Bayesian Value-at-Risk backtesting: The case of annuity pricing

Melvern Leung[a], Youwei Li[b], Athanasios A. Pantelous[a,*], Samuel A. Vigne[c]

[a] *Department of Econometrics and Business Statistics, Monash Business School, Monash University, Wellington Road, Clayton Campus, VIC 3800, Australia*
[b] *Hull University Business School, University of Hull, Cottingham Rd, Hull HU6 7RX, United Kingdom*
[c] *Trinity Business School, The University of Dublin, Luce Hall, Pearse St, Dublin 2, D02 H308, Ireland*

Abstract

We propose new Unconditional, Independence and Conditional Coverage VaR-forecast backtests for the case of annuity pricing under a Bayesian framework that significantly minimise the direct and indirect effects of p-hacking or other biased outcomes in decision-making, in general. As a consequence of the global financial crisis during 2007–09, regulatory demands arising from Solvency II has required a stricter assessment setting for the internal financial risk models of insurance companies. To put our newly proposed backtesting technique into practice we employ linear and nonlinear Bayesianised variants of two typically used mortality models in the context of annuity pricing. In this regard, we explore whether the stressed longevity scenarios are enough to capture the experienced liability over the forecasted time horizon. Most importantly, we conclude that our Bayesian decision theoretic framework quantitatively produce a strength of evidence favouring one decision over the other.

Keywords: Decision analysis; Value-at-Risk; Backtesting; Bayesian framework; Longevity risk

*Corresponding author: Office number: +61(0)399054718, E-mail: Athanasios.Pantelous@monash.edu
Email addresses: Melvern.Leung@monash.edu (Melvern Leung), Youwei.Li@hull.ac.uk (Youwei Li), Athanasios.Pantelous@monash.edu (Athanasios A. Pantelous), vignes@tcd.ie (Samuel A. Vigne)

1. Introduction

After the global financial crisis of 2007–09, regulatory demands arising from Solvency II have required a very strict assessment for internal financial risk models for insurance companies. In practice, the idea of backtesting implies identifying whether risk measurement models are able to accurately determine the risk exposures experienced (e.g., Drenovak et al., 2017). The *Unconditional Coverage* (UC) test was the first *Value-at-Risk* (VaR) backtest developed by Kupiec (1995) to determine whether the risk-measure forecast is able to capture the realized returns at the q-quantile. Several other key innovations include the works of Christoffersen (1998) and Ziggel et al. (2014) who try to relax assumptions in the UC test or to improve the statistical power of the test itself. More specifically, Christoffersen (1998) introduces the idea of the *Independence* test which determines whether sequences of one's and zero's occur independently or clustered together, and the *Conditional Coverage* (CC) test which combines both the Independence and UC tests to establish whether or not sequences of ones and zero's occur at the specified coverage and occur independent to each other.

However, one significant drawback of these backtests is commonly known as the *null hypothesis significance testing* (NHST). As stated and convincingly discussed at the American Finance Association president's address by Harvey (2017), the issue with NHST is that hypothesis testing, which is a significant tool used extensively in the finance literature, and its testing procedures are based on the critical assumption that the null hypothesis is true, and the alternative hypothesis is indirectly inferred. Furthermore, data dredging (known also as data snooping or p-hacking) might exist, and thus dramatically increase the risk of false positive rates. This, in turn, necessitates the development of a backtest under a Bayesian setting, where decision making is based on a measure of strength rather than on a conditioned false positive rate, which as of yet, has not received much attention in the literature.[1]

With the recent developments of Bayesian statistical techniques, there is an increasing motion towards the use of a Bayesian decision framework in hypothesis testing. Bayesian testing began when Berger and Sellke (1987) developed the idea of *Bayes Factor* (BF) to determine a ratio of evidence, and recently, this idea has been extensively applied in different scientific fields. Another method which we consider is the Bayesian Likelihood Ratio test developed by Li et al. (2014), who introduce the idea of averaging the likelihood function over the posterior distribution, as opposed to

[1]For a review of the objectivity that Bayesians have to NHST (see, Harvey, 2017, pp. 1421–1422, and references therein).

the BF which averages over the prior distribution.[2] These two methods are applied to the UC test to facilitate testing under the Bayesian framework. Additionally, for completeness, we show how the BF can be extended to the Independence and CC backtests.[3] We then use the idea of backtesting in annuity liabilities, and since our focus is on the long term longevity trend risk, a longevity stress test scenario is developed accordingly. Under this setting, it is crucial that a suitable backtesting method is identified to determine if the underlying longevity risks associated with a pricing instrument are in fact captured by a mortality model. In essence, the backtesting procedure hinges on the outcome of whether the specified mortality model will be able to produce forecasts such that obligated payments from an annuity can be met with a probability of 99.5%.[4]

The paper is organised as follows. Section 2 presents the related literature in detail. Three distinctive contributions in the corresponding literature are presented in Section 3, and correspondingly, Section 4 is then divided into three main parts. Specifically, in Section 4.1 and Appendix A, we explain the estimation procedure of the Bayesian mortality models under a state-space representation setting. We pay particular interest to the nonlinear dynamics when modelling death counts rather than the crude mortality estimates. Further, Appendices A, B and C (see also, the extensive SI provided) contain the empirical results of fitted LC and CBD models alongside the Bayesian forecast with credible intervals used in the paper. In Section 4.2, and Appendices D and E, we focus on developing the VaR Bayesian backtesting framework for the novel UC, Independent and CC tests. Then, in Section 4.3, we provide the Bayesian backtesting framework developed for longevity stressed scenarios under an annuity calculation. In Section 5 and Appendix F, we determine clearly which mortality model produces the most favourable results under a 99.5% longevity stressed scenario implementing the Solvency II regulation. Finally, Section 6 concludes the paper.

2. Related Literature

Over the past few decades, the popularity of VaR has increased significantly among the financial services industry for setting capital requirements for market risk measure. However, a risk measure is only as good as it is able to accurately predict future risks accordingly, and thus, to measure the accuracy needed when developing effective backtest procedures. These procedures should allow for

[2] Li et al. (2014) also show that this method has the advantage of being less sensitive across varying prior distributions.

[3] However, these tests will not be considered on the annuity pricing aspect, and a sequel paper will follow to assess mortality models and their connection with longevity risks.

[4] Our focus is on the annuity with contingent payments based on the policyholders lifespan, however we should emphasise here that the backtesting framework we develop can also be extended to measure accuracy of any type of risk measures.

the possibility to validate a risk measure given its out-of-sample forecasts and actual realized results (Christoffersen and Pelletier, 2004), such as those found using VaR.[5] By definition, the VaR is the q-quantile of a Profit/Loss distribution, and a backtesting mechanism is used to determine whether the required coverage q is indeed achieved. Although there has been a large emphasis on VaR forecasting in the literature (e.g., Berkowitz and O'Brien, 2002, Glasserman et al., 2002, Christoffersen, 2009, Nieto and Ruiz, 2016), the backtesting literature has since gained traction after the development of the UC backtest by Kupiec (1995). In practice, the idea of backtesting as explained by Kupiec (1995) is a type of "*reality check*" to identify whether risk measurement models are able to accurately determine the risk exposures experienced. Also noteworthy are the extensions to the Kupiec (1995) test, such as those from Christoffersen and Pelletier (2004) which aim at testing Independence of VaR violations, a CC test which combines both the UC and Independence test jointly, a duration approach in Christoffersen (2009), and consequently, Ziggel et al. (2014) and Wied et al. (2016) who aim to improve these tests by applying a new procedure that hinges on Monte Carlo simulations.

Bayesian statistics allows for a rigorous framework in which one can incorporate parameter uncertainty by choosing a prior for the parameter of interest. While many different methods can be used to capture parameter uncertainty, the Bayesian estimation method comes naturally via the Bayes theorem. In particular, with the inclusion of a prior and the combined data (or evidence), a posterior distribution for the parameters can be produced. Recently, with the advancements of Bayesian econometrics, the simulation of the posterior distribution can be obtained efficiently through *Monte-Carlo Markov-Chain* (MCMC) type algorithms. In addition, by combining these algorithms with the *Extended Kalman Filter* (EKF), we can efficiently sample the latent variables in block under a state-space framework. For this reason, we employ a Bayesian state-space framework that treats stochastic mortality models as state-space models (see Wang et al. (2020)). One major downfall with using crude mortality rates as the measured variable is that it is itself estimated from observed death counts and the total population (e.g., Pedroza, 2006, Kogure and Kurachi, 2010). Czado et al. (2005) focus on modelling death counts using Bayesian estimation but not a state-space framework.

In 2016, Solvency II was established with the aim of ensuring that insurance companies meet their obligatory payments. This regulation contains three pillars, and our focus in this paper is on pillar 2, where it contains the risk based solvency requirements. Pillar 2 allows room for internal

[5]While VaR is a widely used risk measure in finance (Jorion, 2000), and in several decision-making processes in general, other risk measures such as Stressed-trends, Expected Shortfalls, Conditional VaR, and an Extreme value approach by Longin (2000) can also be backtested. The results presented in this study can be extended in several other directions, however we will address them in our future research.

models in terms of assessing the financial stability of the insurance company, they have the choice to either use the capital requirements laid out by the supervisory regulators or to keep capital reserves based on their own risk-based models (Hari et al., 2008). There has been a vast amount of research undertaken in various aspects of the *Solvency Capital Requirements* (SCR) under the Solvency II regulation[6]. However, since the SCR hinges on a VaR measure, there currently exists a void in literature for backtesting VaR models under varying longevity stressed scenarios in tandem with Article 101 of Solvency II. Little preliminary work has been conducted testing VaR measures with longevity stressed scenarios. For example, Plat (2011) investigates whether or not longevity shock based models used in Solvency II were able to capture the 99.5% VaR in an annuity portfolio. While Plat (2011) utilizes the VaR measure, they neglect to tackle the issue of backtesting the actual risk measure itself.

3. Contributions

Our paper contributes to the literature in three distinctive ways. Firstly, as of 2016, Solvency II was established with the aim of ensuring that insurance companies meet their obligations to policy-holders (e.g., Eckert and Gatzert, 2018).[7] The idea behind this notion is that the company is required to meet its obligation payments with a probability of 99.5% over 12 months (Hari et al., 2008). The supervisory regulator uses a mortality shock based model which has been criticized, see for instance, Plat (2011), for its over-estimation of longevity trend risk. Moreover, they apply risk measures such as VaR and also longevity-trend stress test scenarios in order to evaluate the solvency capital requirement.[8] In the present paper, we develop a Bayesian estimation method for modelling death counts under a nonlinear state-space framework for the Lee and Carter (1992) (LC) and the Cairns et al. (2006) (CBD) model, which are two commonly used mortality models in the corresponding literature, by utilizing the EKF and MCMC techniques.

Secondly, we propose a new Bayesian framework for the VaR backtest. The Bayesian VaR backtest contains many advantages in the realm of testing, it allows a measure of evidence towards one hypothesis in comparison to another using direct inference, and there is no arbitrary type I error cut-off

[6] European Insurance and Occupational Pensions Authority (2014) Technical specification for the preparatory phase (part I). https://eiopa.europa.eu/Publications/. Accessed 23 November 2020.

[7] This regulation contains three pillars, and our focus in this paper is on pillar 2, where it contains the risk based solvency capital requirements.

[8] p. 122 of EIOPA (2014): Technical specification for the preparatory phase (part I): "*The SCR should correspond to the Value-at-Risk of the basic own funds of an insurance or reinsurance undertaking subject to a confidence level of 99.5% over a one-year period. The parameters and assumptions used for the calculation of the SCR reflect this calibration objective.*"

point. Most importantly, using Bayes rule, we can obtain the probability of a hypothesis being correct given the dataset. In this regard, we first state the assumptions used by the UC test, then develop the Bayesian decision theoretic framework surrounding these assumptions. Most importantly, our Bayesian VaR backtesting framework developed in this study is highly flexible and easy to implement due to the Bayesian conjugacy property which allows a closed-form expression of the posterior. In addition, in the case where Bayesian conjugacy does not exist, we employ recent econometric advancements in Bayesian estimation which also allows for numerical approximations such as MCMC methods. Furthermore, as a robustness test, varying non-informative prior distributions were used to ensure the decision is coherent. In addition we show how the Bayes Factor can be extended to the Independence and CC backtest.[9]

Finally, we develop the idea of backtesting in annuity liabilities which has two main advantages (Leung et al., 2018). Firstly, the backtesting framework allows us to measure the ability for mortality models to capture longevity risk associated with the annuity itself. Secondly, it allows for the ability to determine models most suitable for longevity risk applications.[10] Our focus is on the long term longevity trend risk, and in this case a longevity stress test scenarios would be most suitable. This mainly stems from the fact that longevity trend risk exacerbates over a longer period, and as such a one-year VaR would most likely be unsuitable. Thus, under a longevity stress trend scenario, it is crucial that a suitable backtesting method is developed to determine if the underlying longevity risks associated with a pricing instrument are actually captured by the mortality model used. Although our focus is on the annuity with contingent payments based on the policyholders lifespan, we should emphasise here that the backtesting framework we develop can also be extended to measure accuracy of any type of risk measures.

4. Main results

4.1. Bayesian mortality state-space models, estimation and forecasting

Government interventions such as the introduction of Solvency II regulation has required insurance companies to strictly manage their reserves to reduce the risk of insolvency. Thus it becomes a crucial aspect for insurance services companies to not over- or under-compensate the required reserves which

[9] While we do not emphasize on the application of both these tests, it is still crucial as it forms a stepping stone for combining Bayesian Decision Theory and VaR backtesting.

[10] More discussion about life annuities can be found in the Supplementary Information (SI) Section 1.

is contingent on the underlying mortality assumptions. This is particularly important to pension providers where management of pension payments are crucially dependent on many risk factors including longevity risk (e.g., Konicz and Mulvey, 2015). In this Section, we develop the Bayesian modelling estimation[11] and forecasting procedures for applying the new backtesting technique developed in the previous section for the annuity liability experience in the insurance industry.

One of the more prominent models in mortality modelling is the Lee and Carter (1992) model, which is commonly used as tool for mortality estimation and forecasting due to its simplistic model nature (Debon et al., 2008). Another widely accepted mortality model is the Cairns et al. (2009) model which offers accuracy of mortality rates for higher ages and non-age specific parameters. A possible downfall of the estimation procedure is that it requires a two step approach: firstly, a point estimation stage is produced, secondly a fitting stage is then conducted on the latent dynamics.

In this paper, we focus on a Bayesian estimation of both a linear variant of the LC and CBD models, see details in Appendix A, and a nonlinear variant based on Poisson and Binomial distributed death counts, respectively. A Kalman Filter alongside a Metropolis-Within-Gibbs sampler embedded in a MCMC algorithm will be used as the benefits are twofold. Firstly, the Kalman Filter is a one-step procedure and is able to retain the state dynamics without the need for an extra fitting procedure, and secondly, we are able to retain the MCMC draws for posterior inference and parameter risk analysis.

Let $\mu_x(t)$ denote the force of mortality for an individual aged x at time t. Under the piecewise constant force of mortality assumption we have:

$$\mu_{x+s}(t+s) = m_{x,t} \quad \text{for } 0 \leq s < 1 \text{ and } x \in \mathbb{N},$$

with $q_{x,t} = 1 - e^{-m_{x,t}}$, where $q_{x,t}$ represents the 1-year death probability for an individual aged x at time t. Let us denote the crude central death rate and crude death rate as

$$\tilde{m}_{x,t} = \frac{d_{x,t}}{E_{x,t}}, \quad \tilde{q}_{x,t} = 1 - e^{-\tilde{m}_{x,t}}, \tag{4.1}$$

where $d_{x,t}$ is the number of deaths recorded at age x during year t, and $E_{x,t}$ is the total population at age x during year t. The n-year survival rate of a person aged x at time t can be calculated as

$$S_{x,t}(n) = \prod_{i=1}^{n}(1 - q_{x+i,t+i}) = \exp\left(-\sum_{i=1}^{n} m_{x+i,t+i}\right). \tag{4.2}$$

[11] Arguments and some necessary details about the Bayesian state-space model estimation procedure can be found in SI Section 1.1.

