

Costabile, Massimo; Viviano, Fabio

## Article

# Modeling the future value distribution of a life insurance portfolio

Risks

## Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

*Suggested Citation:* Costabile, Massimo; Viviano, Fabio (2021) : Modeling the future value distribution of a life insurance portfolio, *Risks*, ISSN 2227-9091, MDPI, Basel, Vol. 9, Iss. 10, pp. 1-17, <https://doi.org/10.3390/risks9100177>

This Version is available at:

<https://hdl.handle.net/10419/258261>

## Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

## Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

## Article

# Modeling the Future Value Distribution of a Life Insurance Portfolio

Massimo Costabile <sup>1,\*</sup>  and Fabio Viviano <sup>2,3,†</sup> 

<sup>1</sup> Department of Economics, Statistics and Finance, University of Calabria, Ponte Bucci Cubo 0 C, 87036 Rende, CS, Italy

<sup>2</sup> Department of Economics and Statistics, University of Udine, Via Tomadini 30/A, 33100 Udine, Italy; viviano.fabio@spes.uniud.it

<sup>3</sup> Department of Economics, Business, Mathematics and Statistics “B. de Finetti”, University of Trieste, Piazzale Europa 1, 34127 Trieste, Italy

\* Correspondence: massimo.costabile@unical.it

† These authors contributed equally to this work.

**Abstract:** This paper addresses the problem of approximating the future value distribution of a large and heterogeneous life insurance portfolio which would play a relevant role, for instance, for solvency capital requirement valuations. Based on a metamodel, we first select a subset of *representative policies* in the portfolio. Then, by using Monte Carlo simulations, we obtain a rough estimate of the policies' values at the chosen future date and finally we approximate the distribution of a single policy and of the entire portfolio by means of two different approaches, the ordinary least-squares method and a regression method based on the class of generalized beta distribution of the second kind. Extensive numerical experiments are provided to assess the performance of the proposed models.

**Keywords:** GB2; LSMC; metamodel; regression models; Solvency II

**JEL Classification:** G22



**Citation:** Costabile, Massimo, and Fabio Viviano. 2021. Modelling the Future Value Distribution of a Life Insurance Portfolio. *Risks* 9: 177. <https://doi.org/10.3390/risks9100177>

Academic Editor: Mercedes Ayuso

Received: 19 August 2021

Accepted: 27 September 2021

Published: 2 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In many relevant situations, life insurers face the necessity to determine the distribution of the value of their portfolio of policies at a certain future date. This happens, for example, when regulators need to maintain solvency capital requirements in order to continue to conduct business, as stated in the Solvency II directive or in the Swiss Solvency Test. In particular, Article 101(3) of the European directive requires that the Solvency Capital Requirement “shall 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” (see [European Parliament and European Council 2009](#)). As a consequence, insurers are obliged to assess the value of assets and liabilities at a future date, the so-called risk horizon, in order to derive their full loss distributions. To achieve this, the relevant risk factors must be projected at the risk horizon and then, conditional on the realized values, a market consistent valuation of the insurer's assets and liabilities is required. This has led insurance and reinsurance companies to face a computationally intensive problem. Indeed, due to the complex structure of the insurer's liabilities, in general, closed form formulas are not available and a straightforward approach, common among insurers, is to obtain an estimate through nested Monte Carlo simulations. Unfortunately, this approach is extremely time consuming and becomes readily unmanageable from a computational point of view. In this regard, one possible alternative method proposed in the literature to reduce the computational effort and to preserve the accuracy of the desired estimates is the Least-Squares Monte Carlo (LSMC) method, firstly introduced by [Carrière \(1996\)](#), [Tilley \(1993\)](#), and [Longstaff and Schwartz \(2001\)](#) in the context of American-type Option Pricing. Application of the LSMC method for valuing solvency capital requirements in the

insurance business was proposed in [Cathcart and Morrison \(2009\)](#) and [Bauer et al. \(2010\)](#). Moreover, [Floryszczak et al. \(2016\)](#) and [Krah et al. \(2018\)](#) illustrate a practical implementation of the LSMC in this particular context. The above-mentioned papers, proposed in the actuarial literature, share the common feature of evaluating capital requirements for a single policy.

