $\mathbb{E}\left[\sqrt{v_t}\right]$. The covariance matrix then becomes:

$$\Sigma\left(\boldsymbol{X}_t\right) = \begin{bmatrix} v_t & \rho_{s,v}\sigma v_t & \rho_{s,r}\eta\mathbb{E}\left[\sqrt{v_t}\right] \\ * & \sigma^2 v_t & 0 \\ * & * & \eta^2 \end{bmatrix}. \tag{3}$$

Grzelak & Oosterlee (2011) approximate $\mathbb{E}\left[\sqrt{v_t}\right]$ with the following function:

$$\Lambda_t = \sqrt{c_t\left(\lambda_t - 1\right) + c_t d + \frac{c_t d}{2\left(d + \lambda_t\right)}}, \tag{4}$$

where

$$c_t = \frac{1}{4\kappa}\sigma^2\left(1 - e^{-\kappa t}\right), \quad d = \frac{4\kappa\overline{v}}{\sigma^2}, \quad \lambda_t = \frac{4\kappa v_0 e^{-\kappa t}}{\sigma^2\left(1 - e^{-\kappa t}\right)}.$$

See Grzelak & Oosterlee (2011), and Patel (2019) for the proof.

3.9 Grzelak & Oosterlee (2011) mention that the approximation in equation (4) is still non-trivial and propose a further approximation for $\mathbb{E}\left[\sqrt{v_t}\right]$ with the following function:

$$\mathbb{E}\left[\sqrt{v_t}\right] \approx a + be^{-ct} := \tilde{\Lambda}_t, \tag{5}$$

where

$$a = \sqrt{\overline{v} - \frac{\sigma^2}{8\kappa}},\ b - \sqrt{v_0} - a,\ c = -\log\left(b^{-1}\left(\Lambda\left(1\right) - a\right)\right).$$

As we will show in the next section, the approximation in equation (5) plays an important role in the Heston-Hull-White characteristic function.

3.10 The Heston-Hull-White characteristic function

3.10.1 Grzelak & Oosterlee (2011) assume a constant term structure of interest rates, $\theta_t = \theta$, in their derivation of Heston-Hull-White characteristic function. This assumption simplifies the mathematics and we will maintain this assumption throughout the rest of this paper. From Grzelak & Oosterlee (2011), the characteristic function is given by:

$$\phi_{HHW}\left(u,\mathbf{X}_t,\tau\right)=\exp\left(\tilde{A}\left(u,\tau\right)+B\left(u,\tau\right)x_t+C\left(u,\tau\right)r_t+D\left(u,\tau\right)v_t\right), \tag{6}$$

where

$$B\left(u,\tau\right)=iu,$$

$$C\left(u,\tau\right)=\frac{1}{\lambda}\left(iu-1\right)\left(1-e^{-\lambda\tau}\right),$$

$$D\left(u,\tau\right)=\frac{1-e^{-D_1\tau}}{\sigma^2\left(1-ge^{-D_1\tau}\right)}\left(\kappa-\sigma\rho_{s,v}iu-D_1\right),$$

$$A\left(u,\tau\right)=\lambda\theta I_1\left(u,\tau\right)+\kappa\bar{v}I_2\left(u,\tau\right)+\frac{1}{2}\eta^2 I_3\left(u,\tau\right),$$

$$\tilde{A}\left(u,\tau\right)=A\left(u,\tau\right)+\rho_{s,r}\eta I_4\left(u,\tau\right),$$

with

$$D_1=\sqrt{\left(\sigma\rho_{s,v}iu-\kappa\right)^2-\sigma^2 iu\left(iu-1\right)},$$

$$g=\frac{\kappa-\sigma\rho_{s,v}iu-D_1}{\kappa-\sigma\rho_{s,v}iu+D_1},$$

$$I_1\left(u,\tau\right)=\frac{1}{\lambda}\left(iu-1\right)\left(\tau+\frac{1}{\lambda}\left(e^{-\lambda\tau}-1\right)\right),$$

$$I_2\left(u,\tau\right)=\frac{\tau}{\sigma^2}\left(\kappa-\sigma\rho_{s,v}iu-D_1\right)-\frac{2}{\sigma^2}\log\left(\frac{1-ge^{-D_1\tau}}{1-g}\right),$$

$$I_3\left(u,\tau\right)=\frac{1}{2\lambda^3}\left(i+u\right)^2\left(3+e^{-2\lambda\tau}-4e^{-\lambda\tau}-2\lambda\tau\right),$$

$$I_4\left(u,\tau\right)\approx-\frac{1}{\lambda}\left(iu+u^2\right)\left[\frac{b}{c}\left(e^{-ct}-e^{-cT}+a\tau+\frac{a}{\lambda}\left(e^{-\lambda\tau}-1\right)\right.\right.$$

$$\left.\left.+\frac{b}{c-\lambda}e^{-cT}\left(1-e^{\tau(\lambda-c)}\right)\right)\right].$$

3.10.2 The deterministic approximation proposed by Grzelak & Oosterlee (2011) in equation (5) feeds into the valuation of I_4 above, hence the approximation sign. Since the characteristic function for the Heston-Hull-White model is known, Fourier methods can be used to solve the price of a European call option as we show in the next section.

3.11 Heston-Hull-White calibration

3.11.1 Carr & Madan (1999) showed that European call options can be priced via Fourier transform methods if the characteristic function of the model is known analytically. The authors initially considered constant interest rates in their paper. In the case of stochastic interest rates, the Fourier method must be adapted slightly.

3.11.2 Patel (2019) derived the price of a European call option with strike K and maturity T using the Fourier method under stochastic interest rates as:

$$V_{HHW}(k) = \frac{e^{-\alpha k}}{\pi} \int_0^{\infty} Re\left\{e^{-ivk}\psi(v)\right\} dv, \tag{7}$$

where $k = \log(K)$, $\alpha > 0$, $i = \sqrt{-1}$, $Re\{\cdot\}$ denotes the real part of the complex number, and

$$\psi(v) = \frac{\phi_{HHW}\left(v-(\alpha+1)i\right)}{(\alpha+1+iv)(\alpha+iv)}.$$

3.11.3 The symbol α refers to the dampening coefficient and a suitable value must be chosen in order to compute the option price in equation (7). As mentioned by Patel (2019), choosing α is subjective. The author set $\alpha = 0.75$ and tested the sensitivity of European call option prices to various values of α around 0.75. From the author's testing, it was concluded that $\alpha = 0.75$ is sufficient. We therefore set $\alpha = 0.75$ going forward.

3.11.4 As mentioned, we are not aware of any literature that combines the Heston-Hull-White model with a stochastic mortality model. GMMB and GMDB products are contingent on survival or death; this implies that mortality is an important risk driver when pricing these products. Therefore, we propose an extension of the Heston-Hull-White model by including stochastic mortality rates based on either a continuous-time Cox-Ingersoll-Ross (1985) process or discrete-time AR(1)-ARCH(1) model.

4. STOCHASTIC MORTALITY

This section is divided into two parts. In the first subsection, we introduce the continuous-time Cox-Ingersoll-Ross (1985) model for mortality rates. In the second subsection, we present the discrete-time AR(1)-ARCH(1) model of Syuhada & Hakim (2021). The reason for considering both a continuous-time model and a discrete-time model is to allow for calibration depending on the type of mortality data that is available. A continuous-time model is generally calibrated to a survival probability curve derived from historical data or longevity swaps, whereas a discrete-time model is calibrated to a historical time series of mortality rates.

4.1 Cox-Ingersoll-Ross model

4.1.1 Let $\mu_{x+t,t}$ denote the force of mortality at time t and age $x+t$. The force of mortality represents the instantaneous probability of death at time t for a life aged $x+t$. Cairns et al. (2008) show that the probability of survival up to time t under $\mathbb{Q}$ is given by:

$$ {}_tp_x = S_x(0,t) = \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^t \mu_{x+s,s} ds} \right]. \tag{8} $$

4.1.2 In order to calculate the expectation in equation (8), a suitable stochastic process must be chosen to describe the dynamics of $\mu_{x+t,t}$. Cairns et al. (2008) list certain criteria for mortality rates, including the following:

— mortality rates should be positive;
— long-term dynamics of the model should be reasonable; and
— the model should be simple to implement using analytical methods or numerical methods.

4.1.3 Based on the criteria listed above, we propose the Cox-Ingersoll-Ross (1985) process as a plausible model to describe mortality rates. The reason for choosing this process is the following: the model produces positive values (although care must be taken to ensure the Feller condition[3] is met), the model is mean reverting which means that long-term rates will not explode, and the model has an analytical solution for survival probabilities.

4.1.4 For a fixed age x at inception, the original Cox-Ingersoll-Ross (1985) model is given by the SDE:

$$ d\mu_{x+t,t} = \gamma(\omega - \mu_{x+t,t})dt + \xi\sqrt{\mu_{x+t,t}}dW_t^{\mu}, \qquad \mu_{x,0} > 0, \tag{9} $$

where γ denotes the mean reversion speed of mortality, ω is the long-run mean of mortality rates, and ξ is the volatility of mortality rates.

4.1.5 A shortcoming of this model is that it cannot fit the term structure of mortality. When the force of mortality is plotted against age, it shows an increasing trend. The original Cox-Ingersoll-Ross (1985) model does not capture this effect. Fortunately, the model has been extended to incorporate a term structure.

4.1.6 Brigo & Mercurio (2001) extended the Cox-Ingersoll-Ross (1985) model to incorporate a term structure and created the so-called CIR++ model. The model takes the following form where $\varphi_{x,t}^{CIR}$ is a deterministic function of time:

$$ \mu_{x+t,t} = \varphi_{x,t}^{CIR} + X_{x+t,t}, \tag{10} $$

where

$$ dX_{x+t,t} = \gamma(\omega - X_{x+t,t})dt + \xi\sqrt{X_{x+t,t}}dW_t^{\mu}, \qquad X_{x,0} > 0, $$

3 The Feller condition is met if $2\gamma\omega > \xi^2$, where γ is the mean reversion speed of mortality, ω is the long-run mean of mortality, and ξ is the volatility of mortality.

$$\varphi_{x,t}^{CIR} = f_x^M(0,t) - f_x^{CIR}(0,t).$$

4.1.7 Furthermore, $f_x^M(0,t)$ denotes the market forward mortality curve observed from the initial term structure of mortality (survival curve), and

$$f_x^{CIR}(0,t) = 2\gamma\omega\frac{e^{th}-1}{2h+(\gamma+h)(e^{th}-1)} + X_{x,0}\frac{4h^2e^{th}}{\left[2h+(\gamma+h)(e^{th}-1)\right]^2},$$

with $h = \sqrt{\gamma^2+2\xi^2}$.

4.1.8 The survival probability from time t up to time T is then given by:

$$S_x(t,T) = \frac{S_x^M(0,T)L_x(0,t)e^{-M_x(0,t)X_{x,0}}}{S_x^M(0,t)L_x(0,T)e^{-M_x(0,T)X_{x,0}}}L_x(t,T)e^{-M_x(t,T)\left[\mu_{x+t,t}-\varphi_{x,t}^{CIR}\right]}, \tag{11}$$

where

$$L_x(t,T) = \left[\frac{2he^{(\gamma+h)(T-t)/2}}{2h+(\gamma+h)(e^{(T-t)h}-1)}\right]^{2\gamma\omega/\xi^2},$$

$$M_x(t,T) = \frac{2(e^{(T-t)h}-1)}{2h+(\gamma+h)(e^{(T-t)}h-1)},$$

and $S_x^M(0,\cdot)$ denotes the market observed survival probability curve at $t=0$ for a life aged x at inception.

4.1.9 A major advantage of term structure models, and specifically the CIR++ model, is that they fit the input survival probability curve at $t=0$, i.e, $S_x(0,.)$.

4.1.10 Throughout the rest of this paper, we make the same assumption as Ignatieva et al. (2016) in that the life insurer has an adequately large number of policyholders so that unsystematic mortality risk can be diversified away. This implies that the force of mortality under the physical probability measure $\mathbb{P}$ and risk-neutral probability measure $\mathbb{Q}$ are equal:

$$\mu_{x+t,t}^{\mathbb{P}} = \mu_{x+t,t}^{\mathbb{Q}}.$$

4.1.11 Life insurers in South Africa typically use life tables published by the Actuarial Society of South Africa to calculate survival probabilities. Our assumption that the instantaneous force of mortality is the same under the $\mathbb{P}$ and $\mathbb{Q}$ measures allows us to calibrate the CIR++ model directly to a survival probability curve derived from the life tables.