Let us assume $x \in \{x_1, \ldots, x_n\}$ and $t \in \{t_1, \ldots, t_T\}$, where x_1 represents the initial age of the dataset, x_n represents the ultimate age of our dataset, t_1 represents the initial year, and t_T corresponds to the final year used, for simplicity we will represent $t_1 = 1, ..., t_T = T$, where T is the time horizon.

In the next section we will introduce the idea of MCMC and Bayesian model estimation, for more information regarding a general Bayesian modelling framework see SI Section 1.1.

4.1.1. The Lee-Carter model

Let $y_{x,t} := \ln(m_{x,t})$[12], then the LC model assumes that the central mortality rate is governed by the following process,

$$\boldsymbol{y_t} = \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\varepsilon_t}, \tag{4.3}$$

where $\boldsymbol{y_t} := \{y_{x,t} : x \in (x_1, ..., x_n)\}$; $\boldsymbol{\alpha'} := \{\alpha_x : x \in (x_1, ..., x_n)\}$ and $\boldsymbol{\beta'} := \{\beta_x : x \in (x_1, ..., x_n)\}$ are age dependent variables, κ_t captures the time dynamics of the population common through all ages, and $\boldsymbol{\varepsilon_t} \sim N(0, \mathbb{I}_n\sigma_\varepsilon^2)$. Here, $\mathbb{I}_n$ represents the $n \times n$ identity matrix and a random walk with drift process is used to model the latent state dynamics to facilitate the state-space formulation,

$$\kappa_t = \kappa_{t-1} + \delta + \omega_t, \tag{4.4}$$

where $\omega_t \sim N(0, \sigma_\omega^2)$ and δ represents the drift term of the process, furthermore ω_t and ε_t are assumed to be independent. It is shown in Lee and Carter (1992) that the parametrization in Eq. (4.3) is not unique, which means that for a particular likelihood maximization there is an indefinite number of solutions to the maximum likelihood estimate. To rectify this situation Lee and Carter (1992) imposed the constraints, $\sum_{x=x_1}^{x_n} \beta_x = 1$, and $\sum_{t=1}^{T} \kappa_t = 0$. In our case we also follow these constraints.[13]

For the Poisson model estimation under a nonlinear Bayesian state-space framework, we use a Gibbs sampler, the EKF and a MH, embedded in an MCMC algorithm. Let us first assume that the number of deaths $D_{x,t}$ follows a Poisson distribution with rate $E_{x,t}m_{x,t}$, where $\log(m_{x,t})$ is assumed to be the standard LC model. We have,

$$P(D_{x,t} = \mathrm{d}_{x,t}) = \frac{\exp^{-\mathrm{E}_{x,t}m_{x,t}}(\mathrm{E}_{x,t}\mathrm{m}_{x,t})^{\mathrm{d}_{x,t}}}{\mathrm{d}_{x,t}!},$$

[12] As the central mortality rate cannot be observed, we can instead model the crude central mortality rate given by Eq. (4.1).

[13] For more information regarding the LC model, we refer to SI Section 1.2.

$$\log(E_{x,t}m_{x,t}) = \begin{bmatrix} {}_{NL}\boldsymbol{\alpha} + \log(E_{x,t}) & {}_{NL}\boldsymbol{\beta} \end{bmatrix} \begin{bmatrix} 1 \\ {}_{NL}\kappa_t \end{bmatrix},$$

$${}_{NL}\kappa_t = {}_{NL}\kappa_{t-1} + {}_{NL}\,\delta + {}_{NL}\,\omega_t, \quad {}_{NL}\omega_t \sim N(0, {}_{NL}\,\sigma_\omega^2).$$

Let our static model parameter vector be defined as $\Theta_{\text{PLC}} = \{{}_{NL}\boldsymbol{\alpha},\ {}_{NL}\boldsymbol{\beta},\ {}_{NL}\delta,\ {}_{NL}\sigma_\omega^2,\ {}_{NL}\sigma_\beta^2\}$, then our MCMC algorithm is as follows:

1. Initialise with $\Theta_{\text{PLC}}^{(0)}$ and ${}_{NL}\boldsymbol{\kappa}^{(0)}$.
2. For $i = 1, \ldots, M$,
   (a) Apply the EKF using the function EKF-LC$_\kappa_t({}_{NL}\boldsymbol{\alpha}^{(i-1)},\ {}_{NL}\boldsymbol{\beta}^{(i-1)},\ {}_{NL}\delta^{(i-1)},\ {}_{NL}(\sigma_\omega^2)^{(i-1)})$ to obtain ${}_{NL}\boldsymbol{\kappa}^{*}$.
   (b) Using the function MH-LC$_\kappa_t({}_{NL}\boldsymbol{\kappa}^{*},\ {}_{NL}\boldsymbol{\kappa}^{(i-1)},\ {}_{NL}\boldsymbol{\alpha}^{(i-1)},\ {}_{NL}\boldsymbol{\beta}^{(i-1)},\ {}_{NL}\delta^{(i-1)},\ {}_{NL}(\sigma_\omega^2)^{(i-1)})$ sample ${}_{NL}\boldsymbol{\kappa}^{(i)}$ from $\pi({}_{NL}\boldsymbol{\kappa}|\Theta_{\text{PLC}}^{(i-1)}, \boldsymbol{y}_{1:T})$.
   (c) Using the function MH-LC$_\boldsymbol{\beta}({}_{NL}\boldsymbol{\kappa}^{(i)},\ {}_{NL}\boldsymbol{\alpha}^{(i-1)},\ {}_{NL}\boldsymbol{\beta}^{((i-1)},\ {}_{NL}\delta^{(i-1)},\ {}_{NL}(\sigma_\omega^2)^{(i-1)},\ ({}_{NL}\sigma_\beta^2)^{(i)})$ sample ${}_{NL}\boldsymbol{\beta}^{(i)}$ from $\pi(\boldsymbol{\beta}|\Theta_{\text{PLC}}^{(i-1)}, {}_{NL}\boldsymbol{\kappa}^{(i)}, \boldsymbol{y}_{1:T})$.
   (d) Sample $\Theta_{\text{PLC}}^{(i)}$ from $\pi(\Theta_{\text{PLC}}|_{NL}\boldsymbol{\beta}^{(i)}, {}_{NL}\boldsymbol{\kappa}^{(i)}, \boldsymbol{y}_{1:T})$.

A sample of the conditional distribution $\pi({}_{NL}\boldsymbol{\kappa}|\Theta_{\text{PLC}}, \boldsymbol{y}_{1:T})$ is obtained via a combination of the EKF and the MH algorithm. The full MCMC algorithm is shown in Appendix B. To draw samples from $\pi(\Theta_{\text{PLC}}|_{NL}\kappa_{1:t_T}^{(i)}, \boldsymbol{y}_{1:t_T})$, we assume the following conjugate prior distributions:

$$\pi({}_{NL}\delta) \sim N(\mu_\delta, \sigma_\delta^2),$$

$$\pi({}_{NL}\sigma_\omega^2) \sim I.G(a_\omega, b_\omega), \quad \pi({}_{NL}\sigma_\beta^2) \sim I.G(a_\beta, b_\beta)$$

$$\pi({}_{NL}\alpha_x) \sim \text{LogGamma}(a_\alpha, b_\alpha), \quad \pi({}_{NL}\beta_x) \sim N(\mu_\beta, \sigma_\beta^2) \text{ for } x \in \{(x_1, ..., x_n).$$

Non-informative priors were chosen to ensure the posterior distribution is mainly data driven. The conditional posterior distribution for Θ_{PLC} are as follows[14]:

$$\pi({}_{NL}\alpha_x|\boldsymbol{y},\ {}_{NL}\boldsymbol{\beta},\ {}_{NL}\boldsymbol{\kappa}) \sim \text{LogGamma}(a_\alpha + \textstyle\sum_{t=1}^{T} \mathrm{d}_{x,t}, b_\alpha + \sum_{t=1}^{T} \mathrm{E}_{x,t} \exp({}_{NL}\beta_x\ {}_{NL}\kappa_t)),$$

$$\pi({}_{NL}\sigma_\beta^2|\boldsymbol{y},\ {}_{NL}\boldsymbol{\beta}) \sim \text{I.G}\left(a_\beta + \tfrac{N}{2}, b_\beta + \tfrac{1}{2}\,{}_{NL}\boldsymbol{\beta}\,{}_{NL}\boldsymbol{\beta}'\right),$$

$$\pi({}_{NL}\delta|\boldsymbol{y},\ {}_{NL}\boldsymbol{\kappa},\ {}_{NL}\sigma_\omega^2) \sim N\left((\mu_\delta\ {}_{NL}\sigma_\omega^2 + \sigma_\delta^2 \sum_{t=1}^{T} ({}_{NL}\kappa_t - {}_{NL}\kappa_{t-1}))(\sigma_\delta^2\,{}_{NL}\sigma_\omega^2)^{-1},\ (\sigma_\delta^2\,{}_{NL}\sigma_\omega^2)(T\sigma_\delta^2 + {}_{NL}\sigma_\omega^2)^{-1}\right),$$

$$\pi({}_{NL}\sigma_\omega^2|\boldsymbol{y},\ {}_{NL}\boldsymbol{\kappa},\ {}_{NL}\delta) \sim I.G(a_\omega + \tfrac{T}{2},\ b_\omega + ({}_{NL}\kappa_t - ({}_{NL}\kappa_{t-1} + {}_{NL}\delta))^2),$$

[14]For a derivation of the posterior distributions see Lemmas 1.1-1.5 (with their proofs) in SI Section 1.3.

the sampling procedure for $\pi({}_{NL}\beta_x|\boldsymbol{y}, {}_{NL}\boldsymbol{\kappa}, \Theta_{\text{PLC}})$ was accomplished via a Random Walk MH algorithm (see SI Algorithm 3).

4.1.2. The Cairns-Blake-Dowd model

The CBD model has a wide variety of applications ranging from actuarial pricing, longevity derivative pricing and mortality predictions. Cairns et al. (2006) proposed to model the dynamics of the true 1-year death rates as follows

$$q_{x,t} = \frac{e^{\kappa_{1,t}+\kappa_{2,t}(x-\bar{x})}}{1+e^{\kappa_{1,t}+\kappa_{2,t}(x-\bar{x})}},$$

or equivalently

$$\ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_{1,t} + \kappa_{2,t}(x-\bar{x}), \tag{4.5}$$

where $\bar{x} = n^{-1}\sum_i x_i$ and the latent period factor $\boldsymbol{\kappa_t} := [\kappa_{1,t}, \kappa_{2,t}]'$ is a multivariate random walk with drift process with non-trivial variance-covariance structure:

$$\boldsymbol{\kappa_t} = \boldsymbol{\theta} + \boldsymbol{\kappa_{t-1}} + \boldsymbol{\omega_t}, \quad \boldsymbol{\omega_t} \sim N(\mathbf{0}, \Sigma), \tag{4.6}$$

where $\boldsymbol{\theta} := [\theta_1\, \theta_2]'$ is the drift vector, and Σ is a 2×2 covariance matrix.[15]

The Binomial model for the number of deaths is used for the CBD model due to its canonical link with the generalized dynamic linear model in mortality modelling. Instead of using the crude death rates, we use the observed number of deaths, and assume it follows a Binomial(n,p), with $n = \mathrm{E}_{x,t}$, $p = q_{x,t}$, and $\text{logit}(q_{x,t})$ as defined in Eq. (4.5). The nonlinear state-space framework is given as follows:

$$\begin{aligned}
P(\mathrm{D}_{x,t} = \mathrm{d}_{x,t}) &= \binom{\mathrm{E}_{x,t}}{\mathrm{d}_{x,t}} \mathrm{q}_{x,t}{}^{\mathrm{d}_{x,t}}(1-\mathrm{q}_{x,t})^{\mathrm{E}_{x,t}-\mathrm{d}_{x,t}},\\
\text{logit}(\mathrm{q}_{x,t}) &= \begin{bmatrix} 1 & (x-\bar{x}) \end{bmatrix} \begin{bmatrix} {}_{NL}\kappa_{1,t} \\ {}_{NL}\kappa_{2,t} \end{bmatrix},\\
\begin{bmatrix} {}_{NL}\kappa_{1,t} \\ {}_{NL}\kappa_{2,t} \end{bmatrix} &= \begin{bmatrix} {}_{NL}\theta_1 \\ {}_{NL}\theta_2 \end{bmatrix} + \begin{bmatrix} {}_{NL}\kappa_{1,t-1} \\ {}_{NL}\kappa_{2,t-1} \end{bmatrix} + \begin{bmatrix} \omega_{1,t} \\ \omega_{2,t} \end{bmatrix}, \quad \begin{bmatrix} \omega_{1,t} \\ \omega_{2,t} \end{bmatrix} \sim N(0, {}_{NL}\Sigma).
\end{aligned}$$

[15] For a more detailed analysis of the CBD model see SI Section 1.4.

With the static model parameter vector $\Theta_{\text{BCBD}} = \{{}_{NL}\theta_1, {}_{NL}\theta_2, {}_{NL}\Sigma\}$, the MCMC algorithm is as follows:

1. Initialise $\Theta^{(0)}_{\text{BCBD}}$ and $\boldsymbol{\kappa}^{(0)}_{1:T}$.
2. For $i = 1, \ldots, M$,
   (a) Apply the EKF using the function EKF_$\kappa_t({}_{\boldsymbol{NL}}\boldsymbol{\theta}^{(i-1)}, {}_{NL}\Sigma^{(i-1)})$ to obtain $\boldsymbol{\kappa}^*$.
   (b) Using the function MH_$\kappa_t(\boldsymbol{\kappa}^*,\ {}_{\boldsymbol{NL}}\boldsymbol{\kappa}^{(i-1)}_{1:T},\ {}_{\boldsymbol{NL}}\boldsymbol{\theta}^{(i-1)}, {}_{NL}\Sigma^{(i-1)})$ sample ${}_{\boldsymbol{NL}}\boldsymbol{\kappa}^{(i)}_{1:T} \sim \pi({}_{\boldsymbol{NL}}\boldsymbol{\kappa}_{1:T}|\Theta^{(i-1)}_{\text{BCBD}}, \boldsymbol{y}_{1:T})$.
   (c) Sample $\Theta^{(i)}_{\text{BCBD}}$ from $\pi(\Theta_{\text{BCBD}}|{}_{\boldsymbol{NL}}\boldsymbol{\kappa}^{(i)}_{1:T}, \boldsymbol{y}_{1:T})$.