In the case of an entire portfolio of policies, the nested simulation approach is even more difficult to implement due to the huge computational effort needed. For instance, assuming 10,000 outer trajectories simulated from the current time to the risk horizon for each one of the  $v$  risk factors, and then 2500 inner paths for each outer, with a monthly discretization for 20 years, and considering an insurance portfolio composed of 10,000 contracts, the total number of cash-flow projections needed would be  $10,000 \times v \times 2500 \times 12 \times 20 \times 10,000 = v \times 6 \times 10^{13}$ , which is very hard to manage.

In order to keep the computational complexity of the evaluation problem at a reasonable level, we propose a metamodeling approach. Metamodeling, introduced in system engineering (see [Barton 2015](#)), can be defined as “the practice of using a model to describe another model as an instance” (see [Allemang and Hendler 2011](#)). This approach has also been widely used in the actuarial literature to estimate the price and Greeks of large portfolios of life insurance policies. For instance, [Gan \(2013\)](#) developed a metamodel based on data clustering and machine learning to price large portfolios of variable annuities, while [Gan and Lin \(2015\)](#) tackled a similar problem by developing a functional data approach. In addition, [Gan \(2015\)](#) compares the data clustering approach and Latin hypercube sampling to select representative variable annuities. Finally, [Gan and Valdez \(2018\)](#) proposes a metamodel to estimate partial Greeks of variable annuities with dependence.

In the present paper, the metamodel we propose to approximate the future value distribution of a life insurance portfolio is constructed in different steps:

1. Select a subset of representative policies by means of conditional Latin hypercube sampling;
2. Project the risk factors from the evaluation date to the risk horizon by means of outer simulations;
3. Compute a rough estimate of each representative policy by means of a very limited (say two) number of inner simulations;
4. Create a regression model to approximate the distribution of the value of representative policies;
5. Use the regression model to estimate the future value distribution of the entire portfolio.

We propose two different approaches to develop the regression model in steps 4 and 5. The first approach relies upon the well-established Ordinary Least Squares (OLS) method for approximating the conditional distribution of each representative policy at the risk horizon, and then a second OLS regression is applied to estimate the future value distribution of the entire portfolio. Roughly speaking, we may say that the LSMC method is applied to estimate the distribution of the value of each representative policy at the risk horizon, and then this information is extended to the entire portfolio by means of a simple OLS regression. We call this approach the LSMC method.

The second approach exploits the class of generalized beta of the second kind (GB2) distributions to model the conditional distribution of each representative policy value at the risk horizon and also to estimate the future value distribution of the entire portfolio. We underline that the GB2 regression model has been used in [Gan and Valdez \(2018\)](#) for modeling the fair current market values of guarantees embedded in a large variable annuity portfolio starting from a set of representative policies. Extensive numerical experiments have been conducted in order to assess the performance of the proposed models. The remainder of the paper is structured as follows. Section 2 provides the evaluation framework and Section 3 introduces the metamodeling approach. Section 4 illustrates some numerical results, and finally, in Section 5, conclusions are drawn.

## 2. The Evaluation Framework

We consider a life insurance portfolio with  $M$  contracts underwritten by different policyholders (males and females) of different ages at the inception date  $t = 0$ . We take into account different types of life insurance policies which differ from each other in terms of maturity, policyholders' ages, and sex. In particular, we consider unit-linked products, term life insurance and immediate life annuities. We assume that the unit-linked product pays, upon reaching maturity, and assuming the survival of the insured, the maximum value between the minimum guaranteed benefit and the value of a specific reference asset. The immediate life annuity is assumed to pay 10% of the level of a given reference asset continuously whilst the insured is alive; finally, the term insurance contract pays the total value of the asset upon the death of the policyholder before maturity. Regarding all the possible policy configurations, see Table 1.