4.1.12 In the next subsection, we present the discrete-time mortality model of Syuhada & Hakim (2021).

4.2 AR(1)-ARCH(1) model

4.2.1 Syuhada & Hakim (2021) adopted the AR(1)-ARCH(1) model of Giacometti et al. (2012) and Lin et al. (2015) and applied the model to yearly changes in the log mortality rate. Syuhada & Hakim (2021) explain that the conditional mean of the change in log mortality rate is modelled by a first-order autoregressive AR(1) model, and the conditional variance is modelled by a first-order autoregressive conditional heteroscedastic ARCH(1) stochastic volatility model to capture volatility clustering and the time-varying nature of volatility.

4.2.2 Let $\mu_{x,t}$ denote the force of mortality at age x in year t calculated as:

$$\mu_{x,t} = \frac{D_{x,t}}{E_{x,t}}, \tag{12}$$

where $D_{x,t}$ denotes the number of individuals aged x in year t who die before year $t+1$ or age $x+1$, and $E_{x,t}$ denotes the number of individuals exposed to death risk at age x in year t.

4.2.3 Let $Y_{x,t}$ denote the change in the log mortality rate for each fixed age $x \in \{x_0, x_0+1, x_0+2, ..., X_{max}\}$, that is:

$$Y_{x,t} = \log \mu_{x,t} - \log \mu_{x,t-1} = \log \frac{\mu_{x,t}}{\mu_{x,t-1}}. \tag{13}$$

where $t \in \{t_0+1, t_0+2, ..., T_{max}\}$, with X_{max} and T_{max} denoting the maximum age and year in the historical data respectively.

4.2.3 Syuhada & Hakim (2021) model $Y_{x,t}$ according to an AR(1)-ARCH(1) model:

$$Y_{x,t} = a_x + b_x Y_{x,t-1} + \varepsilon_{x,t}, \tag{14}$$

where

$$\varepsilon_{x,t} = \sqrt{h_{x,t}}\, z_{x,t},$$

$$h_{x,t} = \omega_x + \gamma_x \varepsilon_{x,t-1}^2,$$

with $h_{x,t}$ denoting the conditional variance, $a_x \in \mathbb{R}$, $b_x \in (-1,1)$, $\omega_x \in (0,\infty)$, $\gamma_x \in [0,1)$, and $\varepsilon_{x,t}$ is an innovation parameter with $z_{x,t} \sim N(0,1)$.

4.2.4 For each fixed age x, the model can be calibrated to a time series of historical mortality rates via maximum likelihood estimation.

4.2.5 This concludes the section on stochastic mortality models. In the next section, we introduce the mathematics of the GMMB and GMDB products.

5. GUARANTEED MINIMUM MATURITY AND DEATH BENEFITS

5.1 From now on, we will refer to the Heston-Hull-White-Mortality model when combining the Heston-Hull-White model with either the Cox-Ingersoll-Ross (1985) or AR(1)-ARCH(1) models for mortality. We assume that the force of mortality in the Heston-Hull-White-Mortality model evolves independently from the processes driving S_t, v_t, and

r_t. This implies that the Heston-Hull-White model and CIR++/AR(1)-ARCH(1) stochastic mortality model can be calibrated independently. Furthermore, an independent mortality process will simplify the calculations for the GMMB and GMDB products going forward.

5.2 We make extensive use of the following relationship when deriving pricing formulae for the GMMB and GMDB products: given two independent random variables Y and Z, the expected value of their product is equal to the product of their expected values:

$$\mathbb{E}[YZ] = \mathbb{E}[Y]\mathbb{E}[Z]. \tag{15}$$

5.3 We use the same notation as in the textbook by Feng (2018) when presenting the pricing formulae for the GMMB and GMDB products, adjusting the formulae to account for stochastic interest rates and stochastic mortality.

5.4 The price of a GMMB at $t=0$ from a life insurer's perspective is given by the formula:

$$\begin{aligned}
V_{GMMB,0} &= \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r_s ds}(G-F_T)^+ e^{-\int_0^T \mu_{x+s,s} ds}\right] \\
&\quad -\int_0^T \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^s r_u\,du} m_e F_s e^{-\int_0^s \mu_{x+u,u}\,du}\right] ds \\
&= \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r_s ds}(G-F_T)^+\right]\mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T \mu_{x+s,s} ds}\right] \\
&\quad -m_e\int_0^T \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^s r_u\,du} F_s\right]\mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^s \mu_{x+u,u} du}\right] ds \\
&= \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r_s ds}\left(G-F_T\right)^+\right]{}_T p_x - m_e\int_0^T \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^s r_u\,du} F_s\right]{}_s p_x ds,
\end{aligned} \tag{16}$$

where

— G is the minimum guaranteed amount defined at inception of the contract;
— F_t is the value of the fund account at time t in which the initial premium was invested; and
— m_e is the annualised fee for the GMMB deducted by the insurer from the fund account expressed as a percentage.

5.5 Furthermore, $F_t = F_0 \dfrac{S_t}{S_0} e^{-mt}$, where S_t denotes the value of the underlying equity index at time t, and m denotes the annualised charge payable by the insurer, expressed as a percentage. We assume that $m_e < m$.

5.6 Note from equation (16) that the terms $\mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r_s\,ds}(G-F_T)^+\right]$ and the term $\mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T \mu_{x+s,s}ds}\right]$ were split using the relation in equation (15) since the force of mortality is independent from all other risk drivers. Furthermore, if the bank account is used as numeraire, the discount factor cannot be removed from the expectation since the interest rate process is stochastic and correlated to the stochastic process driving the fund value. Therefore, the terms $\mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r_s\,ds}(G-F_T)^+\right]$ and $m_e\int_0^T \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^s r_u\,du}F_s\right]{}_sp_x$ can be solved using the brute force Monte Carlo method. Ignatieva et al. (2016) proposed a Fourier space time-stepping (FST) algorithm to reduce the computational cost of pricing embedded derivatives. See Ignatieva et al. (2016) for details.

5.7 The price of a GMDB at $t=0$ from a life insurer's perspective is given by the formula:

$$
\begin{aligned}
V_{GMDB,0} &= \int_0^T \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^s r_u\,du}\mu_{x+s,s}e^{-\int_0^s \mu_{x+u,u}\,du}(G_s-F_s)^+\right]ds \\
&\quad -\int_0^T \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^s r_u\,du}m_d e^{-\int_0^s \mu_{x+u,u}\,du}F_s\right]ds \\
&= \int_0^T \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^s r_u\,du}(G_s-F_s)^+\right]\mathbb{E}_{\mathbb{Q}}\left[\mu_{x+s,s}e^{-\int_0^s \mu_{x+u,u}\,du}\right]ds \\
&\quad -m_d\int_0^T \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r_s ds}F_s\right]\mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^s \mu_{x+u,u}\,du}\right]ds \\
&= \int_0^T \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^s r_u\,du}(G_s-F_s)^+\right]\mathbb{E}_{\mathbb{Q}}\left[\mu_{x+s,s}e^{-\int_0^s \mu_{x+u,u}\,du}\right]ds \\
&\quad -m_d\int_0^T \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^s r_s\,ds}F_s\right]{}_sp_x\,ds,
\end{aligned}
\tag{17}
$$

where m_d denotes the annualised fee for the GMDB deducted by the insurer from the fund account. We assume that $m_d < m$.

5.8 Recall the probability of survival formula for a life aged x at inception:

$$
{}_tp_x = S_x(0,t) = \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^t \mu_{x+s,s}\,ds}\right].
$$

5.9 The instantaneous probability of dying at time t and age $x+t$, given survival up to time t, is given by the term $\mathbb{E}_{\mathbb{Q}}\left[\mu_{x+t,t}e^{-\int_0^t \mu_{x+s,s}\,ds}\right]$. This term is used extensively in the GMDB pricing formula and can be solved numerically by the "brute force" Monte Carlo method.

5.10 The pricing formula for the GMDB is more complex than the pricing formula for the GMMB. This is because the payoff of the GMDB product can occur at any time in the interval $[0,T]$ depending on when the policyholder dies, whereas the GMMB only pays an amount of $\max(G,F_T)$ if the policyholder survives to time T.

5.11 In the next section, we introduce the data used in this study.

6. DATA

6.1 Daily FTSE/JSE TOP40 closing prices from September 2005 to November 2020 were downloaded from za.investing.com.[4] The historical closing prices are shown in Figure 1.

6.2 For the Hull-White (1990) component of the Heston-Hull-White model, we require a starting value for r_0. We use the South African 3-month T-Bill rate as a proxy for the short term interest rate. Historical 3-month T-Bill rates were downloaded from the South African Reserve Bank website[5] for the period September 2005 to November 2020 and are shown in Figure 2.

6.3 South African mortality rates were sourced from the Actuarial Society of South Africa's investigation: Report on pensioner mortality 2005–2010,[6] published in February 2017. The investigation was conducted by the Continuous Statistical Investigations (CSI) Committee of the Actuarial Society of South Africa. In the study, the CSI used pensioner data from the start of 2005 to the end of 2010 to construct the force of mortality for each age from 50 to 110, split by male and female.

6.4 Figure 3 shows the survival probability, p_{50}, $t \in \{1,2,...,60\}$, for a 50-year old South African male and female pensioner.

6.5 From Figure 3, the survival probability for a South African 50-year old male and female pensioner decreases the further we look into the future. Furthermore, the survival probability for a 50-year old male is considerably lower than a 50-year old female. The CIR++ model is calibrated to the survival probability curves in Figure 3.

4 https://za.investing.com/indices/ftse-jse-top-40-historical-data
5 https://www.resbank.co.za/en/home/what-we-do/statistics/key-statistics/selected-historical-rates
6 http://legacy.actuarialsociety.org.za/Societyactivities/CommitteeActivities/ContinuousStatisticalInvestigation(CSI).aspx