A sample from $\pi({}_{NL}\boldsymbol{\kappa}_{1:T}|\Theta_{\text{CBD}}, \boldsymbol{y}_{1:T})$ can be obtained via an EKF with MH algorithm. To draw samples from the posterior distributions, $\pi(\Theta_{\text{BCBD}}|{}_{\boldsymbol{NL}}\boldsymbol{\kappa}^{(i)}_{1:T}, \boldsymbol{y}_{1:T})$, we assume the following priors for Θ_{BCBD}:

$$\pi({}_{NL}\theta_i) \sim N(\mu_{\theta_i}, \Sigma_{\theta_i}), \quad i = 1, 2,$$

$$\pi({}_{NL}\Sigma|(\sigma_1^2, \sigma_2^2)) \sim I.W\left((\nu+2)-1, 2\xi \text{diag}\left(\tfrac{1}{\sigma_1^2}, \tfrac{1}{\sigma_2^2}\right)\right),$$

$$\pi(\sigma_k^2) \overset{indep}{\sim} I.G\left(\tfrac{1}{2}, \tfrac{1}{A_k}\right) \quad k = 1, 2.$$

Using the prior distributions described above, the posterior distributions for the static parameters are given by:

$$\pi({}_{NL}\theta|\boldsymbol{y}, {}_{\boldsymbol{NL}}\boldsymbol{\kappa}, {}_{NL}\Sigma) \sim N\left((\Sigma_\theta^{-1} + n_{NL}\Sigma^{-1})^{-1}\left({\Sigma_\theta}^{-1}\mu_\theta + n_{NL}\Sigma^{-1}\textstyle\sum_{t=1}^{T}[{}_{NL}\kappa_t - {}_{NL}\,\kappa_{t-1}]\right), \left(\Sigma_\theta^{-1} + T\,{}_{NL}\Sigma^{-1}\right)^{-1}\right),$$

$$\pi(\sigma_k^2|{}_{NL}\Sigma) \overset{i.i.d}{\sim} I.G(\tfrac{\xi+T}{2}, \xi[{}_{NL}\Sigma^{-1}]_{kk} + \tfrac{1}{(A_k)^2}) \text{ for } k \in (1, 2),$$

$$\pi({}_{NL}\Sigma|\sigma_1^2, \sigma_2^2, \boldsymbol{y}, {}_{\boldsymbol{NL}}\boldsymbol{\kappa}, {}_{\boldsymbol{NL}}\boldsymbol{\theta}) \sim I.W(\xi+T+n-1, 2\xi diag(\tfrac{1}{\sigma_1^2}, \tfrac{1}{\sigma_2^2}) + \textstyle\sum_{t=1}^{T}\left[{}_{\boldsymbol{NL}}\boldsymbol{\kappa_t} - {}_{\boldsymbol{NL}}\boldsymbol{\theta}\right]\left[{}_{\boldsymbol{NL}}\boldsymbol{\kappa_t} - {}_{\boldsymbol{NL}}\boldsymbol{\theta}\right]'),$$

where $[{}_{NL}\Sigma^{-1}]_{kk}$ denotes the (k, k) element of $[{}_{NL}\Sigma^{-1}]$. Derivations of these posteriors are provided in SI Appendix 1.5. Once again, the choice of a hierarchical prior for Σ is to circumvent the issue of the inverse-wishart prior leading to a biased estimator for the correlation coefficient when the variances are small. Details for the MCMC algorithm including EKF are provided in Appendix C.

4.1.3. K-step ahead forecasting

Under the Bayesian method of forecasting, we utilize our posterior draws which retain information about our parameter uncertainty to produce our K-step ahead forecasts. The method to produce the forecasts for the LC and CBD model varies in the dimension of the variance-covariance matrix and drift term. Let us start by denoting k as the k^{th} step ahead forecast, this is consistent with the notion used in Algorithms 1 and 2. Furthermore, let m denote the m^{th} iteration from the MCMC, where M

is the number of kept iterations after the burn-in period.[16] For both Algorithms 1 and 2, the latent

Algorithm 1 Bayesian K step ahead forecasting for the LC model

for $k = 1, ..., K$ **do**
  for $i = 1, ...M$ **do**
  Linear
    $\kappa^i_{T+k} \sim N(\kappa^i_{T+(k-1)} + \delta^i, (\sigma^2_\omega)^i)$
    $\log(\tilde{\mathrm{m}}^i_{x,T+k}) \sim N(\alpha^i_x + \beta^i_x \kappa^i_{T+k}(\sigma^2_\varepsilon)^i)$
  Nonlinear
    ${}_{NL}\kappa^i_{T+k} \sim N({}_{NL}\kappa^i_{T+(k-1)} + {}_{NL}\delta^i, ({}_{NL}\sigma^2_\omega)^i)$
    $\mathrm{m}^i_{x,T+k} = \exp({}_{NL}\alpha^i_x + {}_{NL}\beta^i_x \ {}_{NL}\kappa^i_{T+i})$
  end for
end for

Algorithm 2 Bayesian K step ahead forecasting for the CBD model

for $k = 1, ..., K$ **do**
  for $i = 1, ...M$ **do**
  Linear
    $\boldsymbol{\kappa^i_{T+k}} \sim N(\boldsymbol{\kappa^i_{T+(k-1)}} + \boldsymbol{\theta}^i, (\sigma^2_\nu)^i)$
    $\mathrm{logit}(\tilde{\mathrm{q}}^i_{T+k}) \sim N((x - \bar{x})\boldsymbol{\kappa^i_{T+(k-1)}}, (\sigma^2_\varepsilon)^i)$
  Nonlinear
    ${}_{NL}\boldsymbol{\kappa^i_{T+k}} \sim N({}_{NL}\boldsymbol{\kappa^i_{T+(k-1)}} + {}_{NL}\boldsymbol{\theta}^i, ({}_{NL}\sigma^2_\nu)^i)$
    $\mathrm{q}^i_{T+k} = \frac{\exp((x-\bar{x})_{NL}\kappa^i_{T+k})}{1+\exp((x-\bar{x})_{NL}\kappa^i_{T+k})}$
  end for
end for

states are taken from the *Forward Filtering Backward Sampling* (FFBS) algorithm, where the model static parameters are taken from Gibbs sampling at the m^{th} iteration.[17] The 10-year ahead mortality forecasts for ages 50-90 in increments of 10 years over the years 2001 till 2010 was also produced. For example, $m(90, t)$ will correspond to the realised and forecasted mortality rate at age 90 for a specified country between 1950 and 2010.

4.2. The new Bayesian Backtesting Framework

Financial risk model evaluation plays a major part in risk management and typically this evaluation process is called a *backtesting procedure*, which aims to measure the accuracy of the risk models promised coverage. For instance, a VaR model tries to define a conditional quantile (or coverage) of the return distribution. To evaluate the effectiveness of the VaR model, we can backtest it and determine whether the required coverage rate is met. This is usually accomplished by using ex-post returns on ex-ante VaR forecasts. In this Section, we propose a new UC backtest for VaR-forecasts

[16] For the parameter estimation results and convergence statistics see SI Section 1.7.
[17] SI Section 1.7 shows a 10-step ahead forecast of κ_t from the LC model and $\boldsymbol{\kappa_t}$ from the CBD model.

under a Bayesian framework which is a cornerstone of this paper. In addition, we also include the formulation of the *Independence* and *CC* backtest under this same framework. However, our application will be focussed on the UC backtest and further developments and application of the two other backtests will be available for future research.

Before we proceed with the new UC backtest, let us consider two hypotheses, H_0 and H_1, that we wish to test. Under the standard NHST framework, inference is normally conducted on $P(\boldsymbol{y}|H_0)$, however applying Bayes rule, we obtain the following relation:

$$\pi(H_i|\boldsymbol{y}) = \frac{P(\boldsymbol{y}|H_i)\pi(H_i)}{\pi(\boldsymbol{y})}, \ i = 0, 1,$$

where $\pi(\boldsymbol{y})$ is the marginalizing constant to ensure $\pi(H_i|\boldsymbol{y})$ is a proper probability distribution, and finally the point of interest, the posterior odds ratio, is given by:

$$\frac{\pi(H_0|\boldsymbol{y})}{\pi(H_1|\boldsymbol{y})} = \frac{\pi(\boldsymbol{y}|H_0)\pi(H_0)}{\pi(\boldsymbol{y}|H_1)\pi(H_1)} = \text{BF}_{01}\frac{\pi(H_0)}{\pi(H_1)}, \tag{4.7}$$

where $\text{BF}_{01} = \frac{\pi(\boldsymbol{y}|H_0)}{\pi(\boldsymbol{y}|H_1)}$ is commonly referred to as the BF, and $\frac{\pi(H_0)}{\pi(H_1)}$ is known as the *prior odds*. BF_{01} measures the change in evidence when going from prior to posterior odds. In the decision making process where both hypotheses are given equal weighting, the testing framework focuses solely on BF_{01}, thus a higher positive value for BF_{01} implies $\pi(H_0|\boldsymbol{y}) > \pi(H_1|\boldsymbol{y})$, and concludes an increased support for H_0.

Consider now a point null hypothesis $\{H_0 = \theta_0\}$ and a composite alternative hypothesis $\{H_1 \neq \theta_0\}$. The Bayesian framework then assigns a prior distribution over both hypotheses. Let $\boldsymbol{y} := \{y_1, ..., y_n\}$ be a vector of n observations, the likelihood function of the observed data is given by $l(\boldsymbol{y}|\theta)$, where θ is the parameter of interest. Then, for a given prior, $\pi(\theta)$, our posterior distribution is given by

$$\pi(\theta|\boldsymbol{y}) = \frac{l(\boldsymbol{y}|\theta)\pi(\theta)}{\pi(\boldsymbol{y})}. \tag{4.8}$$

The prior specifications will be as follows: "*under H_0, we assign the (point mass) prior $\pi(\theta) = \theta_0$, whereas for H_1, we assign a prior distribution over the parameter space required*". The decision to accept H_0 is denoted by "a_0" and the decision against H_0 is denoted by "a_1". Overall, for a given loss function, $\mathcal{L}[a_i; \theta], i = 0, 1$, H_0 is rejected when the expected posterior loss for H_0 is sufficiently larger than the expected posterior loss under H_1. The expected posterior loss for the i^{th} decision is

given by

$$E_{\pi(\theta|\boldsymbol{y})}\left[\mathcal{L}(\theta, a_i)\right] = \int_{\theta} \mathcal{L}[a_i;\ \theta]\pi(\theta|\boldsymbol{y})d\theta,\ \text{for } i = 1, 2 \tag{4.9}$$

and we will reject H_0 when

$$\int_{\theta} \left(\mathcal{L}[a_0;\ \theta] - \mathcal{L}[a_1;\ \theta]\right)\pi(\theta|\boldsymbol{y})d\theta > 0. \tag{4.10}$$

If we choose to employ a zero-one loss function[18]

$$\mathcal{L}[a_0;\ \theta] = \begin{cases} 0 & \text{if } \theta = \theta_0, \\ 1 & \text{if } \theta \neq \theta_0, \end{cases} \tag{4.11}$$

with $\mathcal{L}[a_1;\ \theta] = 1 - \mathcal{L}[a_0;\ \theta]$. Given equal probability of H_0 and H_1 occurring, that is $\pi(H_0) = \pi(H_1) = 0.5$, we will have the following decisions to make. For the first decision, when $\theta = \theta_0$, we will accept H_0 with decision a_0, and the second decision, a_1, occurs when $\theta \neq \theta_0$. To tabulate the decision outcome more formally:

$$\text{Choose: } \begin{cases} a_0, & \text{if } \ \theta = \theta_0, \\ a_1, & \text{if } \theta \neq \theta_0. \end{cases}$$

Combining Eqs. (4.9), (4.10) and (4.11), rejection of the H_0 will occur when:

$$\frac{\int_{\theta} \mathcal{L}[a_0;\ \theta]\pi(\theta|\boldsymbol{y})d\theta}{\int_{\theta} \mathcal{L}[a_1;\ \theta]\pi(\theta|\boldsymbol{y})d\theta} = \frac{l(\boldsymbol{y}|\theta = \theta_0)}{\int_{\theta} l(\boldsymbol{y}|\theta)\pi(\theta)d\theta} < 1, \tag{4.12}$$

where the quantity $\frac{l(\boldsymbol{y}|\theta=\theta_0)}{\int_{\theta} l(\boldsymbol{y}|\theta)\pi(\theta)d\theta}$ is the BF_{01}.[19] Further, note that the marginalizing constant $\pi(\boldsymbol{y})$ from Eq. (4.8) disappears in Eq. (4.12), since it appears in both the numerator and the denominator. The *Bayesian* version of the *Likelihood Ratio Test* (BLRT) was pioneered by Li et al. (2014), where instead of having a $0-1$ loss function which corresponds to BF_{01}, they used a continuous loss difference function, defined by

$$\Delta\mathcal{L}[H_0; \theta] = -2\left[\log(\pi(\boldsymbol{y}|\theta_0)) - \log(\pi(\boldsymbol{y}|\theta))\right],$$

[18] A zero-one loss function is a commonly chosen loss function used in Bayesian hypothesis testing, this is simply due to its binary outcome, and is equivalent to either rejecting (a_1) or accepting (a_0), the null hypothesis.

[19] The range of values that BF_{01} can take, represents the different levels of evidence in support of the null or the alternative hypothesis, in Table 1 of Goodman (2001) the strength of evidence against the null hypothesis for a given BF is shown.

under a continuous loss difference function, rejection of H_0 occurs when

$$\int_{\theta} \Delta\mathcal{L}[H_0;\theta]\pi(\theta|\boldsymbol{y})d\theta > 0,$$

and the Bayesian test statistic is given by

$$T_{\text{BLRT}}(\boldsymbol{y},\theta) = \left[-2\int_{\theta}\left[\log(\pi(\boldsymbol{y}|\theta_0)) - \log(\pi(\boldsymbol{y}|\theta))\right]\pi(\theta|\boldsymbol{y})d\theta\right] + 1. \tag{4.13}$$

The main difference between the BLRT and BF_{01} is that the BLRT focuses on averaging over the (log)posterior distribution, whereas BF_{01} averages over the prior distribution. Li et al. (2014) also found that $T_{\text{BLRT}}(\boldsymbol{y},\theta)$ has an asymptotic $\chi^2(1)$-distribution, and a convenient property of the BLRT statistic is that if the integral in Eq. (4.13) has no analytical form, it can be approximated via an MCMC,

$$T_{\text{BLRT}}(\boldsymbol{y},\theta) = -2\sum_{i=1}^{M}\left[\log(\pi(\boldsymbol{y}|\theta_0)) - \log(\pi(\boldsymbol{y}|\theta^i))\right]/M + 1, \tag{4.14}$$

where i represents the i^{th} MCMC draw, and M corresponds to the number of MCMC iterations. In either case, we can produce Bayesianised p-values via

$$p = P(\chi^2(1) \leq T_{\text{BLRT}}).$$

4.2.1. The Unconditional Coverage backtest

The statistical backtest for VaR developed by Kupiec (1995) tests whether a risk model truly generated the correct coverage using the LRT. In this section, we formulate a novel approach of the UC backtest using the Bayesian decision theoretic framework developed in the previous section. Let $\boldsymbol{y}$ denote the daily observed asset or portfolio losses, y_t, for $t \in (1,\ldots,T)$, and the VaR as $P(y_t \geq VaR_{t|\mathcal{F}_{t-1}}(p)) = p$. To produce interval forecasts for each observation, we let $U_{t|\mathcal{F}_{t-1}}(p)$ denote the upper forecast interval produced for time t using information up until $t-1$ with coverage p. Let us define an indicator variable where,

$$\mathcal{I}_p(t) = \begin{cases} 1, & \text{if } y_t \in (-\infty,\ U_{t|\mathcal{F}_{t-1}}(p)) \\ 0, & \text{otherwise,} \end{cases} \tag{4.15}$$

for which $\mathcal{F}_{t-1}$ corresponds to the information set $\mathcal{F}_{t-1} := \{\mathcal{I}_p(1),\ldots,\mathcal{I}_p(t-1)\}$.