**Table 1.** This table shows the parameters used to generate the life insurance portfolio.

Feature	Value
Policyholder age	{55, . . . , 65}
Sex	{Male, Female}
Maturity	{10, 15, 20, 25, 30}
Product type	{Unit-linked, Term Insurance, Life Annuity}

Since our task is to approximate the portfolio value distribution at the risk horizon starting from a set of representative policies, we use the Conditional Latin Hypercube Sampling (CLHS) method (see Minasny and McBratney 2006). Indeed, this approach has already been applied to select subsets of representative policies providing reliable results, e.g., see Gan and Valdez (2018). Therefore, in order to select a set of  $s$  representative contracts, we apply the CLHS method to the design matrix  $\mathbf{X}$ , which contains all the features characterizing each specific policy, i.e., types, maturity, sex and age of the policyholder. Note that the categorical variables are treated as dummy variables.

In order to project the cash-flows generated by the contracts over time, we need to simulate the possible evolution of the risk factors. In this regard, we consider a computational framework where mortality, interest rate and the reference asset are taken into account. Despite insurance companies being exposed to systematic and non-systematic mortality risks, in our setting we consider only the first component for computational purposes due to the big dimension of the portfolio that will be considered.

Let  $(\Omega, \mathbb{F}, \mathbb{P})$  be a filtered probability space large enough to support a process  $X$  in  $\mathbb{R}^k$ , representing the evolution of financial variables, and a process  $Y$  in  $\mathbb{R}^d$ , representing the evolution of mortality. The filtration  $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$  represents the flow of information available as time passes by; this includes knowledge of the evolution of all state variables up to each time  $t$  and of whether the policyholder has died by then. Specifically, we define  $\mathcal{F}_t$  as the  $\sigma$ -algebra generated by  $\mathcal{G}_t \cup \mathcal{H}_t$ , where

$$\mathcal{G}_t = \sigma(Z_s : 0 \leq s \leq t), \quad \mathcal{H}_t = \sigma(\mathbb{I}_{\{\zeta \leq s\}} : 0 \leq s \leq t),$$

and where  $Z = (X, Y)$  is the joint state variables process in  $\mathbb{R}^{k+d}$ . Thus, we have  $\mathbb{F} = \mathbb{G} \vee \mathbb{H}$ , with  $\mathbb{G} = \mathbb{G}^X \vee \mathbb{G}^Y$  and with  $\mathbb{H} = (\mathcal{H}_t)_{t \geq 0}$  being the smallest filtration with respect to which  $\zeta$  is a stopping time and interpreted as the remaining lifetime of an insured. For more detail of modeling mortality under the intensity-based framework, see Biffis (2005).

Under the physical probability measure,  $\mathbb{P}$ , we assume that the financial risk factors (reference asset value  $S$ , and interest rate  $r$ ) dynamics are described by the following stochastic differential equations

$$\begin{aligned} dS(t) &= S(t)(r(t) + \lambda)dt + S(t)\sigma_S dW^{1,\mathbb{P}}(t), \\ S(0) &= S_0, \end{aligned} \quad (1)$$

where  $\lambda$  is the risk premium,  $\sigma_S$  is a positive constant,  $W^{1,\mathbb{P}}(t)$  is a standard Wiener process, and  $r(t)$  is the risk-free interest rate, which is assumed to follow the dynamics

$$\begin{aligned} dr(t) &= \alpha(\theta - r(t))dt + \sigma_r dW^{2,\mathbb{P}}(t), \\ r(0) &= r_0. \end{aligned} \quad (2)$$

Here,  $W^{2,\mathbb{P}}(t)$  is a standard Wiener process, and the coefficients  $\alpha, \theta, \sigma_r$  are positive constants representing the speed of mean reversion, the long-term interest rate, and the interest rate volatility, respectively. Further, we assume that the two Wiener processes,  $W^{1,\mathbb{P}}(t)$  and  $W^{2,\mathbb{P}}(t)$ , are correlated with the correlation coefficient  $\rho$ .