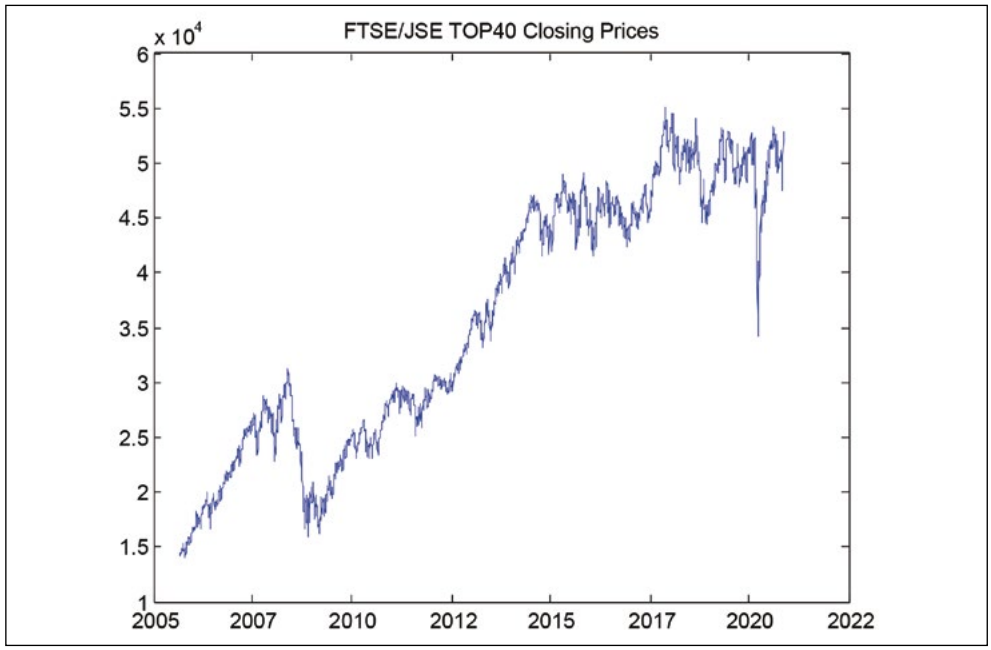

FIGURE 1. Historical FTSE/JSE TOP40 closing prices

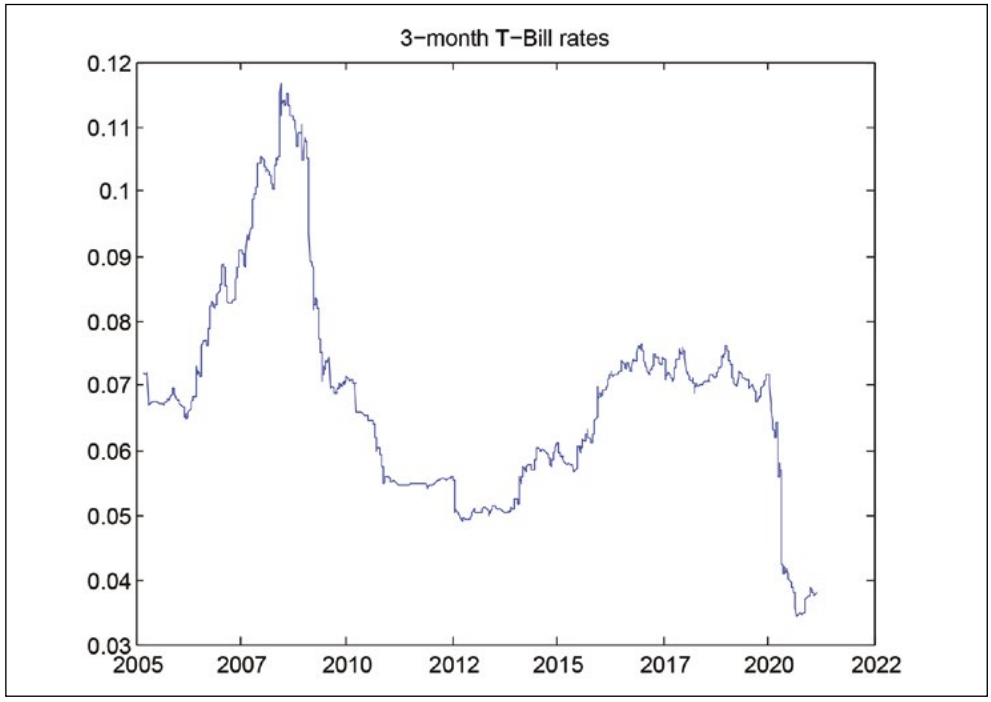

FIGURE 2. Historical 3-month T-bill rates in South Africa

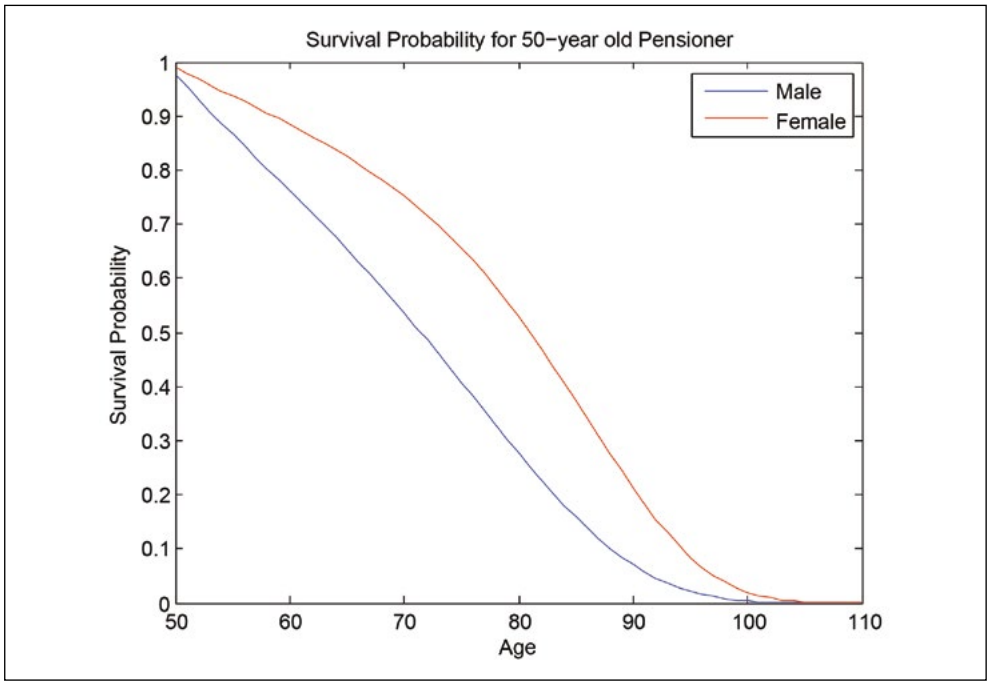

FIGURE 3. Survival probability for 50-year old male and female pensioner

6.6 For the purpose of the AR(1)-ARCH(1) model, we were unable to source historical mortality rates, $\mu_{x,t}$, for South African males and females. However, our goal is not to focus on the quality or availability of historical mortality data, but rather to show how to calibrate and apply the AR(1)-ARCH(1) model. Therefore, we use historical mortality rates for the United Kingdom (UK) from 1922 to 2020 sourced from the Human Mortality Database[7] (HMD), the world's leading source of mortality data for developed countries. The historical mortality rates for the UK are split by age and sex.

6.7 It is important to note that mortality in the UK is not representative of mortality in South Africa due to various social and economic factors such as income, education, employment, and community safety. However, using UK mortality data does not diminish the point we are trying to make. The only reason for using UK mortality data is that it is comprehensive and freely available from HMD. Any life insurer in South Africa with access to South African mortality data can easily adjust the input data to calibrate the AR(1)-ARCH(1) model.

6.6 Figures 4 and 5 show the log mortality rates for UK males and females from 1922 to 2020 for ages 0 to 100.

7 https://www.mortality.org/

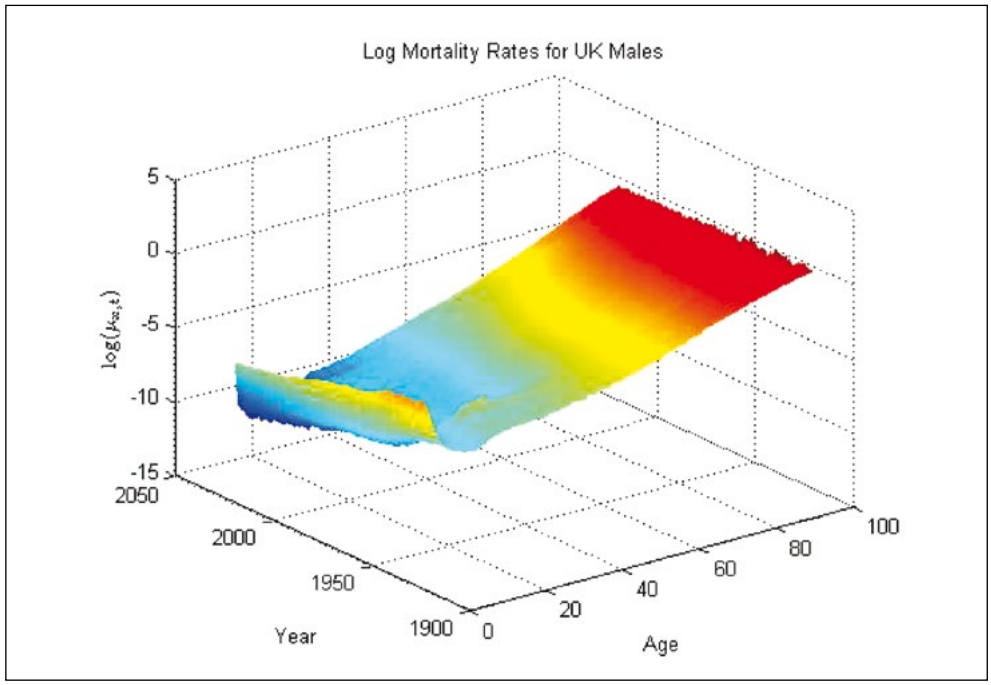

FIGURE 4. Log mortality rates for UK males

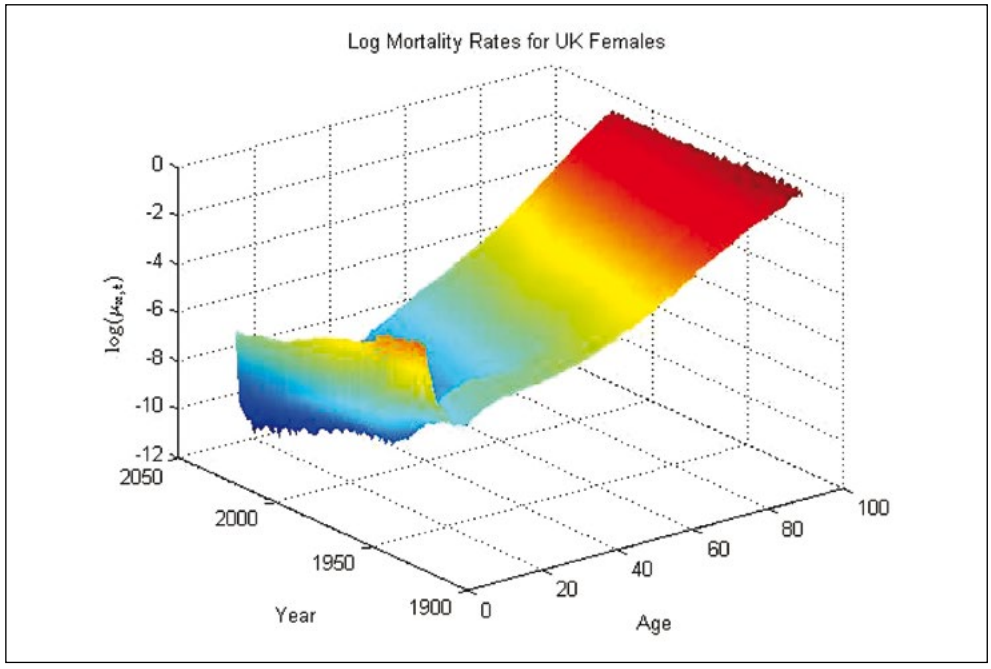

FIGURE 5. Log mortality rates for UK females

6.7 Figures 6 and 7 show the yearly changes in log mortality rates for UK males and females.

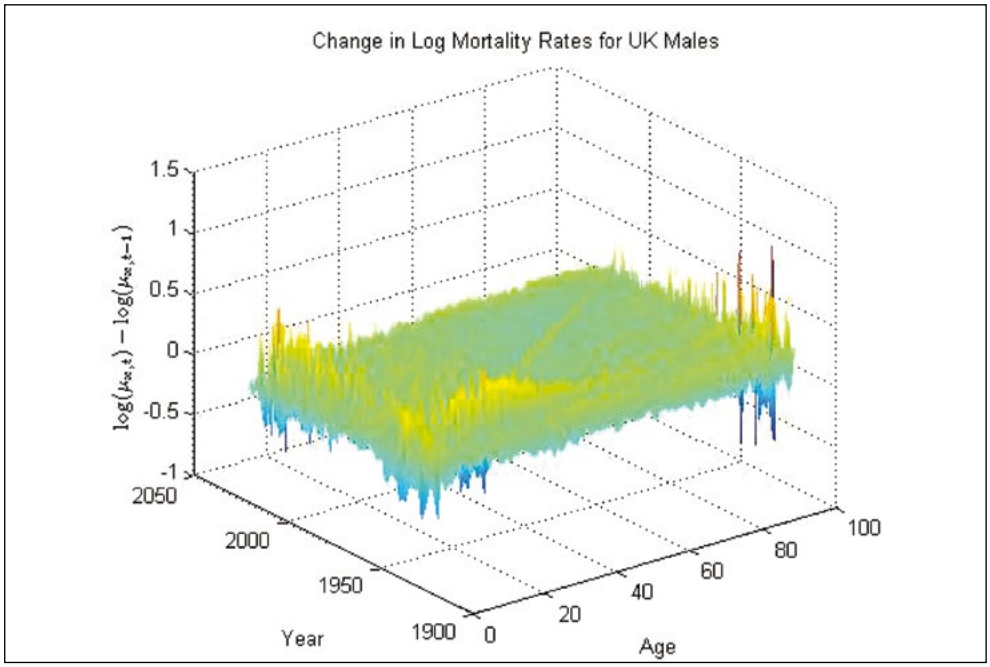

FIGURE 6. Change in log mortality rates for UK males

6.8 The AR(1)-ARCH(1) model of Syuhada & Hakim (2021) can now be applied to the stationary time series of the changes in log mortality rates for each age.

6.9 Weekly FTSE/JSE TOP40 volatility surfaces from September 2005 to November 2020 were provided by Peresec.[8] Each volatility surface covers a forward moneyness range of 75% to 125% and maturities from 1 to 15 months. For more details on the methodology used to construct the volatility surfaces, see Flint & Maré (2017).

6.10 The South African equity options market is illiquid compared to developed markets like Europe and the United States. As mentioned by Flint & Maré (2017), it is uncommon for options to trade beyond an expiry of 15 months in South Africa. This poses a further challenge when pricing GMMB and GMDB products with maturities typically extending far beyond 15 months.