In laymen terms, if the observed daily losses, y_t, is greater than the expected upper bound, $y_t > U_{t|\mathcal{F}_{t-1}}(p)$, then we conclude that the VaR forecasts are violated at time t and we assign a value of 1. Kupiec (1995) examines whether the average non-violations shown in Eq. (4.15) occurs at the required coverage p, mathematically,

$$E\left[(1/T)\sum_{t=1}^{T}\mathcal{I}_p(t)\right] = P(\mathcal{I}_p(t)=1) = p,\ \forall t. \tag{4.16}$$

This also implies that each $\mathcal{I}_p(t) \sim Be(p)$, where $Be(p)$ represents the Bernoulli distribution with probability p of success. Let $\boldsymbol{\mathcal{I}_p} := \{\mathcal{I}_p(t) : t \in (1,...,T)\}$, and let m_1 and m_0 denote the number of one and zero occurrences in $\boldsymbol{\mathcal{I}_p}$, respectively. Then, $\boldsymbol{\mathcal{I}_p}$ will be a vector of size $T = m_1 + m_0$. Our aim is to determine whether or not $E[\mathcal{I}_p(t)] = p^*$, for some predetermined probability p^* and since, $\mathcal{I}_p(t) \sim Be(p)\ \ \forall t$, the joint likelihood function will be given by,

$$l(\mathcal{I}|p) = p^{m_1}(1-p)^{m_0}. \tag{4.17}$$

The Bayesian framework starts by assigning priors on p. Under the $H_0 : p = p^*$ with an assigned point mass prior. Under the alternative $H_1 : p \neq p^*$, since p has a support between (0,1), we use a Beta prior distribution which mimics the support between 0 and 1. Formally, let

$$\pi(p) = \begin{cases} 1 & \text{if } p = p^*, \\ \text{Beta}(a,b), & \text{if } p \neq p^*. \end{cases}$$

The priors chosen here are non-informative and conjugate to the posterior, hence the posterior loss distribution will be mainly data driven and have a closed form expression.[20]

Lemma 4.1. *The BF for the UC backtest is given by,*

$$BF_{01} = \frac{(p^*)^{m_1}(1-p^*)^{m_0}}{\beta(a+m_1, b+m_0)},$$

where β corresponds to the β-function. For a proof of Lemma 4.1 see Appendix D.1. Then using the derived BF_{01} from Eq. (4.1), the decision to reject the H_0 will occur when,

$$\text{BF}_{01} = \frac{(p^*)^{m_1}(1-p^*)^{m_0}}{\beta(a+m_1, b+m_0)} < 1. \tag{4.18}$$

[20]In this paper we will focus on the non-informative prior, however, subjective priors based on past history can also be used to calibrate the hyper-parameters on the Beta(a,b) distribution.

For the BLRT statistic, we can use Eq. (4.13) instead of the simulation method presented in Eq. (4.14). The following Theorem provides an analytical form for the BLRT statistic for the UC backtest, which is extremely useful in what follows.

Theorem 4.1. *The analytical form for the BLRT statistic for the UC backtest is given by,*

$$T_{BLRT}(\boldsymbol{y}, p) = -2[A_{\pi_0} - B_{\pi_1}] + 1, \tag{4.19}$$

where

$$\begin{aligned} A_{\pi_0} &= m_1 \log(p^*) + m_0 \log(1 - p^*), \\ B_{\pi_1} &= m_1(\psi(a + m_1) - \psi(a + m_1 + b + m_0)) - m_0(\psi(b + m_0) - \psi(a + m_1 + b + m_0)), \end{aligned}$$

and ψ corresponds to the digamma function.

For a proof of Theorem 4.1 see Appendix D.2. Let C_{BLRT} be determined using a required tail significance from a $\chi^2(1)$-distribution, then the decision to reject H_0 will occur when $T_{\text{BLRT}}(\boldsymbol{y}, p) > C_{\text{BLRT}}$. A more formal representation for the UC test outcomes is shown in Table 1.

Table 1: Criteria for the rejection or acceptance of the H_0

	Reject H_0	Reject H_1
BF_{01}	$\frac{(p^*)^{m_1}(1-p^*)^{m_0}}{\beta(a+m_1, b+m_0)} \leq 1$	$\frac{(p^*)^{m_1}(1-p^*)^{m_0}}{\beta(a+m_1, b+m_0)} > 1$
T_{BLRT}	$T_{\text{BLRT}} \geq C_{BLRT}$	$T_{\text{BLRT}} < C_{BLRT}$

Finally, a power analysis is included in Table 5 of Appendix E to demonstrate the consistency of the Bayes Factor, and thus to support the adoption of the proposed test. We further provide a short description and comments of the simulation outputs, see Appendix E.

4.2.2. The Independence Test

Christoffersen (1998) developed the independence test for VaR-violations in $\boldsymbol{\mathcal{I}_p}$, which is tested against a first order Markov process. In this situation the chosen VaR coverage p^* will be irrelevant since the main focus is on the clustering of VaR violations and not their required coverage p^*. Let us first consider the transition probability matrix representing a first order Markov process,

$$\Pi_1 = \begin{bmatrix} 1 - p_{01} & p_{01} \\ \\ 1 - p_{11} & p_{11} \end{bmatrix}, \tag{4.20}$$

here, $p_{ij} = P(\mathcal{I}_p(t) = i|\mathcal{I}_p(t-1) = j)$. In this case, VaR violations do indeed contain a Markovian structure and hence we conclude that the violations occur dependently to each other. On the contrary, an independent model would have a transition probability matrix governed by Eq (4.21). In essence, the probability does not change if entering a "zero" state from either a "one" state or "zero" state.

$$\Pi_2 = \begin{bmatrix} 1-p_2 & p_2 \\ \\ 1-p_2 & p_2 \end{bmatrix}. \tag{4.21}$$

A common choice in the Bayesian literature is to use the natural Dirichlet prior since it is often used as a means of modelling transition probabilities. Let us define two different priors under each hypotheses. First, denote $\boldsymbol{\alpha_0}$ and $\boldsymbol{\alpha} := \{\alpha_0, \alpha_1\}$, both be the concentration parameter used in the Dirichlet prior. Let

$$\pi(\boldsymbol{p}) = \begin{cases} \text{Dir}(\boldsymbol{\alpha_0}) & \text{if } \{\boldsymbol{p_0} \in \Theta_0 : \boldsymbol{p_0} = \boldsymbol{p_1}\}, \\ \text{Dir}(\boldsymbol{\alpha}), & \text{if } \{\boldsymbol{p_0} \in \Theta_1 : \boldsymbol{p_0} \neq \boldsymbol{p_1}\}. \end{cases}$$

Lemma 4.2. *The BF for the Independence backtest is given by*

$$BF_{01}^{Ind} = \frac{\beta(\boldsymbol{\alpha_0} + n_{00} + n_{10},\ \boldsymbol{\alpha_0} + n_{01} + n_{11})}{\beta(\alpha_0 + n_{00},\ \alpha_1 + n_{01})\beta(\alpha_0 + n_{10},\ \alpha_1 + n_{11})}.$$

See Appendix D.2 for the proof of Lemma. 4.2.

4.2.3. The Conditional Coverage Test

The CC test is composed of both the UC test and the test for Independence. In this context, our objective is to determine if violations occur independently and are identically distributed with the specified level of probability of p^*. The test statistic for the CC test combines the null model from the UC test and the alternative model from the test for Independence.

$$\pi(\boldsymbol{p_i}) = \begin{cases} 1 & \text{if } \{p \in \Theta_0 : \boldsymbol{p_0} = \boldsymbol{p_1}, \quad p_{11} = p_{01} = p^*\}, \\ \text{Dir}(\boldsymbol{\alpha}), & \text{if } \{\boldsymbol{p_0} \in \Theta_1 : \boldsymbol{p_0} \neq \boldsymbol{p_1}\}. \end{cases}$$

Combining both BF_{01} and $\text{BF}_{01}^{\text{Ind}}$ we have,

$$\frac{\pi(\boldsymbol{y}|p = p^*)}{\int_{p \neq p^*} l(\boldsymbol{y}|p)\pi(p)d\theta} = \frac{(p^*)^{m_1}(1-p^*)^{m_0}}{\int_{\boldsymbol{p_1}} \int_{\boldsymbol{p_0}} \prod_{i=0}^{1} \prod_{j=0}^{1} p_{ij}^{n_{ij}} \text{Dir}\left(\boldsymbol{\alpha}\right) dp_0 dp_1}$$

$$\mathrm{BF}_{01}^{\mathrm{CC}} = \frac{(p^*)^{m_1}(1-p^*)^{m_0}}{\beta(\alpha_0+n_{00},\alpha_1+n_{01})\beta(\alpha_0+n_{10},\alpha_1+n_{11})}. \tag{4.22}$$

A summary table of the BF for the UC, Independence, and CC test is shown in Table 2.

Table 2: Bayes Factor for the UC, Independence, and CC Backtesting

	BF
Unconditional Coverage Test	$\frac{(p^*)^{m_1}(1-p^*)^{m_0}}{\beta(m_1+a,m_0+b)}$
Independence Test	$\frac{\beta(\boldsymbol{\alpha_0}+n_{00}+n_{10},\boldsymbol{\alpha_0}+n_{01}+n_{11})}{\beta(\alpha_0+n_{00},\alpha_1+n_{01})\beta(\alpha_0+n_{10},\alpha_1+n_{11})}$
Conditional Coverage Test	$\frac{(p^*)^{m_1}(1-p^*)^{m_0}}{\beta(\alpha_0+n_{00},\alpha_1+n_{01})\beta(\alpha_0+n_{10},\alpha_1+n_{11})}$

4.3. Annuity liabilities backtesting

To backest annuity liabilities, we apply a stress test procedure on the longevity trend with the aim of capturing longevity risk. Our first step is to obtain an upper quantile liability estimate. This is achieved by stressing the mortality forecasts at the lower quantile, which thus signifies improvements in life expectancy (liability). Assume now we have a \$1 continuously paid temporary annuity to a person currently aged x for the next N years. Let the price of a zero coupon bond which matures in n years be denoted as $B(0,n)$, we then have the liability for a \$1 annuity paid to a person aged x at time t for the next N years to be,[21]

$$L_x(N) = \sum_{n=1}^{N} B(0,n) S_{x,t}(n). \tag{4.23}$$

We intentionally choose to not use market based annuity rates since besides longevity risk the premium will include company dependent factors such as profits and expenses. Eq. (4.23) will only be affected by longevity improvements over time and as such allows us to focus on longevity trend risk. Let $N = \omega - x$, where ω is our limiting age. Using a set of mortality forecast intervals, a mean and upper bound on $L_x(N)$ can be obtained. Denote the mean of $L_x(N)$ as $L_x^{\mathrm{mean}}(N) = E[L_x(N)]$ and the upper bound as $L_x^{\mathrm{upper}}(N)$, where $L_x^{\mathrm{upper}}(N)$ is the liability calculated at the required VaR quantile p^*.[22] Furthermore, the capital requirement is a ratio which determines the extra capital amount

[21] Here, $B(0,n) := (\frac{1}{1+i})^n$

[22] It is calculated using a $(1-p^*)$ quantile of the mortality forecasts since a lower quantile estimate of mortality represents an increase of the annuity liability.

needed to be held at time t for someone aged x. It is determined using,

$$\text{CapR} = \left(\frac{L_x^{\text{upper}}(N)}{L_x^{\text{mean}}(N)} - 1 \right) \times 100\%. \tag{4.24}$$

The capital requirement ratio allows us to determine the percentage of extra capital needed to achieve the p^* quantile of annuity liabilities compared with its mean.

The idea of backtesting in the context of annuity pricing is to determine whether the stressed longevity scenarios were enough to capture the experienced liability over a required forecast time horizon. Our objective is to then test whether the "hits" (when $\mathcal{I} = 1$) follows a Be(p^*) distribution, where p^* represents the quantile of interest. A backtesting procedure using BF and BLRT test statistic shown in Eqs. (4.18) and (4.19) can be used to determine the strength of evidence for the null and alternative models. This procedure statistically examines whether the frequency of exceptions for N-year annuity liabilities is in line with the regulations of Solvency II. That is, whether companies are able to hold reserves capable of sustaining the liabilities in the long term. In this section we define "long term" to be capped at ω years, however this assumption can be relaxed. Let j represent either varying countries analysed or a separate male/female population. Then, denote ${}_jL_x^*(N)$ for $x \in (x_1, ..., x_n)$ as the sample path of realised liabilities. Let us create an indicator variable, $\mathcal{I}_x^j$ where for a given interval forecast $(-\infty,\ {}_jL_x^{\text{upper}}(N))$, we have

$$\mathcal{I}_p^j(x) = \begin{cases} 1, & \text{if } {}_jL_x^*(N) \in (-\infty,\ \ {}_jL_x^{\text{upper}}(N)), \\ 0, & \text{if } {}_jL_x^*(N) \notin (-\infty,\ \ {}_jL_x^{\text{upper}}(N)). \end{cases} \tag{4.25}$$

We wish to test whether $\mathcal{I}_p(x) \sim Be(p^*)$. Let $\boldsymbol{\mathcal{I}_{p^*}} := \{\mathcal{I}_p^j(t) : j \in (1, ..., J),\ x \in (x_1, ..., x_n)\}$, and let m_1 and m_0 denote the number of one and zero occurrences in $\boldsymbol{\mathcal{I}_{p^*}}$ respectively. Then $\boldsymbol{\mathcal{I}_{p^*}}$ will be a vector of size $T = m_1 + m_0$, and the joint likelihood function will be given by,

$$l(\boldsymbol{\mathcal{I}_p}|p) = p^{m_1}(1-p)^{m_0}. \tag{4.26}$$

We assign the point mass prior under $H_0 : p = p^*$ and under the alternative $H_1 : p \neq p^*$ we use a Beta(a,a) uninformative prior distribution. Let

$$\pi(p) = \begin{cases} 1 & \text{if } p = p^*, \\ \text{Beta}(a, a), & \text{if } p \neq p^*. \end{cases}$$

with,

$$\mathrm{BF}_{01} = \frac{(p^*)^{m_1}(1-p^*)^{m_0}}{\beta(a+m_1, a+m_0)}.$$

For the BLRT statistic we find $T_{\mathrm{BLRT}}(\boldsymbol{y}, \theta) = -2[A_{\pi_0} - B_{\pi_1}] + 1$ as defined in Eq. (4.19). Furthermore, we can also back-solve BF_{01} to obtain an implicit $\hat{p}$, which represents the true frequency rate implied by the model. $\hat{p}$ can be found by:

$$\hat{p} = \min_{p \in (0,1)} \frac{(p)^{m_1}(1-p)^{m_0}}{\beta(a+m_1, b+m_0)}. \tag{4.27}$$

5. Empirical results

In this section, we compare the results obtained from estimating the LC and the CBD models under the linear and nonlinear variants. We used the Human Mortality Database for the total population aged between 50 and 95 of the following list of countries: Australia, United Kingdom, Italy, France, Spain, New Zealand, Sweden, Germany, and Russia. For countries where the mortality data does not date back to 1950, the earliest year was used instead. A total of $20,000$ MCMC iterations are conducted, and the first $5,000$ was used as the burn-in period. Hyper-parameters were chosen to be non-informative, such that our posterior distribution will be mainly data driven. The hyper-parameter specifications are shown in SI Table 1, and they were identical for all countries. The Geweke statistic is a tool used in Bayesian statistics to determine whether the last iterations of the MCMC draws from the full conditional posteriors are different from the first half of the iterations, if there is no statistical evidence of a difference we say that the chain has reached a stationary state. The Geweke Statistic shown in SI Tables 2 to 19, indicate that most parameters reached a stationary state with a 95% confidence. Furthermore, the trace-plots shown in SI Figures 10 and 11 indicate no apparent signs of serial correlation, once again confirming our hypothesis that the chain has reached convergence.[23]