In the absence of arbitrage opportunities, an equivalent martingale measure  $\mathbb{Q}$  exists, under which all financial security prices are martingales after deflation by the money market account. We refer the readers to Biffis (2005) for more detail. Under the risk-neutral probability measure,  $\mathbb{Q}$ , the dynamics in Equations (1) and (2) can be re-written as

$$dS(t) = S(t)r(t)dt + S(t)\sigma_S dW^{1,\mathbb{Q}}(t),$$

and

$$dr(t) = \alpha\left(\theta - \frac{\sigma_r}{\alpha}\gamma - r(t)\right)dt + \sigma_r dW^{2,\mathbb{Q}}(t),$$

where  $\gamma$  is the market price of risk. Note that  $W^{1,\mathbb{Q}}(t)$  and  $W^{2,\mathbb{Q}}(t)$  are two correlated standard Wiener processes with the coefficient of correlation  $\rho$  under  $\mathbb{Q}$ .

Concerning mortality, following Fung et al. (2014), we assume that the force of mortality,  $\mu_{x+t}(t)$ , under the physical probability measure  $\mathbb{P}$  for an individual aged  $x$  at time  $t = 0$ , evolves accordingly to the following one-factor, non-mean-reverting and time-homogeneous affine process:

$$\begin{aligned} d\mu_{x+t}(t) &= [a + b\mu_{x+t}(t)]dt + \sigma_\mu \sqrt{\mu_{x+t}(t)} dW^{3,\mathbb{P}}(t), \\ \mu_x(0) &> 0, \end{aligned} \quad (3)$$

where  $a \neq 0, b > 0, \sigma_\mu > 0$  represent the volatility of the mortality intensity and  $W^{3,\mathbb{P}}(t)$  is a standard Wiener process which is assumed to be independent with respect to  $W^{1,\mathbb{P}}(t)$  and  $W^{2,\mathbb{P}}(t)$ . As pointed out by Fung et al. (2014), the important advantages of the mortality model defined in Equation (3) are its tractability since analytical expressions are available to evaluate survival probabilities, and also its simplicity since the model dynamics can be easily simulated. Furthermore, this model guarantees that, under specific conditions, the force of mortality is strictly positive (i.e., if  $a \geq \sigma_\mu^2/2$ ).

The dynamics in Equation (3) under  $\mathbb{Q}$  can be defined as

$$\begin{aligned} d\mu_{x+t}(t) &= [a + (b - \delta\sigma_\mu)\mu_{x+t}(t)]dt + \sigma_\mu \sqrt{\mu_{x+t}(t)} dW^{3,\mathbb{Q}}(t), \\ \mu_x(0) &> 0, \end{aligned}$$

where  $W^{3,\mathbb{Q}}(t)$  is a standard Wiener process under the risk-neutral measure and  $\delta$  is the market price of the systematic mortality risk.

Note that the parameters in the stochastic mortality model are estimated by calibrating the implied survival curve to the one obtained from the Italian population data of year

2016 (assumed to be  $t = 0$ ) collected from the Human Mortality Database (see [Fung et al. 2014](#)). The calibration procedure was conducted for all policyholder ages and genders reported in Table 1.

Finally, it is worth noting that, due to the flexibility of the methodology that will be proposed, different and/or more complex dynamics to describe the evolution of the risk factors may be assumed with respect to the ones assumed above.

### 3. Problem and Methodology

Under the framework defined in Section 2, we need to evaluate the streams of payments embedded in each policy inside the insurance portfolio. Before discussing the methodology, let us recall some results provided by [Biffis \(2005\)](#) related to the time- $\tau$  fair values of the most common payoffs embedded in typical life insurance products, i.e., survival and death benefits.