6.11 Due to the illiquid equity options market in South Africa, there is no benchmark volatility that can be used for calibration beyond the 15-month mark. This implies that the

8 https://www.peresec.com/

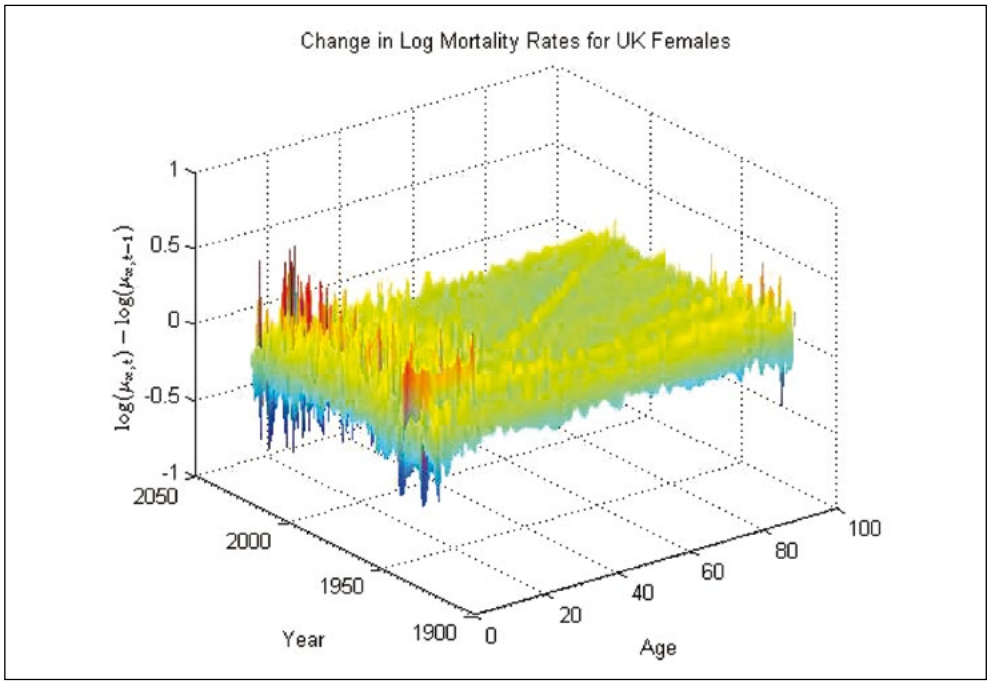

FIGURE 7. Change in log mortality rates for UK females

volatility term structure produced by the model can yield nonsensical values for long-dated maturities. Flint et al. (2014) tested several models that can produce long-term volatility estimates and showed that different models yield substantial differences for the volatility term structure.

6.12 In this paper, we allow the Heston-Hull-White-Mortality model to generate the volatility term structure beyond the 15-month mark. We will test the model by plotting the term structure of volatility to ensure that the model leads to plausible estimates for long-term volatility. For more detail on long-term volatility estimation, see Flint et al. (2014).

6.13 Figure 8 shows the implied volatility surface constructed from FTSE/JSE TOP40 index options as at 16 November 2020.

6.14 Note from Figure 8 the volatility skew observed across forward moneyness levels and the non-constant volatility across term to maturity. APN 110 requires that the investment return projection model reproduce this surface as closely as possible.

6.15 In the next section, we implement the Heston-Hull-White-Mortality model to price GMMB and GMDB products written on the FTSE/JSE TOP40 index, and backtest a delta-hedging strategy for the GMMB.

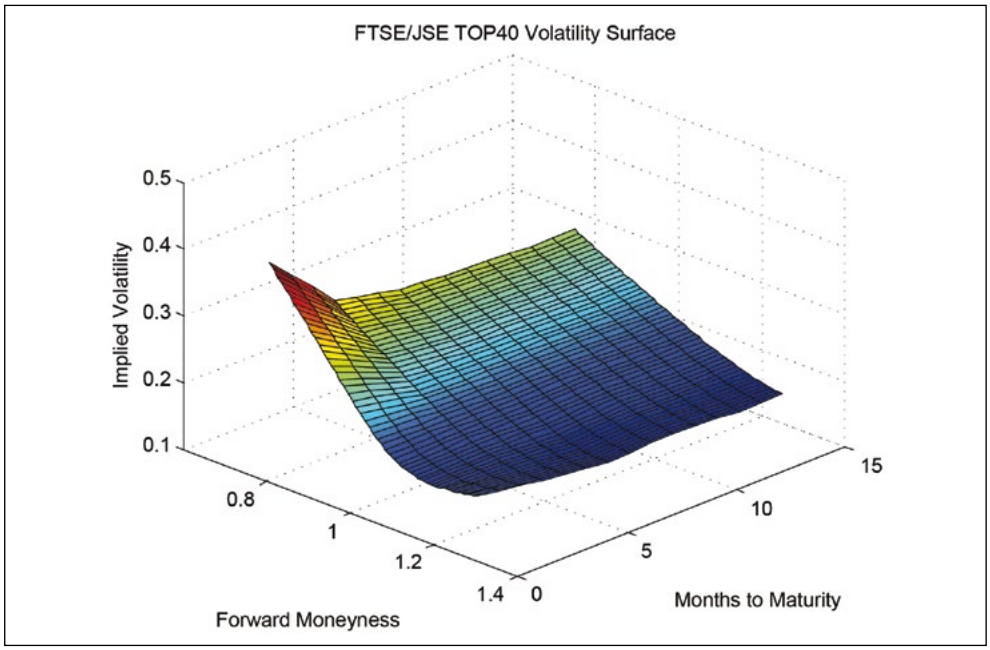

FIGURE 8. FTSE/JSE TOP40 volatility surface as at 16 November 2020

7. NUMERICAL RESULTS

This section has two subsections. The first subsection shows the pricing results obtained from the Heston-Hull-White-Mortality model for the GMMB and GMDB products, and the sensitivity of the prices to various model parameters. In the second subsection, we perform a backtesting experiment over weekly periods from September 2005 to November 2020 to test the hedging performance of the Heston-Hull-White-Mortality for the GMMB liability.

7.1 Pricing

7.1.1 The first step in pricing the GMMB and GMDB liabilities is to calibrate the Heston-Hull-White parameters to market data. We optimise the set of parameters on the following search space:

$$\begin{aligned}\Omega^{Search} &= D_{\kappa} \times D_{v_0} \times D_{\bar{v}} \times D_{\sigma} \times D_{\rho_{s,v}} \times D_{\lambda} \times D_{\eta} \times D_{\rho_{s,r}} \\ &= [1,1] \times [0,1] \times [0,1] \times [0,3] \times [-1,1] \times [0,10] \times [0,1] \times [-1,1],\end{aligned}$$

where D denotes the search domain for each parameter.

7.1.2 Grzelak & Oosterlee (2011) suggest that the Heston-Hull-White model be calibrated in two stages. First, the parameters for the Hull-White (1990) short rate process should be estimated independently from the equity (Heston, 1993) component. Then, keeping the parameters fixed for the short-term interest rate process, the remaining parameters are estimated.

7.1.3 The Hull-White (1990) mean reversion parameter, λ, is generally estimated from historical data in practice. Once λ has been estimated, the volatility parameter, η, is calibrated to a set of at-the-money swaptions.

7.1.4 The purpose of this paper is not to focus on the calibration of the Hull-White (1990) model, since this in itself can become rather complex, see, for example, Gurrieri et al. (2009). Rather, we aim to show the sensitivity of the GMMB and GMDB prices to various short-term interest rate parameters.

7.1.5 Once the parameters for the Hull-White (1990) component of the Heston-Hull-White model have been estimated, the next step is to calibrate the volatility parameters for the Heston (1993) component of the Heston-Hull-White model. We implemented the Fourier method of Carr & Madan (1999) discussed in Section 3.11.

7.1.6 The Heston-Hull-White characteristic function in Section 3.10 was implemented in Matlab. The next step involved solving the infinite integral in the Fourier transform pricing formula, $\int_0^\infty Re\left\{e^{-ivk}\psi(v)\right\}dv$, via numerical integration.

7.1.7 Choosing a suitable upper bound, $v_{\max}$, for the integral, and discretising the integration grid uniformly, the integral was evaluated using the trapezoidal rule of integration. We set $v_{\max} = 50$ and $dv = 0.01$ and used the trapz[9] function in Matlab.

7.1.8 For the characteristic function, we fixed the parameters $\{\kappa, \lambda, \eta, \rho_{s,r}\}$, and left the parameters $\{v_0, \overline{v}, \sigma, \rho_{s,v}\}$ free to estimate, bounded by the constraints in Ω^{Search}. We used the least squares optimisation algorithm lsqnonlin[10] in Matlab to minimise the sum of squared differences between the model and European call option prices. We fixed $\kappa = 1$ since the calibration results became unstable when we left this parameter free to estimate. Similar findings were reported by van Dijk et al. (2018).

7.1.9 The details of the calibration for the Heston-Hull-White model to the FTSE/JSE TOP40 volatility surface as at 16 November 2020 are shown in Table 1.

TABLE 1. Calibration details as at 16 November 2020

Number of strikes included	21
Number of maturities included	3
Number of options included	63
θ	0.07
λ	0.05
η	0.02
$\rho_{s,r}$	0.3
κ	1
v_0	Free to estimate
$\overline{v}$	Free to estimate
σ	Free to estimate
$\rho_{s,v}$	Free to estimate

9 https://www.mathworks.com/help/matlab/ref/trapz.html
10 https://www.mathworks.com/help/optim/ug/lsqnonlin.html

Using the information in Table 1 to calibrate the Heston-Hull-White model to the implied volatility surface in Figure 8, we estimated the parameters $v_0 = 0.0433$, $\overline{v} = 0.05$, $\sigma = 0.3817$, and $\rho_{s,v} = -0.9208$. Figure 9 shows the calibration results compared to the market data.

7.1.10 Figure 9 shows that the Heston-Hull-White model reproduces the European call option prices well. The model is therefore aligned with the guidelines set by APN 110 in terms of being market-consistent.

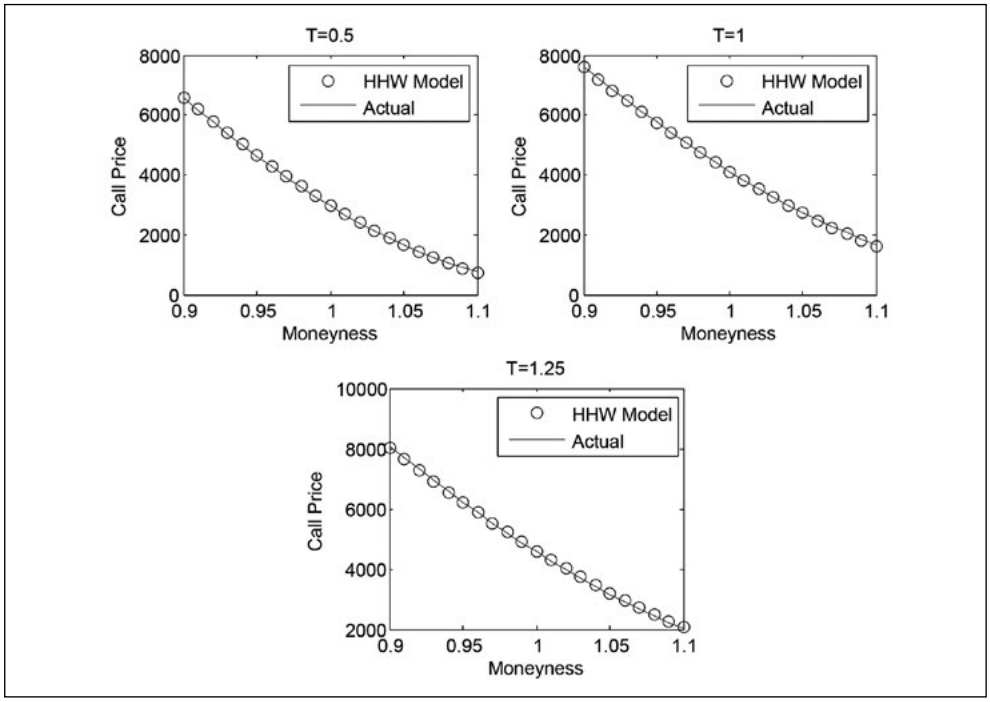

FIGURE 9. Heston-Hull-White model fit to market data on 16 November 2020

7.1.11 As mentioned earlier, the South African equity options market is short-dated. From Figure 8, the longest maturity observable is 15 months. Since the GMMB and GMDB liabilities have maturities that extend far beyond this point, we must test whether the calibrated Heston-Hull-White model produces a plausible volatility term structure for maturities beyond 15 months. Figure 10 shows the volatility term structure for at-the-money options as at 16 November 2020.

7.1.12 From Figure 10, the 30-year annualised volatility produced by the Heston-Hull-White model as at 16 November 2020 is approximately 30%. This is a reasonable estimate and falls within the range of values reported by Flint et al. (2014) in their testing of various models. However, more advanced methods for long-term volatility estimation are available and the reader is referred to Flint et al. (2014) for further detail.