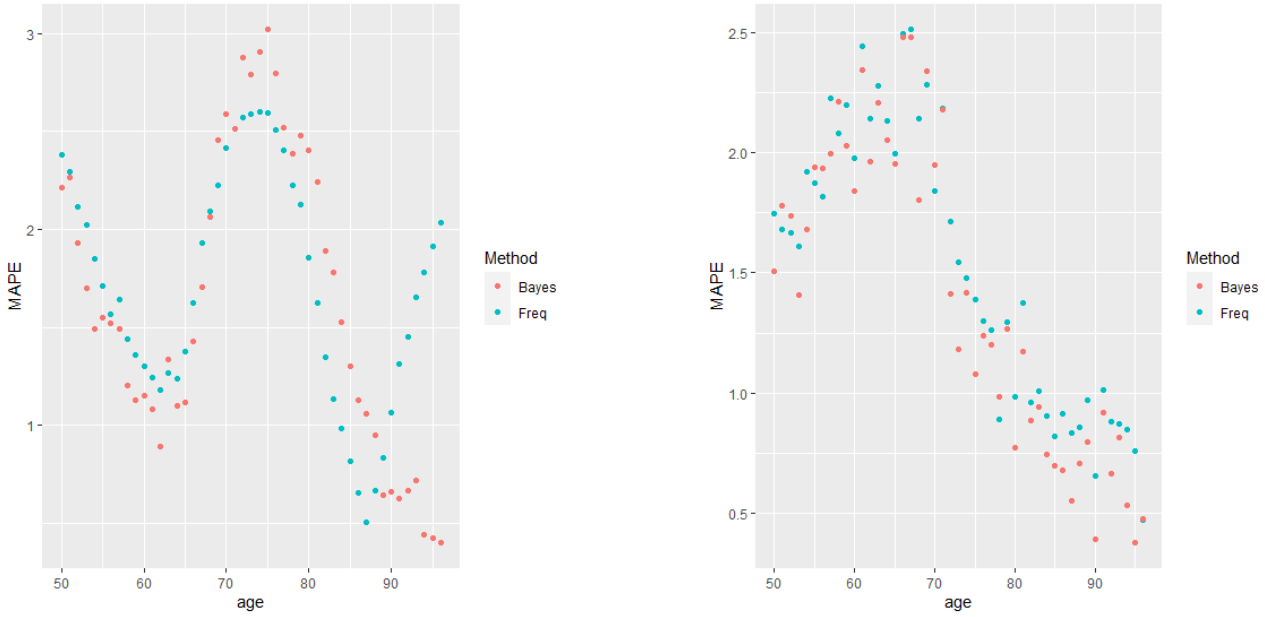
In order to generate our out-of-sample forecasts, we use ages $x_1 = 50$ to $x_n = \omega - 1 = 95$, where the period of interest spans from 2001 to 2013. The forecasts are obtained using the methods described in section 4.1.3 applied to $J = 9$ different countries. Using these forecast intervals we obtain the following set: $\{({}_jL_x^{\mathrm{mean}}(N), {}_jL_x^{\mathrm{upper}}(N)) : j \in (1, ..., J),\ x \in (x_1, ..., x_n)\}$, where ${}_jL_x^{\mathrm{upper}}(N)$ is calculated using the $p^* = 0.5\%$ quantile of the mortality forecasts[24] applied to Eq. (4.2), and thus would represent the liability estimated at the upper 99.5% quantile.

[23] For more details on the LC and CBD parameter implications see SI Section 1.6.1.

[24] A lower quantile estimate of mortality forecast represents an increase of the annuity liability.

Table 3: Comparison of the MAPE forecast measure under the Bayesian nonlinear model against the frequentist 2-step estimation averaged across all countries used.

Mean Absolute Percentage Error under the CBD model									
Age Group	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-95
Bayesian	1.95	1.48	1.03	1.36	2.38	2.54	2.18	1.25	0.297
Frequentist	2.13	1.54	1.25	1.85	2.54	2.37	1.39	0.694	1.45
Mean Absolute Percentage Error under the LC model									
Age Group	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-95
Bayesian	1.68	1.79	2.08	2.24	1.94	1.22	1.02	0.72	0.69
Frequentist	1.73	2.04	2.20	2.29	1.75	1.23	1.05	0.88	0.85



(a) CBD model

(b) LC model

Figure 1: MAPE measured over a 13 year forecast horizon under the CBD and LC model. The horizontal axis corresponds to the MAPE across ages and the vertical axis corresponds to the MAPE averaged across all countries. The red dotted line corresponds to the nonlinear Bayesian estimation and the blue dotted line represents the two-step frequentist approach.

To determine whether our new Bayesian nonlinear estimation methods produce plausible results, we compare them against the commonly used two-step frequentist estimation method. We use the "R" package StMoMo (Millossovich et al., 2017) to obtain frequentist forecasts, and in order to retain the non linearity estimation we used the Binomial and Poisson link functions for the CBD and LC model respectively. To test whether our proposed Bayesian nonlinear estimation benefits from the classical approach, we consider the Mean Average Percentage Error (MAPE) forecast measure. While

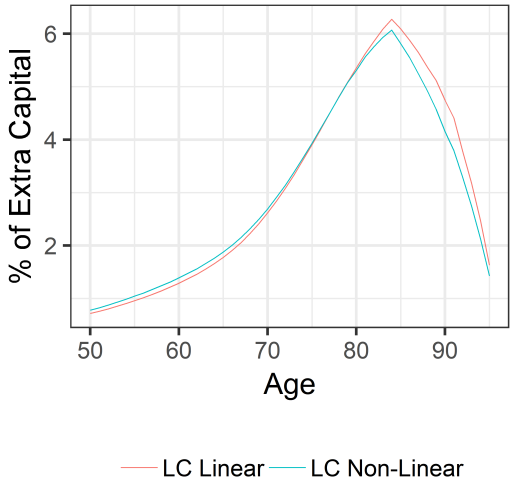
many forecast measures exist, see for instance (Dowd et al., 2010), the MAPE measure is among the most popular within the actuarial literature and in particular, Leung et al. (2018) uses it in the Bayesian linear context. The MAPE measure is given by:

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{A_t - F_t}{A_t}\right|, \tag{5.1}$$

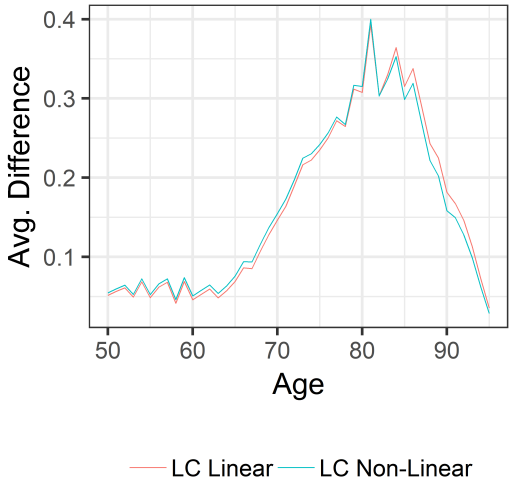
where A_t and F_t correspond to the realised and forecasted mortality rates at time t. Table 3 shows the Average of MAPE across all countries and we observe that under the CBD model our proposed method performs significantly better at the earlier and higher age groups, whilst the frequentist method outperforms in the age brackets 75-79, 80-84 and 85-89. Furthermore, under the LC model specification, our method performs better across all age groups barring 70-74. Figure 1 compares the MAPE across all ages, and we observe that the red dots, corresponding to our proposed method, are mostly *lower* than the blue dots representing the frequentist estimation. This further emphasizes that our proposed methods seems to perform better at earlier and later age groups.

Figures 2 and 3 represent the percentage of extra capital required for a 99.5% longevity stressed scenario. It is interesting to note that the linear and nonlinear variant of the LC model produced similar capital requirements and average difference across all countries for the realised annuity and 99.5% upper bound. Whereas on the contrary, the nonlinear CBD model produced largely varying results. The graph for the extra capital amount shows that the linear CBD model requires a larger amount for all ages when compared with its nonlinear counterpart. Furthermore, we see that the average difference across all ages shows that it peaks for the higher age groups, indicating that the linear CBD model overestimates the upper 99.5% annuity price. The findings can be summarized as follows. Firstly, the linear and nonlinear LC models show similar structures in the extra capital required and average difference of the upper annuity liability compared with the realised one, thus not much difference can be seen between the two models. The linear CBD model seems to over estimate the annuity liabilities and hence has the highest peak for the average difference curve, this is also reflected in the larger extra capital required to ensure that a 99.5% of the annuity obligation can be met.

To backtest the liabilities, we assign the point mass prior $H_0 : p = p^* = 0.5\%$ and the Beta(0.5,0.5) uninformative prior distribution for $H_1 : p \neq p^* \neq 0.5\%$. For the BLRT statistic we find $T_{\text{BLRT}}(\boldsymbol{y}, \theta) = -2[A_{\pi_0} - B_{\pi_1}] + 1$ with $p^* = 0.5\%$, $a = 0.5$, and $b = 0.5$. Table 6 in the Appendix shows the violations across age groups and countries, and also the extra capital required and differences from realised

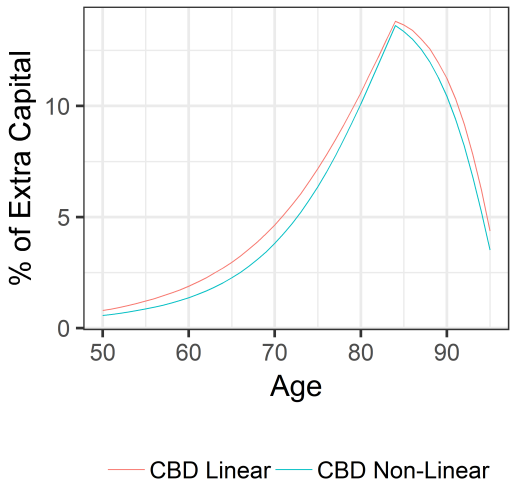


(a) Average over countries of the Percentage of extra capital required across ages $50 - 95$.

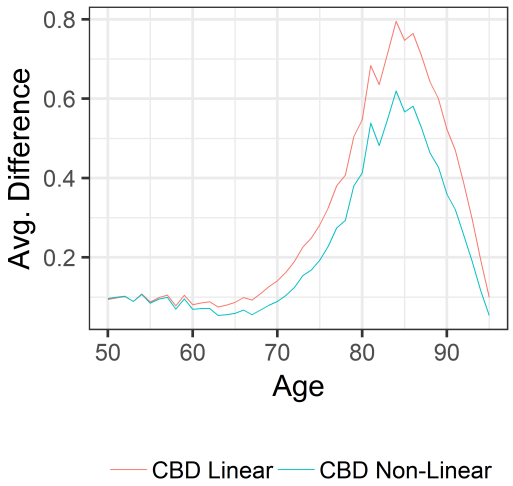


(b) Average difference of Realised annuity and the upper 99.5% predicted annuity liability across ages $50 - 95$ under both Lee-Carter model variants

Figure 2: LC model



(a) Average over countries of the Percentage of extra capital required across ages $50 - 95$.



(b) Average difference of Realised annuity and the upper 99.5% predicted annuity liability across ages $50 - 95$ under both CBD model variants

Figure 3: CBD model

annuity liability.[25]

Table 4 contains the BF and BLRT statistic for each variant of the LC and CBD model by tallying the "ones" from Appendix Table 6. A higher value for BF_{01} represents evidence favouring H_0, meaning that indeed the violations occur at frequency of p^*. For values smaller than 1, it shows evidence towards H_1, that violations do not follow a frequency of p^*. For the BLRT statistic, we compare it against the χ^2-distribution with 1 degrees of freedom. At the 5% region, the critical value (C_{BLRT}) is 2.841. A variety of prior specifications was conducted as a means of robustness check, however the goal of choosing a diffuse prior still applies. Besides the Beta(0.5,0.5) distribution corresponding to Jeffreys Prior, the Beta($\bar{\epsilon}$,$\bar{\epsilon}$) known as Haldane's prior and the Neutral-Information (NI) Beta($\frac{1}{3}$,$\frac{1}{3}$) prior was also used.[26]

Table 4: Derived BF and BLRT under the linear and nonlinear LC and CBD models across varying prior hyper-parameter specifications.

LC model						
	Linear			**Nonlinear**		
Hyper-Parameters	Haldanes-Beta($\bar{\epsilon}$,$\bar{\epsilon}$)	NI-Beta($\frac{1}{3}$, $\frac{1}{3}$)	Jeffreys-Beta(0.5,0.5)	Haldanes-Beta($\bar{\epsilon}$,$\bar{\epsilon}$)	NI-Beta($\frac{1}{3}$, $\frac{1}{3}$)	Jeffreys-Beta(0.5,0.5)
Bayes Factor	8.472522×10^{-09}	2.556865×10^{-08}	4.431379×10^{-08}	2.8695×10^{-06}	9.2722×10^{-09}	1.6618×10^{-05}
BLRT	37.0473	37.0613	37.0635	25.1922	25.2094	25.2120
$\hat{p}$	0.039	0.039	0.039	0.031	0.031	0.031
CBD model						
	Linear			**Nonlinear**		
Hyper-Parameters	Haldanes-Beta($\bar{\epsilon}$,$\bar{\epsilon}$)	NI-Beta($\frac{1}{3}$, $\frac{1}{3}$)	Jeffreys-Beta(0.5,0.5)	Haldanes-Beta($\bar{\epsilon}$,$\bar{\epsilon}$)	NI-Beta($\frac{1}{3}$, $\frac{1}{3}$)	Jeffreys-Beta(0.5,0.5)
Bayes Factor	0.0000	0.3492	1.4415	0.5394	3.3818	8.2784
BLRT	5.1504	4.4837	4.1510	-0.0790	0.0350	0.0430
$\hat{p}$	NA	NA	NA	0.005	0.005	0.005

In Table 4, the outcomes of both test statistics and $\hat{p}$ is shown. Both the LC models are presented with a rejection of H_0 under the BF and BLRT, meaning that a frequency of p^* was rejected. In particular since the BF can be compared between models, we see that the nonlinear version of the LC model presents a stronger model compared with its linear variant due to its larger BF, with its resulting $\hat{p}$ closer to p^*. The nonlinear CBD model performed the best out of the group, this can be seen from the BF and BLRT statistic giving evidence for H_0 and the BF being the largest of the models. Its implicit $\hat{p}$ of 0.005 in Table 4 also shows that the model achieved the correct mortality coverage. It is interesting to note that although the linear CBD model was not able to achieve the correct coverage, it was due to it having no violations across all countries and ages. This meant that the linear CBD model overestimated the mortality rate and was able to capture the realised longevity liability in all cases. From a pricing perspective, although the linear variant of the CBD

[25]For the full table of results see SI Section 2.

[26]Here, $\bar{\epsilon}$ represents a small positive real number ($\bar{\epsilon} \in \mathbb{R}^+$).

model captured all realised liabilities under its mortality forecasts, the liability estimates over the long run would over compensate for more than 0.5% under the longevity stress trend scenario. In a scenario where we are trying to minimize the capital required in order to achieve a 0.5% coverage, the nonlinear CBD model out-performed all models. With varying prior specifications, the test outcomes stayed mostly consistent. Under Jeffreys prior, the linear CBD model also seems to fail to reject the H_0 hypothesis, however if we compare that to the nonlinear CBD model, failure to reject the H_0 is more than 8 times more likely. As pointed out by Li et al. (2014), the BLRT seems to have the least deviations with prior changes, mainly due to its loss function being averaged over the posterior rather than the prior distribution as in the BF. Furthermore, as seen in our current application, the data obtained does not follow a time series. However, if we were to use time series such as (monthy) mortality data from one specific country, we could use the Independence or CC tests to detect mortality cohort effects.