7.1.13 In order to price the GMMB and GMDB liabilities, we use a brute force Monte Carlo simulation to generate sample paths for each state variable. We use an Euler

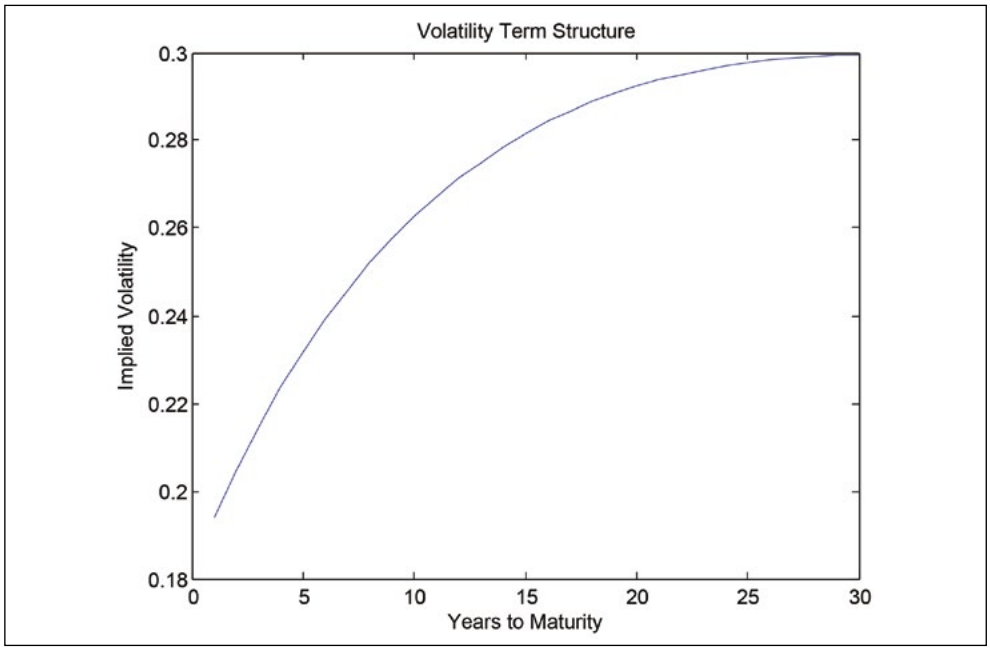

FIGURE 10. Heston-Hull-White volatility term structure as at 16 November 2020

discretisation scheme to discretise the time grid. The discretised Heston-Hull-White model is given by:

$$\begin{cases} S_{t+1} &= S_t + r_t S_t \delta_t + \sqrt{v_t} S_t \left(W_{t+1}^s - W_t^s \right), & S_0 > 0, \\ v_{t+1} &= v_t + \kappa\left(\overline{v} - v_t\right)\delta_t + \sigma\sqrt{v_t}\left(W_{t+1}^v - W_t^v \right), & v_0 > 0, \\ r_{t+1} &= r_t + \lambda\left(\theta - r_t\right)\delta_t + \eta\left(W_{t+1}^r - W_t^r \right), & r_0 \in \mathbb{R}, \end{cases}$$

where δ_t denotes the spacing of the time grid, and $W_{t+1} - W_t \sim N\left(0, \delta_t\right)$. To generate correlated Brownian motion increments for the state variables, we use Cholesky factorisation.[11]

7.1.14 Using a time spacing of $\delta_t = 1/252$, Figure 11 shows a single simulation from the Heston-Hull-White model for each state variable.

7.1.15 In Figure 11, the left y-axis refers to the share price, and the right y-axis refers to the volatility and interest rate. Note the inverse relationship between the volatility and share price. From our calibration, we estimated $\rho_{s,v} = -0.9208$, hence, the inverse relationship between the two state variables. This relationship is generally observed in equity markets and Black (1976) refers to this as the leverage effect. Recall that $\rho_{v,r} = 0$ so that the volatility and interest rate processes move independently. In our pricing experiment, we test the sensitivity of the GMMB and GMDB prices to various Heston-Hull-White model parameters.

11 https://www.mathworks.com/help/matlab/ref/chol.html

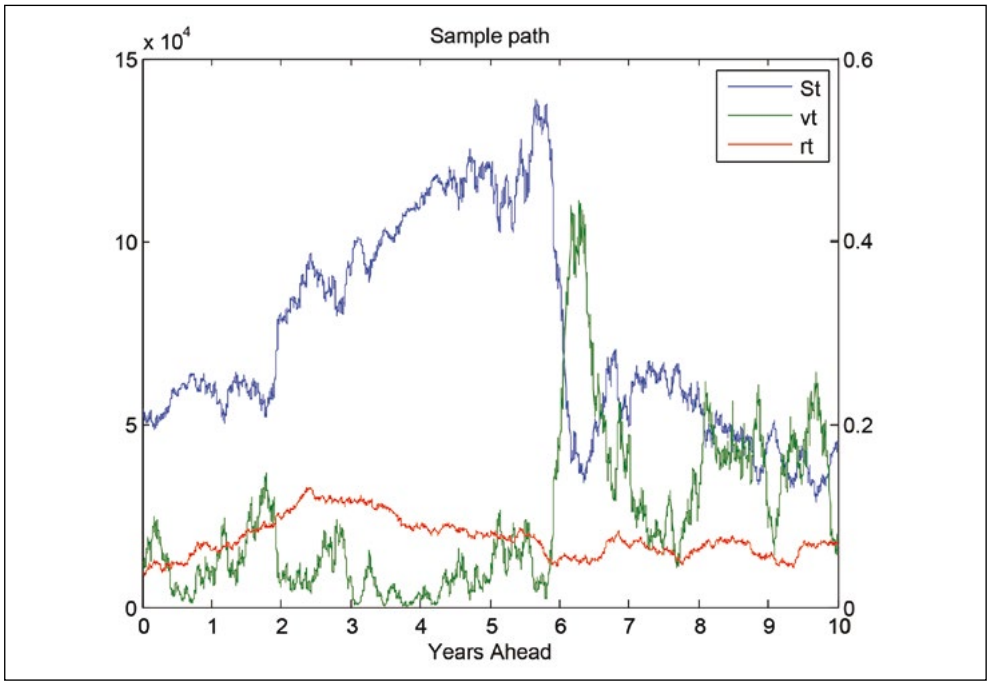

FIGURE 11. Single sample path for each state variable

7.1.16 As mentioned, the CIR++ model is a term structure model that can fit the input survival probability curve. To illustrate the mechanics of the CIR++ model, we consider a 50-year old South African male and female pensioner with parameters shown in Table 2.

TABLE 2. CIR++ parameters for 50-year old male and female pensioners

Parameter	Male	Female
γ	0.90	1.00
ω	0.05	0.03
η	0.03	0.01

7.1.17 Figure 12 shows the simulation for the force of mortality, $\mu_{50+t,t}$, for a 50-year old South African male and female pensioner from the CIR++ model. Note that the force of mortality increases with age. Furthermore, the distribution for $\mu_{50+t,t}$ is wider for males than females. This is driven by the volatility of mortality, set to $\eta = 0.03$ for males and $\eta = 0.01$ for females.

7.1.18 Figure 13 shows the stochastic survival probability curves calculated from the formula $e^{-\int_0^t \mu_{50+s,s} ds}$.

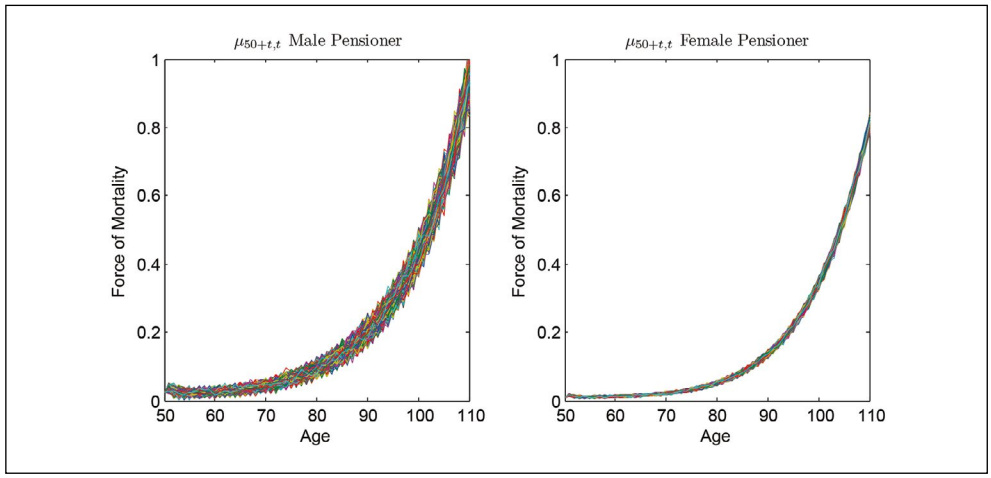

FIGURE 12. CIR++ simulation for force of mortality

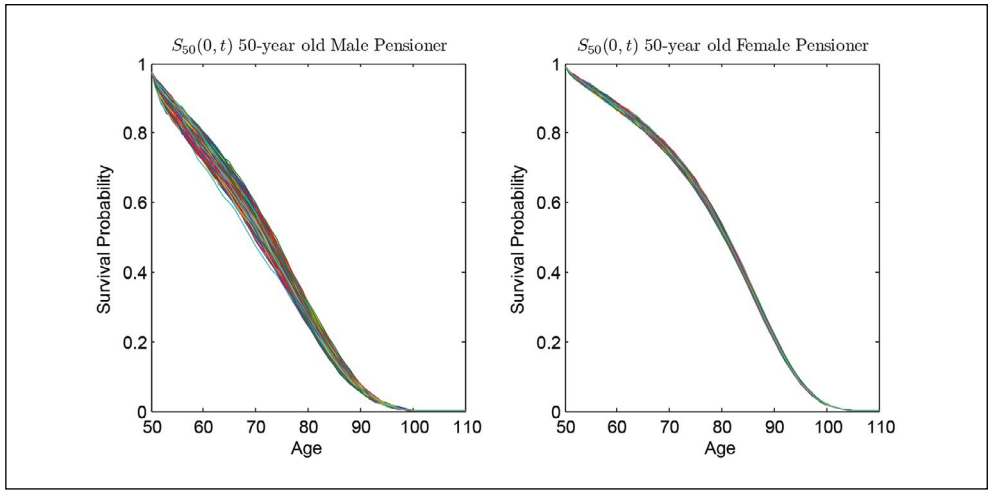

FIGURE 13. Stochastic survival probability curves

7.1.19 Figure 14 shows the mean survival probability curve, $\mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^t \mu_{50+s,s}ds}\right]$, compared to the input survival probability curve for a 50-year old South African male and female pensioner. Figure 14 shows that the CIR++ model reproduces the input survival probability curves. Although the CIR++ model has the attractive feature of matching the supplied survival probability curve and incorporating volatility for mortality, we recommend that further research be done to test the impact of non-mean reverting models on mortality.

7.1.20 To illustrate the performance of the AR(1)-ARCH(1) model, we calibrated the model to yearly changes in the log mortality rate for the UK female population in Figure 7.

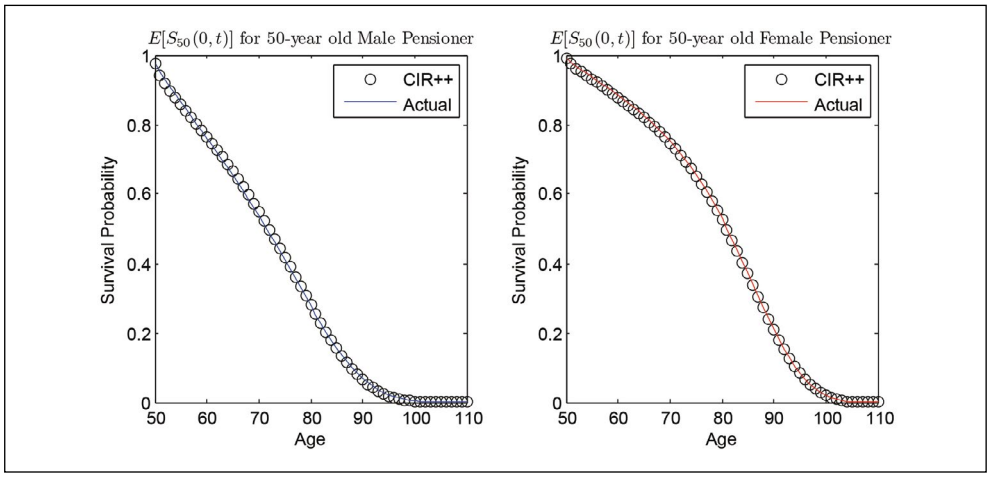

FIGURE 14. Mean survival probability curve versus input survival probability curve

For each fixed age, we estimated the parameters over the period 1922 to 2020 using the ar[12] and garch[13] packages in Matlab. The calibrated parameters are shown in Figure 15.

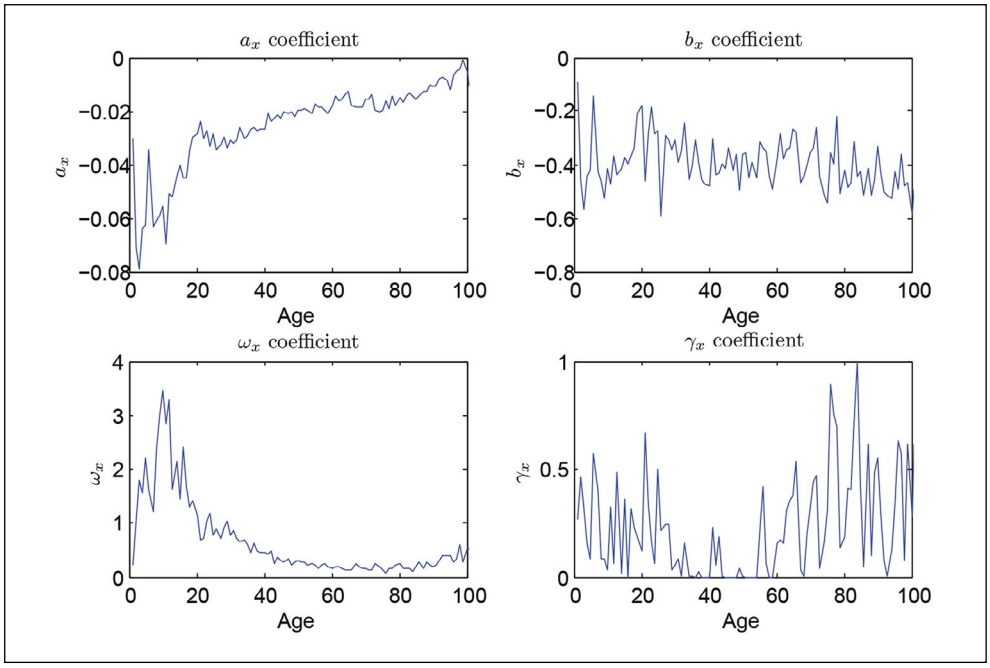

FIGURE 15. Calibrated AR(1)-ARCH(1) parameters for UK females

12 https://www.mathworks.com/help/ident/ref/ar.html
13 https://www.mathworks.com/help/econ/garch.html

7.1.21 Figure 16 shows the predicted force of mortality for each year from 1922 to 2020 for an 85-year old UK female. Visually, the AR(1)-ARCH(1) model predicts the force of mortality well. Furthermore, the actual mortality rates generally lie within the 95% confidence interval.

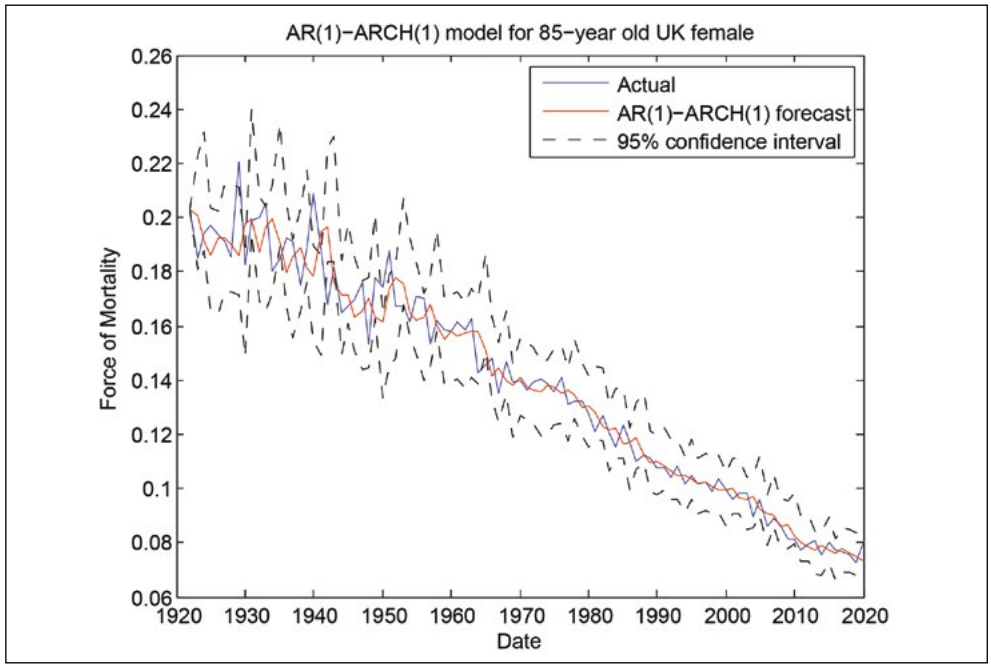

FIGURE 16. AR(1)-ARCH(1) forecast for 85-year old UK female

7.1.22 To test the goodness-of-fit, the standardised residuals (assuming normal innovations for the AR(1)-ARCH(1) model) are compared to the standard normal density function in Figure 17.

7.1.23 The Jarque-Bera test can be used to test the normality of the data—the null hypothesis is that the standardised residuals are normally distributed. We calculated the p-value for the Jarque-Bera test as 0.08, hence, we do not reject the null hypothesis at a 5% level of significance. We conclude that the AR(1)-ARCH(1) model with normally distributed innovations captures the mean and variance for 85-year old UK females well.

7.1.24 To construct a survival probability curve, the AR(1)-ARCH(1) model is fit to a time series of mortality rates for each fixed age using the age specific parameters in Figure 15.

7.1.25 As mentioned, we are not aware of historical mortality data that is publicly available in South Africa. Therefore, we showed how to calibrate the AR(1)-ARCH(1) model to mortality data sourced from the UK. The purpose was to show the mechanics of the AR(1)-ARCH(1) model, and the reader should note that the model can easily be applied to historical South African mortality rates by just changing the input data.

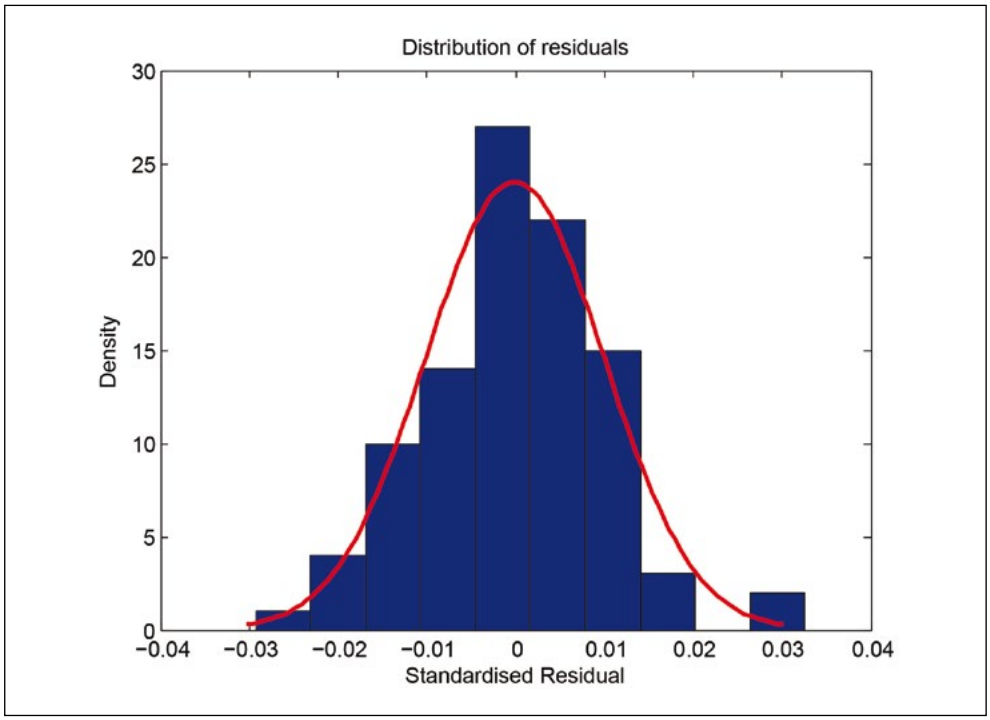

FIGURE 17. Standardised residuals versus standard normal density

7.1.26 The focus of this paper is specifically on the South African market, hence, we apply the CIR++ model going forward since survival probability curves for South African male and female pensioners are readily available from the CSI Committee's "Report on pensioner mortality 2005–2010".

7.1.27 Before we present the sensitivity of the GMMB and GMDB prices to various Heston-Hull-White-Mortality model parameters, we first create a base set of parameters. We use the parameters from Tables 1 and 2 as our base. Table 3 shows the base parameters for a 50-year old male and female pensioner in South Africa.

7.1.28 In the results below, we run 10,000 simulations and consider a single premium of R100,000 invested in the GMMB and GMDB respectively. It is simple to extend the analysis to take into account recurring premiums, see Feng (2018). We report our results as a ratio of the GMMB and GMDB price to the initial premium invested at $t = 0$:

$$R_{GMMB} = \frac{V_{GMMB,0}}{Initial \quad Premium},$$

$$R_{GMDB} = \frac{V_{GMDB,0}}{Initial \quad Premium}.$$

TABLE 3. Base parameters as at 16 November 2020

Hull-White component	θ	0.07
	r_0	0.04
	λ	0.05
	η	0.02
	$\rho_{s,r}$	0.3
Heston component	κ	1.00
	v_0	0.04
	$\overline{v}$	0.05
	σ	0.38
	$\rho_{s,v}$	–0.92
CIR++ component	$\mu_{50,0,Male}$	0.02
	$\mu_{50,0,Female}$	0.01
	γ_{Male}	0.90
	γ_{Female}	1.00
	ω_{Male}	0.05
	ω_{Female}	0.03
	ξ_{Male}	0.03
	ξ_{Female}	0.01

7.1.29 First, we consider three scenarios for the GMMB price from a life insurer's perspective: 1) deterministic interest rates by setting $\eta = 0$; 2) an independent interest rate process by setting $\rho_{s,r} = 0$; and 3) constant equity volatility by setting $\sigma = 0$. We keep all other parameters in Table 3 fixed. Figure 18 shows R_{GMMB} for all three scenarios with a guaranteed rate of $g = 6\%$ per annum, where $G = G_0(1+g)^T$, with $G_0 = 100,000$.

7.1.30 From Figure 18, we observe the following:

— From a life insurer's perspective, a 6% guarantee per annum relates to a life contingent put option that is written in-the-money, hence, there may be a significant liability;
— The GMMB price for a 50-year old female is higher than the price for a 50-year old male. This is driven solely by the fact that the survival probability for South African females is higher than males;
— Deterministic interest rates, $\eta = 0$, yield GMMB prices that are significantly lower than the base model. The difference is amplified as the term to maturity increases;
— Independent equity and interest rates, $\rho_{s,r} = 0$, yield GMMB prices that are lower than base model, but higher than the deterministic interest rate model; and
— Constant equity volatility, $\sigma = 0$, yields GMMB prices that are very close to the base model.

Our findings suggest that interest rate risk is the dominating factor when pricing GMMB products. Assuming deterministic rates may lead to GMMB prices that are significantly understated and, hence, pose a large risk to the life insurer. Assuming independent interest rate

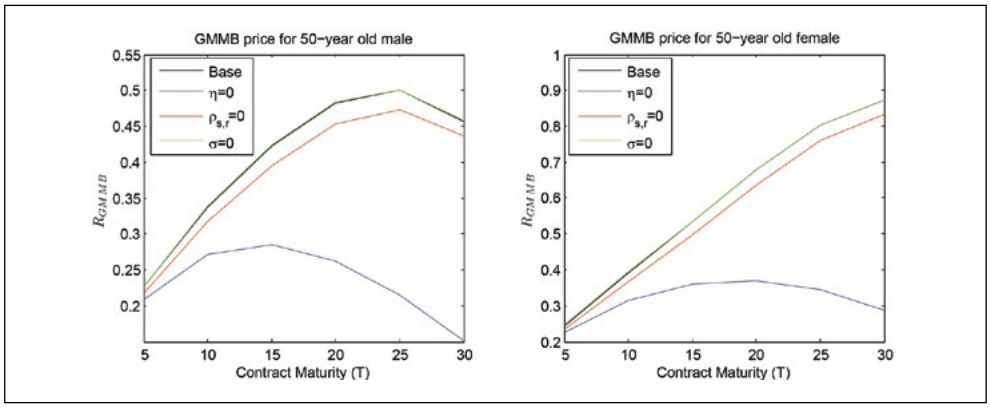

FIGURE 18. GMMB price for a 50-year old South African pensioner

and equity processes may also not be sufficient. An interesting observation is that stochastic volatility makes nearly no impact on the price of the GMMB.