6. Conclusion

The work in this paper contributes to three gaps in the literature. Firstly, parameter uncertainty is of particular importance when undergoing analysis in risk management. While there are many methods for capturing parameter uncertainty, the Bayesian estimation method comes naturally via Bayes theorem. In addition, we impose the Poisson and Binomial nonlinear Bayesian state-space framework for the LC and CBD models in which we efficiently sample the latent variables in block. We showed that we are able to achieve efficient simulation of the posterior by utilizing both the EKF and MH step within the MCMC. Lastly, we utilized tools from Bayesian decision theory to determine test outcomes from the UC test and as an application we used annuity liabilities in a longevity stressed scenario from two prominent mortality models, the LC model and CBD model.

Secondly, there has been a wide range of VaR backtesting procedures established to evaluate certain VaR models. One example is the UC test by Kupiec (1995) which aims to find whether VaR violations occur at the correct VaR coverage. Other examples are the Independence and CC tests proposed by Christoffersen (1998). In this regard, we develop a novel approach to backtesting under the Bayesian paradigm which can be an excellent tool for decision and policy makers and its demonstrated effectiveness establishes the foundations for a much broader area of applications.

As a natural extension of our work, one could consider the multivariate version of our newly proposed backtest which would have to account for possible correlations in VaR-violations across different types of annuities and time. Moreover, we want to extend the tests for i.i.d. VaR-violations,

and of course, for conditional coverage to annuity pricing under the same Bayesian decision theoretic framework. As a result in doing so, we could propose new means to determine cohort effects through the detection of consecutive VaR violations. As these issues are beyond the scope of the present paper, we will address them in our future research.

Acknowledgement: We are extremely grateful to the four anonymous reviewers and handling editor Prof. Emanuele Borgonovo who have afforded us considerable assistance in enhancing both the quality of the findings and the clarity of their presentation. The authors would like to thank the participants at the Quantitative Finance and Risk Analysis (QFRA) 2019 and seminars at Boston University, University of Liverpool, Monash University and Shanghai University for helpful comments. Any remaining errors are ours.

Supplementary Information: 72-page supplementary material associated with this article can be found in the online version.

Berger, J. O. and Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of p values and evidence. *Journal of the American Statistical Association*, 82(397):112–122.

Berkowitz, J. and O'Brien, J. (2002). How accurate are value-at-risk models at commercial banks? *The Journal of Finance*, 57(3):1093–1111.

Cairns, A., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from england and wales and the united states. *North American Actuarial Journal*, 13(1):1–35.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81(3):541–553.

Christoffersen, P. (2009). Value–at–risk models. In *Handbook of Financial Time Series*, pages 753–766. Springer.

Christoffersen, P. and Pelletier, D. (2004). Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics*, 2(1):84–108.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 4(39):841–862.

Czado, C., Delwarde, A., and Denuit, M. (2005). Bayesian poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics*, 36(3):260–284.

Debon, A., Montes, F., and Puig, F. (2008). Modelling and forecasting mortality in Spain. *European Journal of Operational Research*, 189(3):624 – 637.

Dowd, K., Blake, D., Cairns, A., Coughlan, G., Epstein, D., and Khalaf-Allah, M. (2010). Evaluating the goodness of fit of stochastic mortality models. *Insurance: Mathematics and Economics*, 47(3):255–265.

Drenovak, M., Ranković, V., Ivanović, M., Urošević, B., and Jelic, R. (2017). Market risk management in a post-Basel II regulatory environment. *European Journal of Operational Research*, 257(3):1030–1044.

Eckert, J. and Gatzert, N. (2018). Risk-and value-based management for non-life insurers under solvency constraints. *European Journal of Operational Research*, 266(2):761–774.

Fung, M. C., Peters, G. W., and Shevchenko, P. V. (2017). A unified approach to mortality modelling using state-space framework: characterisation, identification, estimation and forecasting. *Annals of Actuarial Science*, 11(2):343–389.

Glasserman, P., Heidelberger, P., and Shahabuddin, P. (2002). Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance*, 12(3):239–269.

Goodman, S. N. (2001). Of p-values and bayes: a modest proposal. *Epidemiology*, 12(3):295–297.

Hari, N., De Waegenaere, A., Melenberg, B., and Nijman, T. E. (2008). Longevity risk in portfolios of pension annuities. *Insurance: Mathematics and Economics*, 42(2):505–519.

Harvey, C. R. (2017). Presidential address: the scientific outlook in financial economics. *The Journal of Finance*, 72(4):1399–1440.

Jorion, P. (2000). *Value at risk*. McGraw-Hill Professional Publishing.

Kogure, A. and Kurachi, Y. (2010). A bayesian approach to pricing longevity risk based on risk-neutral predictive distributions. *Insurance: Mathematics and Economics*, 46(1):162–172.

Konicz, A. K. and Mulvey, J. M. (2015). Optimal savings management for individuals with defined contribution pension plans. *European Journal of Operational Research*, 243(1):233–247.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2):73–84.

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

Leung, M., Fung, M. C., and O'Hare, C. (2018). A comparative study of pricing approaches for longevity instruments. *Insurance: Mathematics and Economics*, 82:95 – 116.

Li, Y., Zeng, T., and Yu, J. (2014). A new approach to bayesian hypothesis testing. *Journal of Econometrics*, 178:602–612.

Longin, F. M. (2000). From value at risk to stress testing: The extreme value approach. *Journal of Banking & Finance*, 24(7):1097–1130.

Millossovich, P., Villegas, A. M., and Kaishev, V. K. (2017). Stmomo: An r package for stochastic mortality modelling. *Journal of Statistical Software*.

Nieto, M. R. and Ruiz, E. (2016). Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting*, 32(2):475–501.

Pedroza, C. (2006). A bayesian forecasting model: predicting us male mortality. *Biostatistics*, 7(4):530–550.

Plat, R. (2011). One-year value-at-risk for longevity and mortality. *Insurance: Mathematics and Economics*, 49(3):462–470.

Wang, P., Pantelous, A. A., and Vahid, F. (2020). Multi-population mortality projection: The augmented common factor model with structural breaks. *Available at SSRN 3487386*.

Wied, D., Weiß, G. N., and Ziggel, D. (2016). Evaluating value-at-risk forecasts: A new set of multivariate backtests. *Journal of Banking & Finance*, 72:121–132.

Ziggel, D., Berens, T., Weiß, G. N., and Wied, D. (2014). A new set of improved value-at-risk backtests. *Journal of Banking & Finance*, 48:29–41.

Appendix A Linear variants of the LC and CBD model

A.1 Linear variant of the LC model

The LC model in linear state-space form is given by

$$\boldsymbol{y_t} = \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\varepsilon_t}, \quad \boldsymbol{\varepsilon_t} \sim N(0, \mathbb{I}_n \sigma_\varepsilon^2) \tag{A.1}$$

$$\kappa_t = \kappa_{t-1} + \delta + \omega_t, \quad \omega_t \sim N(0, \sigma_\omega^2), \tag{A.2}$$

with the static model parameter vector $\Theta_{\text{LC}} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \delta, \sigma_\omega^2, \sigma_\varepsilon^2\}$. Recall that in the Bayesian setting, our aim is to draw samples from the joint posterior density $\pi(\kappa_{1:T}, \Theta_{\text{LC}} | \boldsymbol{y}_{1:T})$, using Gibbs sampling, our MCMC procedure consists of

1. Initialise $\Theta_{\text{LC}}^{(0)}$ and $\kappa_{1:T}^{(0)}$.
2. For $i = 1, \ldots, M$,
   (a) Sample $\boldsymbol{\kappa}^{(i)}$ from $\pi(\boldsymbol{\kappa} | \Theta_{\text{LC}}^{(i-1)}, \boldsymbol{y}_{1:T})$.
   (b) Sample $\Theta_{\text{LC}}^{(i)}$ from $\pi(\Theta_{\text{LC}} | \boldsymbol{\kappa}^{(i)}, \boldsymbol{y}_{1:T})$.

A sample of the conditional distribution $\pi(\boldsymbol{\kappa} | \Theta_{\text{LC}}, \boldsymbol{y}_{1:T})$ can be obtained via forward-backward sampling using Kalman filtering (Carter and Kohn, 1994). To draw samples from $\pi(\Theta_{\text{LC}} | \boldsymbol{\kappa}, \boldsymbol{y}_{1:T})$, we assume the following conjugate prior distributions:

$$\pi(\delta) \sim N(\mu_\theta, \sigma_\theta^2),$$

$$\pi(\sigma_\varepsilon^2) \sim I.G(a_\varepsilon, b_\varepsilon)^{27}, \quad \pi(\sigma_\omega^2) \sim I.G(a_\omega, b_\omega)$$

$$\pi(\alpha_x) \sim N(\mu_\alpha, \sigma_\alpha^2), \quad \pi(\beta_x) \sim N(\mu_\beta, \sigma_\beta^2) \text{ for } x \in \{(x_1, ..., x_n).$$

The prior distributions were chosen such that when multiplied by the likelihood function, the resulting posterior distribution will be of the same family; this is known as the conjugacy property and it facilitates the Gibbs sampling procedure. In the case where no conjugacy is involved, the Metropolis-Hastings (MH) algorithm can be applied. The full conditional posterior distribution for Θ_{LC} are as follows[28]:

$$\pi(\alpha_x | \boldsymbol{y}, \boldsymbol{\kappa}, \boldsymbol{\beta}, \sigma_\varepsilon^2) \sim N\left(\mu_\alpha \sigma_\varepsilon^2 + \sigma_\alpha^2 \sum_{t=1}^{T} (y_{x,t} - \beta_x \kappa_t)(\sigma_\alpha^2 T + \sigma_\varepsilon^2)^{-1}, (\sigma_\alpha^2 T + \sigma_\varepsilon^2)(\sigma_\alpha^2 \sigma_\varepsilon^2)^{-1}\right)$$

$$\pi(\beta_x | \boldsymbol{y}, \boldsymbol{\kappa}, \boldsymbol{\alpha}, \sigma_\varepsilon^2) \sim N\left((\mu_\beta \sigma_\varepsilon^2 + \sigma_\beta^2 \sum_{t=1}^{T} (y_{x,t} - \alpha_x)\kappa_t)(\sigma_\beta^2 \sum_{t=1}^{T} (\kappa_t^2 + \sigma_\varepsilon^2))^{-1}, (\sigma_\beta^2 \sigma_\varepsilon^2)(\sigma_\beta^2 \sum_{t=1}^{T} (\kappa_t^2 + \sigma_\varepsilon^2))^{-1}\right)$$

[27] $I.G.$ represents the Inverse Gamma distribution hereafter.
[28] For a full derivation of posterior parameters and MCMC algorithm see Fung et al. (2017).

$$\pi(\sigma_\varepsilon^2|\boldsymbol{y},\boldsymbol{\kappa},\boldsymbol{\beta},\boldsymbol{\alpha}) \sim I.G(a_\varepsilon + \tfrac{Tn}{2},\ \ b_\varepsilon + \tfrac{1}{2}\sum_{t=1}^{T}\sum_{x=x_1}^{x_n}(y_{x,t} - (\alpha_x + \beta_x\kappa_t))^2)$$

$$\pi(\delta|\boldsymbol{y},\boldsymbol{\kappa},\sigma_\omega^2) \sim N\left((\mu_\delta\sigma_\omega^2 + \sigma_\delta^2\sum_{t=1}^{T}(\kappa_t - \kappa_{t-1}))(\sigma_\delta^2\sigma_\omega^2)^{-1},\ (\sigma_\delta^2\sigma_\omega^2)(T\sigma_\delta^2 + \sigma_\omega^2)^{-1}\right)$$

$$\pi(\sigma_\omega^2|\boldsymbol{y},\boldsymbol{\kappa},\delta) \sim I.G(a_\omega + \tfrac{T}{2},\ b_\omega + (\kappa_t - (\kappa_{t-1} + \delta))^2)$$

A.2 Linear variant of the CBD model

Since the true death probabilities, $q_{x,t}$, are unobservable, we can instead model the observable crude death probabilities, $\tilde{q}_{x,t}$, estimated using Eq. (4.1), which allows the CBD model to directly follow a linear structure shown in Eqs. (4.5) and (4.6). For convenience, let $\Theta_{\text{CBD}} = (\theta_1, \theta_2, \sigma_\nu^2, \Sigma)$ denote the static parameter vector for the CBD model in Eq. (4.5) and with the introduction of an error component. Let $y_{x,t} := \ln(\tilde{q}_{x,t}/(1-\tilde{q}_{x,t}))$, then the CBD model in linear state-space representation is given by

$$\begin{bmatrix} y_{x_1,t} \\ \vdots \\ y_{x_n,t} \end{bmatrix} = \begin{bmatrix} 1 & (x_1 - \bar{x}) \\ \vdots & \vdots \\ 1 & (x_n - \bar{x}) \end{bmatrix} \begin{bmatrix} \kappa_{1,t} \\ \kappa_{2,t} \end{bmatrix} + \begin{bmatrix} \nu_{x_1,t} \\ \vdots \\ \nu_{x_n,t} \end{bmatrix}, \quad \begin{bmatrix} \nu_{x_1,t} \\ \vdots \\ \nu_{x_n,t} \end{bmatrix} \sim\ N(0, \mathbb{I}_n\sigma_\nu^2), \tag{A.3}$$

$$\begin{bmatrix} \kappa_{1,t} \\ \kappa_{2,t} \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} \kappa_{1,t-1} \\ \kappa_{2,t-1} \end{bmatrix} + \begin{bmatrix} \omega_{1,t} \\ \omega_{2,t} \end{bmatrix}, \quad \begin{bmatrix} \omega_{1,t} \\ \omega_{2,t} \end{bmatrix} \sim N(0, \Sigma), \tag{A.4}$$

where $\mathbb{I}_n$ represents the $n \times n$ identity matrix. Eqs. (A.3) and (A.4) correspond to the measurement equation and the state equation respectively. A measurement error term, $\nu_{x,t}$, was included in Eq. (A.3) to facilitate the linear Gaussian state-space model estimation. Since model (A.3) and (A.4) belongs to the class of linear and Gaussian state-space models, we can perform MCMC estimation of the model utilizing a multivariate Kalman filter. Similar to the case in the LC model, our aim is to draw samples from the joint posterior density $\pi(\boldsymbol{\kappa}_{1:T}, \Theta_{\text{CBD}}|\boldsymbol{y}_{1:T})$ using Gibbs sampling which is as follows:

1. Initialise $\Theta_{\text{CBD}}^{(0)}$ and $\boldsymbol{\kappa}_{1:T}^{(0)}$.
2. For $i = 1, \ldots, M$,
   (a) Sample $\boldsymbol{\kappa}_{1:T}^{(i)}$ from $\pi(\boldsymbol{\kappa}_{1:T}|\Theta_{\text{CBD}}^{(i-1)}, \boldsymbol{y}_{1:T})$.
   (b) Sample $\Theta_{\text{CBD}}^{(i)}$ from $\pi(\Theta_{\text{CBD}}|\boldsymbol{\kappa}_{1:T}^{(i)}, \boldsymbol{y}_{1:T})$.