7.1.31 Next, we present three scenarios for the GMDB price from a life insurer's perspective: 1) deterministic interest rates by setting $\eta = 0$; 2) an independent interest rate process by setting $\rho_{s,r} = 0$; and 3) constant equity volatility by setting $\sigma = 0$. As with the GMMB product, we keep all other parameters in Table 3 fixed. Figure 19 shows R_{GMDB} for all three scenarios with a guaranteed roll-up rate of $g = 6\%$ per annum, where $G_t = G_0 e^{gt}$, with $G_0 = 100,000$.

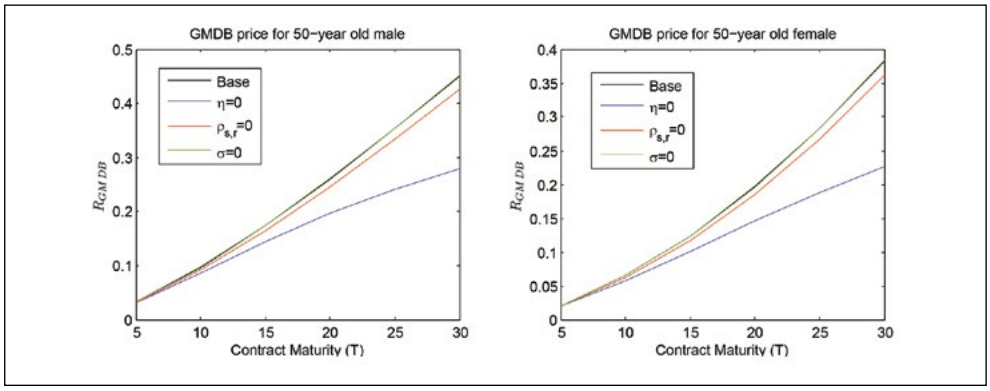

FIGURE 19. GMDB price for a 50-year old South African pensioner

7.1.32 From Figure 19, we observe the following:

— The GMDB price increases as the term to maturity increases, whereas the GMMB price starts to decrease as the contract maturity becomes larger;
— The GMDB price for a 50-year old male is higher than the price for a 50-year old female. This is because the force of mortality for South African males is higher than females;

— Deterministic interest rates, $\eta = 0$, yield GMDB prices that are considerably lower than the base model. The difference is amplified as the term to maturity increases;
— Independent equity and interest rates, $\rho_{s,r} = 0$, yield lower prices than the base model, but higher prices than the deterministic interest rate model; and
— Constant volatility, $\sigma = 0$, yields GMDB prices that are very close to the base model.

Once again, our findings suggest that interest rates are the dominating factor when pricing GMDB products.

7.1.33 Stochastic mortality has application in risk management problems and can be applied to Value-at-Risk (VAR) (see Syuhada & Hakim, 2021) and longevity stress testing (see Browne et al., 2009), for example.

7.1.34 In the next section, we apply the Heston-Hull-White-Mortality model to hedge the GMMB.

7.2 Hedging

7.2.1 Standard Black-Scholes (1973) theory suggests that a contingent claim can be replicated with a combination of shares and cash. This is referred to as delta-hedging. Since the GMMB and GMDB payoffs resemble the payoff of a put option (although life contingent), Black-Scholes (1973) theory can be applied to hedge these products.

7.2.2 Let Δ_t denote the number of shares that an insurer must hold in a hedging portfolio to hedge against movements in the underlying equity index. Furthermore, let B_t denote the money market account from which money can be borrowed or invested. No upfront premium is payable for the GMMB. Therefore, the value of the money market account at $t = 0$ is given by:

$$B_0 = -\Delta_0 S_0, \tag{18}$$

where $\Delta_0 \in [-1, 0]$ in the case of a put option.

7.2.3 For $t > 0$, the money market account grows at the risk-free rate, r. We assume that the risk-free rate is the 3-month T-Bill rate in South Africa. The change in the money market account over a discrete time interval is given by:

$$B_t = B_{t-1} e^{r\delta t} - (\Delta_t - \Delta_{t-1}) S_t, \quad B_0 = -\Delta_0 S_0. \tag{19}$$

Note that we consider a discrete hedging portfolio since it is impractical to hedge continuously in practice. At every time step, the hedging portfolio must be rebalanced to reflect the sensitivity of the derivative to the underlying asset price as new information becomes available.

7.2.4 The hedging portfolio at time t consists of a combination of shares and cash and is given by:

$$\Pi_t = \Delta_t S_t + B_t. \tag{20}$$

7.2.5 In order to calculate Δ_t in the Heston-Hull-White-Mortality model, we approximate Δ_t with a forward finite difference scheme as follows:

$$\Delta_t \approx \frac{V_{GMMB}\left(t, S_t + 0.01, v_t, r_t, \mu_t\right) - V_{GMMB}\left(t, S_t, v_t, r_t, \mu_t\right)}{0.01}. \tag{21}$$

7.2.6 Before we proceed with our backtesting experiment, we calibrate the Heston-Hull-White-Mortality model to weekly FTSE/JSE TOP40 implied volatility surfaces from September 2005 to November 2020. The calibrated parameters are shown in Figure 20. Note the clear spikes in volatility during the global financial crisis of 2007/2008 and the COVID-19 pandemic. In the calibration, we set $\kappa = 1$ since an unrestricted κ produced unstable results.

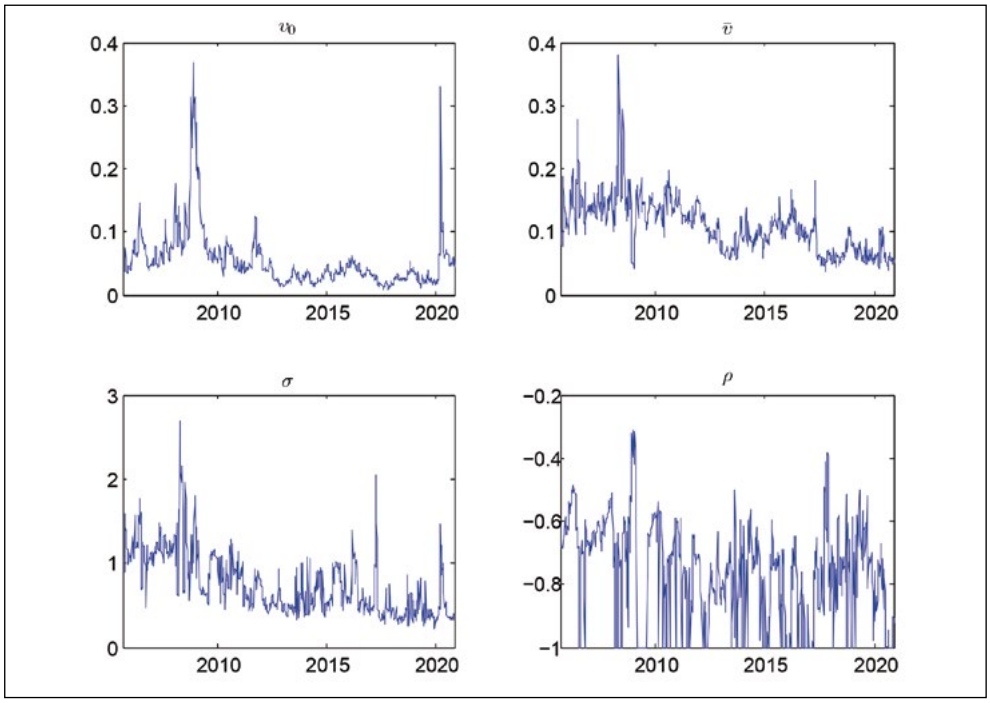

FIGURE 20. Calibrated parameters over time

7.2.7 For this experiment, we rebalance the hedging portfolio once every week. We consider a 6% per annum GMMB product sold to a 50-year old male pensioner in September 2005, expiring in November 2020. We further assume that the investor is alive at each valuation date for this backtesting experiment.

7.2.8 Figure 21 shows the value of the GMMB liability over time calculated from the Heston-Hull-White-Mortality model. Note the extreme volatility in the GMMB price over time. The volatility is driven by movements in the underlying FTSE/JSE TOP40 index, interest rates, and mortality. Furthermore, the GMMB liability displays significant jumps in the 2007/2008 global financial crisis and COVID-19 periods. Hedging is a key tool that can be used by life insurers to manage the risk of selling embedded derivatives.

7.2.9 Figure 22 shows the value for Δ over time calculated from equation (21). In periods of stress and poor equity performance, Δ approaches –1. This means that equity

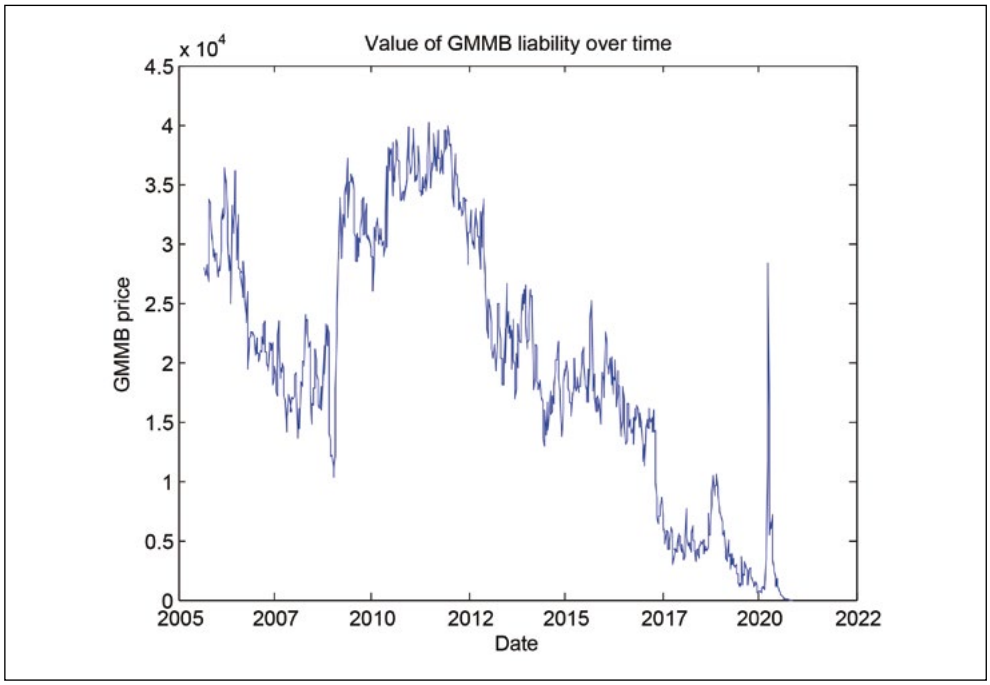

FIGURE 21. GMMB liability over time

exposure in the hedging portfolio increases. Note that Δ is close to –1 in the time of the global financial crisis and the COVID-19 pandemic. Alternatively, when equity performs well, Δ approaches 0 (equity exposure decreases). Hedging the GMMB requires active management by consistently updating the allocation in a short equity position and cash.

7.2.10 From Feng (2018), we define $V^*_{GMMB,t}$ as:

$$V^*_{GMMB,t} := V_{GMMB,t} - V_{GMMB,0}e^{rt}. \tag{22}$$

$V^*_{GMMB,t}$ represents the deviation from the risk-neutral expectation of the GMMB calculated at inception of the contract. Figure 23 shows the performance of the delta-hedging strategy in equation (20) compared to equation (22).

7.2.11 Delta-hedging captures the general trend of the movements in the GMMB liability, but fails to capture the volatility and jumps. This makes sense since the market is incomplete under the Heston-Hull-White-Mortality model. Volatility, interest rates, and mortality are all factors that are not directly tradable in the South African market. To capture the effect of volatility, interest rates, and mortality on the GMMB liability, more products that derive their value from these risk factors need to be added to the hedging portfolio.

7.2.12 Although not perfect, delta-hedging does seem to provide some benefit for the GMMB liability. The hedging strategy reduces the variability of GMMB liability since it captures the downward trend through time.

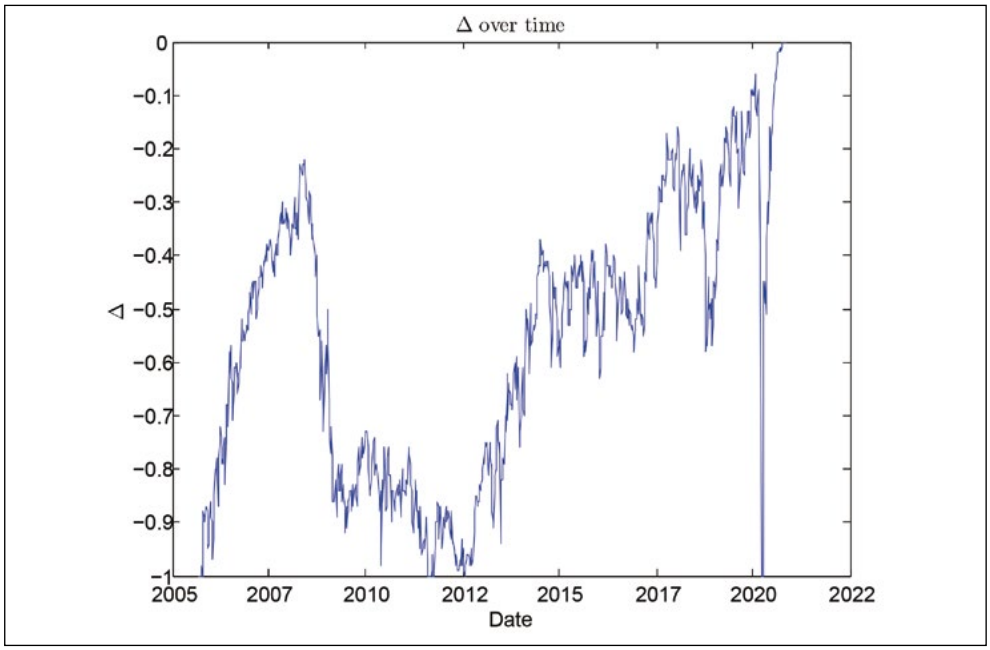

FIGURE 22. GMMB Δ over time

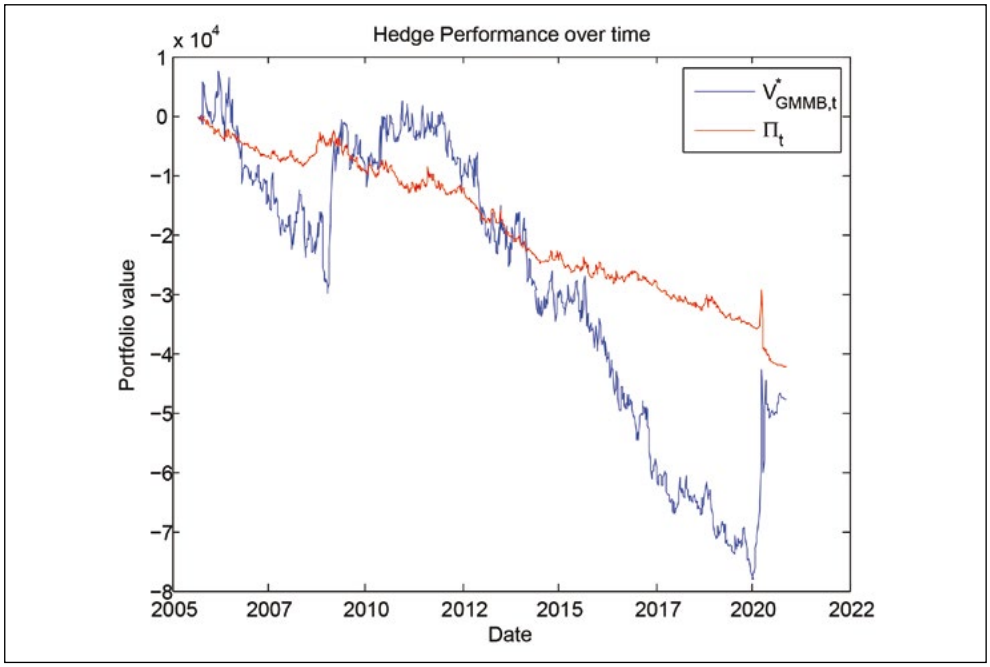

FIGURE 23. Hedge performance over time

7.2.13 Hedging is a complex problem and highly relevant in the life insurance industry. Therefore, further research in the field of embedded derivative hedging is of significant importance.

8. CONCLUSION

8.1 In this paper, we showed the importance of incorporating stochastic interest rates when reserving for GMMB and GMDB products since interest rate risk is the dominating factor when pricing these claims. Ignoring interest rate risk can lead to prices that are significantly understated from a life insurer's perspective. Stochastic volatility had no significant impact on the GMMB and GMDB liabilities.

8.2 Furthermore, we proposed the continuous-time CIR++ and discrete-time AR(1)-ARCH(1) models for mortality. The stochastic mortality models performed very well. The CIR++ model is able to reproduce the input survival probability curve, whereas the AR(1)-ARCH(1) model provides a good fit to historical data. Both models can be applied to risk management problems such as VAR and longevity stress testing, and we suggest further research on this topic.

8.3 Insurance liabilities change through time as the underlying risk factors change. To hedge these risks, we implemented a simple delta-hedging strategy based on the Heston-Hull-White-Mortality model and showed that a delta-hedging strategy captures the general trend of the movements in the liability value. However, under the Heston-Hull-White-Mortality model, the market is incomplete. Further research should be conducted to test whether a volatility (delta-vega) or interest rate (delta-rho) hedging strategy can produce better hedging results than a simple delta-hedging strategy. This is a significant challenge faced in the life insurance industry and finding a good hedge may lead to less volatility in the profit and loss (P&L) in the insurer's income statement.

8.4 The proposed Heston-Hull-White-Mortality model aligns with the requirements set by APN 110 and is a good starting point for life insurers when deciding on a model for embedded derivatives.

ACKNOWLEDGEMENTS

The authors would like to thank Emlyn Flint from Peresec for providing us with historical FTSE/JSE TOP40 volatility surface data.

REFERENCES

Ballotta, L, Eberlein, E, Schmidt, T & Zeineddine, R (2020). Variable annuities in a Lévy-based hybrid model with surrender risk. *Quantitative Finance* **20**(5), 867–86

Black, F & Scholes, M (1973). The pricing of options and corporate liabilities. *Journal of Political Economy* **81**(3), 637–54

Black, F (1976). Studies of stock price volatility changes. Proceedings of the 1976 Meeting of the American Statistical Association, Business and Economic Statistics Section 177–81

Brigo, D & Mercurio, F (2001). On deterministic shift extensions of short rate models. Available at SSRN: https://ssrn.com/abstract=292060 or http://dx.doi.org/10.2139/ssrn.292060

Browne, B, Duchassaing, J & Suter, F (2009). Longevity: A 'simple' stochastic modelling of mortality. *British Actuarial Journal* **15**(S1), 249–65

Cairns, A, Blake, D & Dowd, K (2008). Modelling and management of mortality risk: A review. *Scandinavian Actuarial Journal* **2**(3) ,79–113

Carr, P & Madan, D (1999). Option valuation using the fast Fourier transform. *Journal of Computational Finance* **2**(4), 61–73

Cont, R (2001) Empirical properties of asset returns stylized facts and statistical issues. *Quantitative Finance* **1**(2), 223–36

Cox, J, Ingersoll, J & Ross, S (1985). A theory of the term structure of interest rates. *Econometrica* **53**(2), 385–408

Duffie, D, Pan, J & Singleton, K (2002). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* **68**(6), 1343–76

Fang, F & Janssens, B (2007). Characteristic function of the hybrid Heston–Hull–White model. In: Proceedings of the 58th European Study Group Mathematics with Industry. Utrecht. http://www.math.uu.nl/swi2007/proc2007.pdf. pp 107–15

Fang, F & Oosterlee, C (2008). A novel pricing method for European options based on Fourier-Cosine series expansions. *SIAM Journal on Scientific Computing* **31**(2), 826–48

Feng, R (2018). *An introduction to computational risk management of equity-linked insurance*. Taylor & Francis

Flint, E & Maré E (2017). Estimating option-implied distributions in illiquid markets and implementing the Ross recovery theorem. *South African Actuarial Journal* **17**(1), 1–28

Flint, E, Ochse, E & Polakow, D (2014). Estimating long-term volatility parameters for market-consistent models. *South African Actuarial Journal* **14**(1), 19–72

Giacometti R, Betocchi M, Rachev S & Fabozzi F (2012). A comparison of the Lee-Carter model and AR-ARCH model for forecasting mortality rates. *Insurance: Mathematics and Economics* **50**(1), 85–93

Grzelak, L & Oosterlee, C (2011). On the Heston model with stochastic interest rates. *SIAM Journal on Financial Mathematics* **2**(1), 255–86

Gurrieri, S, Nakabayashi, M & Wong, T (2009). Calibration methods of Hull-White model. Available at SSRN: https://ssrn.com/abstract=1514192 or http://dx.doi.org/10.2139/ssrn.1514192

Hardy, M (2003). *Investment guarantees modelling and risk management for equity-linked life insurance.* John Wiley & Sons

Heston, S (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies* **6**(2), 327–43

Hull, J & White, A (1990). Pricing interest rate derivatives securities, *Review of Financial Studies* **3**(4), 573–92

Ignatieva, K, Song, A & Ziveyi, J (2016). Pricing and hedging of guaranteed minimum benefits under regime-switching and stochastic mortality. *Insurance: Mathematics and Economics* **70**(C), 286–300

in 't Hout, K, Bierkens, J, van der Ploeg, A & in 't Panhuis, J (2007). A semi closed-form analytic pricing formula for call options in a hybrid Heston–Hull–White model. In: Proceedings of the 58th European Study Group Mathematics with Industry. Utrecht. http://www.math.uu.nl/swi2007/proc2007.pdf. pp 101–6

Kammeyer, H & Kienitz, J (2009). An implementation of the hybrid-Heston-Hull-White model. Available at SSRN: https://ssrn.com/abstract=1399389 or http://dx.doi.org/10.2139/ssrn.1399389

Kammeyer, H & Kienitz, J (2012). The Heston-Hull-White model Part I: Finance and analytics. *Wilmott Magazine* 2012(57), 46–53

Lee, R & Carter, L (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association* **87**(419), 659–71

Lin T, Wang C & Tsai C (2015). Age-specific copula-AR-GARCH mortality models. *Insurance: Mathematics and Economics* **61**, 110–24

Maze, S (2014). Efficient implementation of the Heston-Hull & White model. Available at SSRN: https://ssrn.com/abstract=2378955 or http://dx.doi.org/10.2139/ssrn.2378955

Muskulus, M (2007). Three approaches to extend the Heston model. Proceedings of the 58th European Study Group Mathematics with Industry. Utrecht. http://www.math.uu.nl/swi2007/proc2007.pdf. pp 93–100

Ngugi, A, Maré, E & Kufakunesu, R (2015). Pricing variable annuity guarantees in South Africa under a variance-gamma model. *South African Actuarial Journal* **15**(1), 131–70

Oosterlee, C (2007). The ING problem: a problem from the financial industry. In: Proceedings of the 58th European Study Group Mathematics with Industry. Utrecht. http://www.math.uu.nl/swi2007/proc2007.pdf. pp 91–2

Patel, R (2019). Approximating the Heston-Hull-White model. MPhil thesis. University of Cape Town. Available at: http://hdl.handle.net/11427/30881 (Accessed: 31 January 2021)

Syuhada, K & Hakim, A (2021). Stochastic modeling of mortality rates and mortality-at-risk forecast by taking conditional heteroscedasticity effect into account. *Heliyon* **7**(10), e08083. doi: 10.1016/j.heliyon.2021.e08083

Truter, G (2012). The valuation and hedging of default-contingent claims in multiple currencies. MSc thesis. University of the Witwatersrand. Available at: http://hdl.handle.net/10539/11955 (Accessed: 31 January 2021)

van Dijk, M, de Graaf, C & Oosterlee, C (2018). Between $\mathbb{P}$ and $\mathbb{Q}$: The $\mathbb{P}^{\mathbb{Q}}$ measure for pricing in asset liability management. *Journal of Risk and Financial Management* **11**(4), 67

Vasicek, O (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* **5**(2), 177–88

Veilleux, P (2016). On the impact of stochastic volatility, interest rates and mortality on the hedge efficiency of GLWB guarantees. MSc thesis. Laval University. Available at: http://hdl.handle.net/20.500.11794/26668 (Accessed: 31 January 2021)

Wang, G (2011). An equity and foreign exchange Heston-Hull-White model for variable annuities. MSc thesis. Delft University of Technology. Available at: http://resolver.tudelft.nl/uuid:b8b2276c-aeee-49ed-b9b7-0bc78b794d76 (Accessed: 31 January 2021)