A sample from $\pi(\boldsymbol{\kappa}_{1:T}|\Theta_{\text{CBD}}, \boldsymbol{y}_{1:T})$ can be obtained via a multivariate forward-backward sampling. To draw samples from the full conditional posterior distributions, we assume the following priors for Θ_{CBD},

$$\pi(\sigma_\nu^2) \sim I.G(a_\nu, b_\nu), \quad \pi(\theta_i) \sim N(\mu_{\theta_i}, \Sigma_{\theta_i}), \quad i = 1, 2,$$

$$\pi(\Sigma|(\sigma_1^2, \sigma_2^2)) \sim I.W\left((\nu+2)-1, 2\xi \text{ diag}\left(\tfrac{1}{\sigma_1^2}, \tfrac{1}{\sigma_2^2}\right)\right),$$

$$\pi(\sigma_k^2) \overset{indep}{\sim} I.G\left(\tfrac{1}{2}, \tfrac{1}{A_k}\right), \quad k = 1, 2,$$

where $I.W$ corresponds to the Inverse Wishart distribution, A_k are hyper-parameters, and the notation $\overset{indep}{\sim}$ corresponds to "independently distributed". For more information on the MCMC algorithm and posterior derivations, see Leung et al. (2018). Using the prior distributions described above, the posterior distributions for the static parameters are given by:

$$\pi(\sigma_\nu^2|\boldsymbol{y}, \boldsymbol{\kappa}_{1:T}) \sim I.G(a_\nu + \tfrac{Tn}{2}, \;\; b_\nu + \tfrac{1}{2}\sum_{t=1}^{T}\sum_{x=x_1}^{x_n}(y_{x,t} - (\kappa_{1,t} + (x-\bar{x})\kappa_{2,t}))^2),$$

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{\kappa}_{1:T}, \Sigma) \sim N\left((\Sigma_\theta^{-1} + n\Sigma^{-1})^{-1}\left({\Sigma_\theta}^{-1}\mu_\theta + n\Sigma^{-1}\textstyle\sum_{t=1}^{T}[\boldsymbol{\kappa}_t - \boldsymbol{\kappa}_{t-1}]\right), \left(\Sigma_\theta^{-1} + T\Sigma^{-1}\right)^{-1}\right),$$

$$\pi(\sigma_k^2|\Sigma) \overset{i.i.d}{\sim} I.G(\tfrac{\xi+T}{2}, \xi[\Sigma^{-1}]_{kk} + \tfrac{1}{(A_k)^2}) \text{ for } k \in (1, 2),$$

$$\pi(\Sigma|\sigma_1^2, \sigma_2^2, \boldsymbol{y}, \boldsymbol{\kappa}_{1:T}, \boldsymbol{\theta}) \sim I.W(\xi + T + n - 1, 2\xi diag(\tfrac{1}{\sigma_1^2}, \tfrac{1}{\sigma_2^2}) + \textstyle\sum_{t=1}^{T}\left[\boldsymbol{\kappa_t} - \boldsymbol{\theta}\right]\left[\boldsymbol{\kappa_t} - \boldsymbol{\theta}\right]'),$$

where $[\Sigma^{-1}]_{kk}$ denotes the (k,k) element of $[\Sigma^{-1}]$. Derivations of these posteriors are provided in SI Section 1.5. The choice of a hierarchical prior for Σ is to circumvent the issue of the Inverse-Wishart prior leading to a biased estimator for the correlation coefficient when the variances are small.[29]

Appendix B MCMC Algorithm for the Lee-Carter Poisson model

(i) Set initial values for the parameter vector $(\boldsymbol{\alpha}^0,\ \boldsymbol{\beta}^0, \boldsymbol{\kappa}^0,\ \delta^0,\ (\sigma_\omega^2)^0,\ (\sigma_\beta^2)^0)$.

(ii) Conditional on $(\boldsymbol{\alpha}^{(i-1)},\ \boldsymbol{\beta}^{(i-1)},\ \delta^{(i-1)})$, apply the EKF and Backward Smoother shown in Algorithm 1 of SI to obtain $\boldsymbol{\kappa}^*$ as the candidate for step (iii).

(iii) Simulate $\boldsymbol{\kappa}^{(i)}$ using the MH step for κ_t shown in Algorithm 2 of SI, the candidate draw is taken from step (ii) to obtain the density f^* and $\boldsymbol{\kappa}^{(i-1)}$ is used to compute the density f.

(iv) Simulate $\boldsymbol{\beta}^{(i)}$ using the MH step for β_x shown in Algorithm 3 of SI.

(v) Draw $\boldsymbol{\alpha}^{(i)}$ from the transformed posterior $\text{LogGamma}(a_\alpha + \sum_t \text{d}_{x,t}, b_\alpha + \sum_t \text{E}_{x,t}\exp(\beta_x^{(i-1)}\kappa_t^{(i-1)}))$.

(vi) Draw $\delta^{(i)}$ from its posterior distribution $N(\mu, \sigma^2)$, with $\mu = 1/(\frac{1}{100} + \frac{T-1}{(\sigma_\omega^2)^{(i-1)}})(1/((\sigma_\omega^2)^{(i-1)}))\sum_t(\kappa_t^{(i)} - \kappa_{t-1}^{(i)})$ and, $\sigma^2 = (\frac{1}{100} + \frac{T-1}{(\sigma_\omega^2)^{(i-1)}})(1/(\sigma_\omega^2)^{(i-1)})$.

[29] For more details the reader is referred to Section 2 of Leung et al. (2018).

(vii) Draw $(\sigma_\beta^2)^{(i)}$ from its posterior distribution $I.G(a_\beta + n/2, b_\beta + (1/2)\boldsymbol{\beta}\boldsymbol{\beta}')$.

(viii) Draw $(\sigma_\omega^2)^{(i)}$ from $I.G(a_\omega + (T-1)/2, b_\omega + (1/2)\sum_t((\kappa_t^{(i)} - \kappa_{t-1}^{(i)}) - \delta^{(i)})^2)$.

(ix) Conditional on $\boldsymbol{\kappa}^{(i)}, \boldsymbol{\alpha}^{(i)}, \boldsymbol{\beta}^{(i-1)}, (\sigma_\omega^2)^{(i)}, (\sigma_\beta^2)^{(i)}$, simulate $\boldsymbol{\beta}^{(i)}$ using Algorithm 3 with tuning parameter σ_x^2.

Appendix C MCMC Algorithm for the CBD Binomial model

(i) Set initial values for the parameter vector $(\boldsymbol{\kappa_1^0}, \boldsymbol{\kappa_2^0}, \ \boldsymbol{\theta}^0, \ \Sigma^0)$.

(ii) Conditional on $(\boldsymbol{\theta}^{(i-1)}, \ \Sigma^{(i-1)})$, apply the EKF and Backward Smoother shown in Algorithm 4 of SI to obtain $\boldsymbol{\kappa^*}$ as the candidate for step (iii).

(iii) Simulate $\boldsymbol{\kappa}^{(i)}$ using the MH step for $\boldsymbol{\kappa}$ shown in Algorithm 5 of SI. The candidate draw, $\boldsymbol{\kappa^*}$, is taken from step (ii) to obtain the density f^* and $\boldsymbol{\kappa}^{(i-1)}$ is used to compute the density f.

(v) Draw $\boldsymbol{\theta}^{(i)}$ from its posterior distribution $\pi(\boldsymbol{\theta^{(i)}}|\boldsymbol{y}, \boldsymbol{\kappa_1}^{(i)}, \boldsymbol{\kappa_2}^{(i)}, \Sigma^{(i-1)})$.

(vi) Draw $\Sigma_{kk}^{(i)}$ from its posterior distribution $\pi(\Sigma_{kk}^{(i)}|\boldsymbol{y}, \boldsymbol{\kappa_1}^{(i)}, \boldsymbol{\kappa_2}^{(i)}, \boldsymbol{\theta^{(i)}})$ for $k \in 1, 2$.

(vii) Draw $\Sigma^{(i)}$ from its posterior distribution $\pi(\Sigma_\omega^{(i)}|\Sigma_{kk}^{(i)}, \boldsymbol{y}, \boldsymbol{\kappa_1}^{(i)}, \boldsymbol{\kappa_2}^{(i)}, \boldsymbol{\theta^{(i)}})$.

Appendix D Bayesian Testing

D.1 Proof of Lemma 4.1

When testing two hypotheses, H_0 and H_1, where H_0 is a point null hypothesis and H_1 is a composite hypothesis, the BF is given by:

$$\frac{l(\boldsymbol{y}|\theta = \theta_0)}{\int_\theta \pi(\boldsymbol{y}|\theta, H_1)\pi(\theta|H_1)d\theta}.$$

Under the UC backtest assumptions, the sequence of ones and zeros are assumed to follow a Bernoulli distribution, hence the likelihood function for the data corresponds to

$$l(\mathcal{I}|p) = \prod_{i=1}^{m} p^{y_i}(1-p)^{1-y_i} = p^{m_1}(1-p)^{m_0},$$

where m is the number of data points, m_1 represents the number of "ones" in the data, likewise m_0 corresponds to the number of "zeros". The null hypothesis is defined as $p = p^*$ for a given p^*, and

the alternative hypothesis as $p \neq p^*$. With the following prior specifications for p:

$$\pi(p) = \begin{cases} 1 & \text{if } p = p^*, \\ \text{Beta}(a, b), & \text{if } p \neq p^*. \end{cases}$$

We derive an analytical form for the BF due to the conjugate property of the Beta-Binomial distribution, i.e.,

$$\begin{aligned} \frac{l(\boldsymbol{y}|p = p^*)}{\int_{p \neq p^*} l(\boldsymbol{y}|\theta)\pi(\theta)d\theta} &= \frac{(p^*)^{m_1}(1-p^*)^{m_0}}{\int_{p \neq p^*} p^{m_1}(1-p)^{m_0}\text{Beta}(a,b)dp} \\ &= \frac{(p^*)^{m_1}(1-p^*)^{m_0}}{\beta(m_1+a, m_0+b)\int_{p \neq p^*} \frac{p^{m_1+a-1}(1-p)^{m_0+b-1}}{\beta(m_1+a, m_0+b)}dp} \\ &= \frac{(p^*)^{m_1}(1-p^*)^{m_0}}{\beta(m_1+a, m_0+b)}. \end{aligned}$$

D.2 Proof of Theorem 4.1

To derive the analytical form for the BLRT for the UC test, we begin with the test statistic,

$$T_{\text{BLRT}}(\boldsymbol{y}, \theta) = -2\left[\underbrace{\int_\theta \log(l(\boldsymbol{y}|\theta_0))\pi(\theta|\boldsymbol{y})d\theta}_{(1)} - \underbrace{\int_\theta \log(l(\boldsymbol{y}|\theta))\pi(\theta|\boldsymbol{y})d\theta}_{(2)}\right] + 1.$$

Here, it can be shown that integral (1) can be decomposed to the following:

$$\begin{aligned} \int_\theta \log(l(\boldsymbol{y}|\theta_0))\pi(\theta|\boldsymbol{y})d\theta &= \int_\theta \log((\theta^*)^{m_1}(1-\theta^*)^{m_0})\frac{\theta^{a+m_1-1}(1-\theta)^{b+m_0-1}}{\beta(a+m_1, b+m_0)}d\theta \\ &= m_1\log(\theta^*) + m_0\log(1-\theta^*). \end{aligned} \tag{D.1}$$

Now, expanding integral (2), we have:

$$\int_\theta \log(l(\boldsymbol{y}|\theta))\pi(\theta|\boldsymbol{y})d\theta = \int_\theta \left[m_1\log(\theta) + m_0\log(1-\theta)\right]\frac{\theta^{a+m_1-1}(1-\theta)^{b+m_0-1}}{\beta(a+m_1, b+m_0)}d\theta.$$

$$\int_\theta m_1\log(\theta)\frac{\theta^{a+m_1-1}(1-\theta)^{b+m_0-1}}{\beta(a+m_1, b+m_0)}d\theta = \frac{m_1}{\beta(a+m_1, b+m_0)}\int_\theta \log(\theta)\theta^{a+m_1-1}(1-\theta)^{b+m_0-1}d\theta$$

$$
\begin{aligned}
&= \frac{m_1}{\beta(a+m_1, b+m_0)} \int_\theta \frac{\partial \theta^{a+m_1-1}}{\partial a} (1-\theta)^{b+m_0-1} d\theta \\
&= \frac{m_1}{\beta(a+m_1, b+m_0)} \frac{\partial \beta(a+m_1, b+m_0)}{\partial a} \\
&= \frac{m_1 \partial \log(\beta(a+m_1, b+m_0))}{\partial a} \\
&= \frac{m_1 \partial \log(\Gamma(a+m_1))}{\partial a} - \frac{m_1 \partial \log(\Gamma(a+m_1+b+m_0))}{\partial a} \\
&= m_1(\psi(a+m_1) - \psi(a+m_1+b+m_0)).
\end{aligned} \tag{D.2}
$$

$$
\begin{aligned}
\int_\theta m_0 \log(1-\theta) \frac{\theta^{a+m_1-1}(1-\theta)^{b+m_0-1}}{\beta(a+m_1, b+m_0)} d\theta &= \int_\theta m_0 \log(1-\theta) \frac{\theta^{a+m_1-1}(1-\theta)^{b+m_0-1}}{\beta(a+m_1, b+m_0)} d\theta \\
&= \frac{m_0}{\beta(a+m_1, b+m_0)} \int_\theta \log(\theta) \theta^{a+m_1} (1-\theta)^{b+m_0} d\theta \\
&= \frac{m_0}{\beta(a+m_1, b+m_0)} \int_\theta \frac{\partial \theta^{a+m_1-1}}{\partial a} (1-\theta)^{b+m_0-1} d\theta \\
&= \frac{m_0}{\beta(a+m_1, b+m_0)} \frac{\partial \beta(a+m_1, b+m_0)}{\partial a} \\
&= \frac{m_0 \partial \log(\beta(a+m_1, b+m_0))}{\partial a} \\
&= \frac{m_0 \partial \log(\Gamma(a+m_1))}{\partial a} - \frac{m_1 \partial \log(\Gamma(a+m_1+b+m_0))}{\partial a} \\
&= m_0(\psi(b+m_0) - \psi(a+m_1+b+m_0)).
\end{aligned} \tag{D.3}
$$

Combing Eqs. (D.1), (D.2), and (D.3), we retrieve the analytical form for the BLRT statistic.

$$
\begin{aligned}
T_{\text{BLRT}}(\boldsymbol{y}, \theta) = -2[m_1 \log(\theta^*) + m_0 \log(1-\theta^*) - (m_1(\psi(a+m_1) - \psi(a+m_1+b+m_0)) + \\
m_0(\psi(b+m_0) - \psi(a+m_1+b+m_0)))] + 1,
\end{aligned} \tag{D.4}
$$

as required.

D.3 Proof of Lemma 4.2

Let the Multinomial data generating process be defined as,

$$
f(\boldsymbol{\mathcal{I}}|P) = \prod_{i=0}^{1} \prod_{j=0}^{1} p_{ij}^{n_{ij}},
$$

where n_{ij} corresponds to the number of times a transition from state i to state j occurred. Let P be the transition matrix, with elements p_{ij} for $i, j = (0, 1)$ being the transition probability from state i to state j. We impose the Dirichlet prior on p_{ij} for $i, j = (0, 1)$, as it is a natural conjugate for the Multinomial distribution. The Dirichlet distribution is given by,

$$f^{\text{DIR}}(\boldsymbol{x}) = \frac{1}{\boldsymbol{B}(\boldsymbol{\alpha})} \prod_{i=0}^{N-1} x_i^{\alpha_i - 1},$$

where $\boldsymbol{B}(\boldsymbol{\alpha})$ is the Multivariate Beta function evaluated by computing $\frac{\sum_i \Gamma(\alpha_i)}{\Gamma(\sum_i \alpha_i)}$. Let $p_j = p_{0j} + p_{1j}$, then, $n_j = n_{0j} + n_{1j}$ for $j = 0, 1$. Assuming we want to test for $p_0 = p_1$, we have

$$\begin{aligned}
\pi(P|\mathcal{I}) = & \frac{f(\boldsymbol{\mathcal{I}}|p)\pi(p)}{f(y)} \propto f(\mathcal{I}|p)\pi(p) \\
\propto & \prod_{i=0}^{1}\prod_{j=0}^{1} p_{ij}^{n_{ij}} \frac{1}{\boldsymbol{B}(\boldsymbol{\alpha})} \prod_{i=0}^{1}\prod_{j=0}^{1} p_{ij}^{\alpha_0 - 1} \\
\propto & (p_{00} + p_{10})^{n_{00}+n_{10}+\alpha_0-1}(1 - (p_{00} + p_{10}))^{n_{01}+n_{11}+\alpha_0-1} \qquad \text{(D.5)}
\end{aligned}$$

Eq. (D.5) contains the kernel of a $\beta(n_{00} + n_{10} + \alpha_0,\ n_{01} + n_{11} + \alpha_0)$ distribution. Now, assuming $p_0 \neq p_1$, we have

$$\begin{aligned}
\pi(P|\mathcal{I}) = & \frac{f(\boldsymbol{\mathcal{I}}|p)\pi(p)}{f(y)} \\
\propto & f(\mathcal{I}|p)\pi(p) \\
\propto & \prod_{i=0}^{1}\prod_{j=0}^{1} p_{ij}^{n_{ij}} \frac{1}{\boldsymbol{B}(\boldsymbol{\alpha})} \prod_{i=0}^{1}\prod_{j=0}^{1} p_{ij}^{\alpha_i - 1} \\
\propto & \prod_{i=0}^{1}\prod_{j=0}^{1} p_{ij}^{n_{ij}} \frac{1}{\boldsymbol{B}(\boldsymbol{\alpha})} \prod_{i=0}^{1}\prod_{j=0}^{1} p_{ij}^{\alpha_i - 1} \\
\propto & (1 - p_{01})^{n_{00}-\alpha_0-1}(p_{01})^{n_{10}-\alpha_0-1}(1 - p_{11})^{n_{01}+\alpha_1-1}(p_{11})^{n_{11}-\alpha_1-1} \qquad \text{(D.6)}
\end{aligned}$$

Eq. (D.6) contains the kernel of two Beta distributions given by, $\beta(n_{00} + \alpha_0,\ n_{01} + \alpha_0)$ and $\beta(n_{10} + \alpha_1,\ n_{11} + \alpha_1)$.

Appendix E Simulation Results for the Bayes Factor under Unconditional Coverage

In this section we examine the performance of the new Bayesian backtest method in a finite sample situation. Our study aims to mimic a realistic scenario of sample sizes, namely, 1-year ($T = 252$),

4-year ($T = 1000$), 10-year ($T = 2500$) and lastly 20-year ($T = 5000$) periods. This simulation study is setup to determine whether our method is capable of detecting changes in VaR violations ranging from as expected (when the null is correct) till higher than expected (when the alternative is correct). For the UC test, the aim is to determine whether or not VaR violations occur with probability $p = p^*$. The first step is to simulate 30,000 sets of Bernoulli random variables (see Algorithm 3):

$$I_t(p) \sim \text{Bern}(\gamma \cdot p), \text{for } t = 1, .., T$$

of length T and probability p^* of success. In this case, a success in $I_t(p)$ would imply a VaR violation

Algorithm 3 Unconditional Coverage Simulation

1: Generate a T-length violation sequence ($\text{I}_{1:T}$), based on independent Bernoulli($p^* \cdot \gamma$) distribution.
2: Compute BF_{01} using Eq. (4.18), with $m_1 = \sum_{t=1}^{T} \text{I}_t$ and $m_0 = T - m_1$.
3: Rejection if $\text{BF}_{01} < 1$

has occurred. The γ coverage variable allows us to determine whether or not we correctly rejected the null hypothesis. When $\gamma = 1$, the VaR violations occur at the rate p^* which implies the null model was correct. As γ increases from 1 to 1.5, it increases the chance of VaR violations occurring more often than the null model of p^*.

The simulation results of the UC test are shown in Table 5. In our simulation study we choose $\gamma = 1$, 1.25, and, 1.5. Our VaR levels will be represented by $p^* = 0.01$ and 0.05. At $\gamma = 1$, we see that the BF false rejection rate decreases as sample size T increases. This is a consistency we should expect with the BF; as opposed to NHST, where a particular type-1 error α would need to be chosen and the false rejection rate of the null cannot be decreased with sample size. When the VaR is low at $p^* = 0.01$, Jeffreys prior seems to reject the null hypothesis less often than both Haldanes prior and NI prior. However, once we increase the VaR to 5% ($p^* = 0.05$), all three priors show consistent results. Lastly, we see that when VaR violation rates occur more frequently than p^*, γ increases, we see that chance of rejecting the null also increases in tandem. The outcomes of our simulation study shows there is some robustness with varying prior distributions.

Table 5: Shows the rejection rates for the BF under UC after 30,000 simulations using Algorithm 3. The three priors, Haldanes, NI and Jeffreys, were used to conduct the simulation.

Bayes Factor							
		VaR at 1%			VaR at 5%		
γ	**T**	**Haldanes-Beta**$(\varepsilon, \varepsilon)$	**NI-Beta**$(\frac{1}{3}, \frac{1}{3})$	**Jeffreys-Beta**$(0.5, 0.5)$	**Haldanes-Beta**$(\varepsilon, \varepsilon)$	**NI-Beta**$(\frac{1}{3}, \frac{1}{3})$	**NI-Beta**$(\frac{1}{3}, \frac{1}{3})$
1	252	0.08263	0.00457	0.00457	0.01497	0.01883	0.00797
	1000	0.01223	0.00543	0.00173	0.00703	0.00703	0.00460
	2500	0.00463	0.00463	0.00137	0.00447	0.00383	0.00273
	5000	0.00507	0.00323	0.00113	0.00297	0.00297	0.00197
1.25	252	0.05573	0.01387	0.01387	0.04577	0.07420	0.04533
	1000	0.03087	0.02967	0.01557	0.17993	0.17993	0.15043
	2500	0.05317	0.05317	0.02543	0.44973	0.44973	0.41850
	5000	0.12787	0.10250	0.06477	0.81030	0.81030	0.77647
1.5	252	0.05943	0.03793	0.03793	0.19177	0.26163	0.19170
	1000	0.12577	0.12570	0.08340	0.74127	0.74127	0.70160
	2500	0.30410	0.30410	0.20137	0.99000	0.99000	0.98783
	5000	0.65333	0.60763	0.51697	1.00000	1.00000	1.00000

Appendix F Annuity Liability Results

Table 6: Table showing the percentage of extra capital required, difference from realised annuity, and the indicator variable ${}_j\mathcal{I}_p^x(N)$ over all countries used under the linear and nonlinear LC and CBD models.

LC model: Linear variant

	Australia			United Kingdom			Italy			France			Spain			New Zealand			Sweden			Germany			Russia		
x	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$
62	1.0871	0.0345	0	1.3976	-0.0027	1	1.4167	0.0346	0	1.3211	0.0754	0	1.6937	0.1345	0	1.2647	0.0457	0	0.7653	0.0139	0	1.1181	0.0007	0	2.9916	0.1953	0
63	1.1863	0.0400	0	1.5136	-0.0126	1	1.5539	0.0459	0	1.4437	0.0793	0	1.8621	0.1560	0	1.3784	0.0342	0	0.8516	0.0229	0	1.2135	-0.0093	1	2.9308	0.0746	0
64	1.2891	0.0432	0	1.6438	-0.0277	1	1.7054	0.0484	0	1.5849	0.0882	0	2.0546	0.1594	0	1.4972	0.0355	0	0.9434	0.0225	0	1.3135	0.0030	0	2.8399	0.1390	0
65	1.4045	0.0434	0	1.7866	-0.0260	1	1.8837	0.0725	0	1.7383	0.1028	0	2.2676	0.1615	0	1.6369	0.0540	0	1.0441	0.0450	0	1.4354	0.0105	0	2.7367	0.1525	0
66	1.5266	0.0409	0	1.9488	-0.0182	1	2.0850	0.0875	0	1.9075	0.1149	0	2.5059	0.1859	0	1.7883	0.0644	0	1.1604	0.0651	0	1.5661	0.0700	0	2.6445	0.1601	0
67	1.6674	0.0451	0	2.1263	-0.0152	1	2.2985	0.1191	0	2.1026	0.1409	0	2.7705	0.1993	0	1.9546	0.0775	0	1.2778	0.0583	0	1.7244	0.0553	0	2.5420	0.0837	0
68	1.8089	0.0494	0	2.3145	-0.0063	1	2.5403	0.1292	0	2.3119	0.1550	0	3.0570	0.2183	0	2.1262	0.1207	0	1.4140	0.0878	0	1.8940	0.0893	0	2.5299	0.1208	0
78	3.6317	0.1937	0	4.8381	0.3167	0	6.0905	0.4478	0	5.3598	0.3220	0	6.3595	0.3362	0	4.3679	0.3027	0	3.3127	0.2586	0	4.6747	0.2946	0	4.5228	-0.0945	1
79	3.7922	0.2315	0	5.1090	0.3803	0	6.4520	0.4760	0	5.7091	0.3397	0	6.5175	0.3137	0	4.6064	0.4092	0	3.4933	0.3098	0	5.0275	0.3529	0	4.9417	-0.0125	1
83	4.4613	0.3334	0	5.9857	0.4174	0	7.6583	0.5039	0	6.7247	0.3379	0	6.6271	0.2436	0	5.6131	0.4988	0	4.1458	0.3883	0	6.1867	0.2771	0	7.3121	-0.0343	1
89	3.3419	0.2215	0	4.3346	0.2995	0	5.5666	0.2744	0	4.8763	0.2055	0	3.4448	-0.0029	1	4.5050	0.3767	0	3.2734	0.2785	0	5.2542	0.1695	0	11.1593	0.1917	0
90	3.0166	0.1994	0	3.8785	0.2400	0	5.0126	0.2081	0	4.3462	0.1607	0	2.9596	-0.0044	1	4.1780	0.3092	0	2.9890	0.2129	0	4.9491	0.1366	0	11.0033	0.1582	0
91	2.7074	0.1729	0	3.4274	0.2243	0	4.3573	0.1942	0	3.7929	0.1780	0	2.3229	-0.0517	1	3.8098	0.2965	0	2.6991	0.2141	0	4.4998	0.1203	0	11.6785	0.1490	0
93	2.0600	0.1206	0	2.4038	0.1287	0	2.9633	0.1280	0	2.5141	0.1001	0	1.6598	-0.0240	1	2.8808	0.2030	0	2.0125	0.1506	0	3.4585	0.0721	0	8.3450	0.1345	0
94	1.6045	0.0888	0	1.8104	0.0776	0	2.2519	0.0757	0	1.8733	0.0527	0	1.3483	-0.0252	1	2.2760	0.1381	0	1.5699	0.0976	0	2.7775	0.0496	0	6.5192	0.0965	0

LC model: Nonlinear variant

	Australia			United Kingdom			Italy			France			Spain			New Zealand			Sweden			Germany			Russia		
x	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$
63	1.2695	0.0503	0	1.7360	-0.0052	1	1.6612	0.0471	0	1.5021	0.0803	0	2.0136	0.1692	0	1.4916	0.0420	0	1.0001	0.0261	0	0.8158	-0.0071	1	3.4137	0.0808	0
64	1.3791	0.0545	0	1.8832	-0.0182	1	1.8260	0.0508	0	1.6402	0.0890	0	2.2261	0.1744	0	1.6240	0.0450	0	1.1102	0.0267	0	0.8634	0.0025	0	3.2759	0.1433	0
65	1.4949	0.0559	0	2.0537	-0.0133	1	2.0143	0.0753	0	1.8048	0.1044	0	2.4505	0.1777	0	1.7636	0.0643	0	1.2315	0.0505	0	0.9456	0.0114	0	3.1094	0.1537	0
66	1.6245	0.0554	0	2.2286	-0.0037	1	2.2267	0.0908	0	1.9833	0.1172	0	2.7054	0.2038	0	1.8948	0.0736	0	1.3728	0.0714	0	1.0541	0.0751	0	2.9430	0.1604	0
91	2.3049	0.1302	0	3.2832	0.1903	0	4.3565	0.1690	0	3.6473	0.1583	0	1.9845	-0.0045	1	3.6316	0.2798	0	2.5641	0.1658	0	1.8977	0.0736	0	10.3803	0.1823	0
93	1.7617	0.0904	0	2.2035	0.1038	0	2.8844	0.1119	0	2.4057	0.0867	0	1.3278	-0.0075	1	2.9968	0.2011	0	1.8977	0.1172	0	1.3124	0.0389	0	7.7978	0.1469	0
94	1.4047	0.0676	0	1.6078	0.0588	0	2.1721	0.0629	0	1.7553	0.0420	0	1.0020	-0.0190	1	2.4746	0.1397	0	1.5227	0.0749	0	1.0246	0.0259	0	6.0321	0.1054	0
63	1.2695	0.0503	0	1.7360	-0.0052	1	1.6612	0.0471	0	1.5021	0.0803	0	2.0136	0.1692	0	1.4916	0.0420	0	1.0001	0.0261	0	0.8158	-0.0071	1	3.4137	0.0808	0
76	3.3241	0.1787	0	4.8855	0.3000	0	5.6138	0.4152	0	4.8224	0.3075	0	6.3248	0.4009	0	3.9225	0.2331	0	3.3571	0.2020	0	2.9556	0.2970	0	2.7350	-0.0271	1
78	3.6884	0.1930	0	5.4783	0.3447	0	6.4470	0.4454	0	5.5334	0.3251	0	7.0404	0.4116	0	4.4158	0.2503	0	3.8241	0.2429	0	3.4005	0.3212	0	3.3081	-0.1337	1
79	3.8234	0.2251	0	5.7610	0.4048	0	6.8284	0.4707	0	5.8599	0.3392	0	7.2263	0.3956	0	4.6646	0.3493	0	4.0534	0.3223	0	3.6003	0.3748	0	3.8562	-0.0345	1
80	3.9549	0.2345	0	6	0.3577	0	7.1986	0.4754	0	6.1692	0.3513	0	7.4181	0.4106	0	4.7713	0.3364	0	4.2201	0.3082	0	3.7534	0.3606	0	4.2202	-0.0017	1
83	4.3148	0.2917	0	6.5218	0.4124	0	7.9736	0.4780	0	6.8316	0.3269	0	7.3116	0.3324	0	5.4876	0.4766	0	4.5474	0.3669	0	3.9718	0.2714	0	6.4005	-0.0374	1

CBD model: Nonlinear variant

	Australia			United Kingdom			Italy			France			Spain			New Zealand			Sweden			Germany			Russia		
x	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$	CapR(%)	ΔR	${}_j\mathcal{I}_p^x(N)$
70	3.1158	0.0927	0	3.7519	0.1045	0	4.1374	0.1496	0	3.5182	0.0263	0	4.6068	0.1828	0	3.2907	0.1245	0	2.1797	-0.0052	1	1.7275	0.0502	0	7.9876	0.0733	0
82	11.7376	0.5207	0	12.3340	0.3435	0	15.3102	0.7242	0	12.1693	0.5594	0	15.0026	0.7515	0	11.7768	0.6471	0	8.0685	0.2912	0	5.5538	-0.0247	1	14.4498	0.5227	0