**Proposition 1.** (Survival benefit.) Let  $C$  be a bounded  $\mathbb{G}$ -adapted process. Then, the time- $\tau$  fair value  $SB_\tau(C_T; T)$  of the time- $T$  survival benefit of amount  $C_T$ , with  $0 \leq \tau \leq T$ , is given by:

$$SB_\tau(C_T; T) = \mathbb{E} \left[ e^{-\int_\tau^T r_s ds} \mathbb{I}_{\{\zeta > T\}} C_T \mid \mathcal{F}_\tau \right] = \mathbb{I}_{\{\zeta > \tau\}} \mathbb{E} \left[ e^{-\int_\tau^T (r_s + \mu_s) ds} C_T \mid \mathcal{G}_\tau \right].$$

In particular, if  $C$  is  $\mathbb{G}^X$ -adapted, the following holds:

$$SB_\tau(C_T; T) = \mathbb{I}_{\{\zeta > \tau\}} \mathbb{E} \left[ e^{-\int_\tau^T r_s ds} C_T \mid \mathcal{G}_\tau^X \right] \mathbb{E} \left[ e^{-\int_\tau^T \mu_s ds} \mid \mathcal{G}_\tau^Y \right].$$

**Proposition 2.** (Death benefit.) Let  $C$  be a bounded  $\mathbb{G}$ -predictable process. Then, the time- $t$  fair value  $DB_\tau(C_\zeta; T)$  of the death benefit of amount  $C_\zeta$ , payable in case the insured dies before time  $T$ , with  $0 \leq \tau \leq T$ , is given by

$$DB_\tau(C_\zeta; T) = \mathbb{E} \left[ e^{-\int_\tau^\zeta r_s ds} C_\zeta \mathbb{I}_{\{\tau < \zeta \leq T\}} \mid \mathcal{F}_\tau \right] = \mathbb{I}_{\{\zeta > \tau\}} \int_\tau^T \mathbb{E} \left[ e^{-\int_\tau^u (r_s + \mu_s) ds} \mu_u C_u \mid \mathcal{G}_\tau \right] du.$$

In particular, if  $C$  is  $\mathbb{G}^X$ -predictable, the following holds

$$DB_\tau(C_\zeta; T) = \mathbb{I}_{\{\zeta > \tau\}} \int_\tau^T \mathbb{E} \left[ e^{-\int_\tau^u r_s ds} C_u \mid \mathcal{G}_\tau^X \right] \mathbb{E} \left[ e^{-\int_\tau^u \mu_s ds} \mu_u \mid \mathcal{G}_\tau^Y \right] du.$$

We refer the readers to [Biffis \(2005\)](#) for the corresponding proofs and further details. Therefore, as we can see from Propositions 1 and 2, evaluating life insurance policies at future times implies solving conditional expectations for which often analytical formulas do not exist. Due to this, simulation-based approaches are extensively used (see [Boyer and Stentoft 2013](#)), among which we mention the nested simulations method where a high number of inner simulations branch out from another huge set of outer scenarios. However, the simulations within simulations approach is computationally challenging, especially when several policies are considered, as in our case. Therefore, in the following, we are going to discuss two methodologies to evaluate the streams of payments embedded in each policy inside the insurance portfolio. For this purpose, we project the relevant risk factors affecting the policy (i.e.,  $S$ ,  $r$ , and  $\mu$ ) under the physical probability measure from time  $t = 0$  up to the risk horizon  $\tau$ , and then for each outer scenario another set of inner trajectories is simulated under the risk-neutral measure.

In order to avoid the huge computational cost of a pure nested model, as in the LSMC approach, we simulate  $n$  possible outer trajectories of the risk factors and then for each of them we further simulate  $\bar{n} \ll n$  inner paths. Following this approach, let  $\mathbf{Z}^i$  be an  $n \times v$  matrix, where the row vector  $\mathbf{z}_k^i$  contains the  $k$ th outer scenario of the  $v$  risk factors affecting the value of the  $i$ th representative policy. For each vector  $\mathbf{z}_k^i$  and for time  $\tau < t \leq T$ , we simulate  $\bar{n}$  trajectories under the risk-neutral probability measure. To simplify the notation,



we focus on the  $i$ th representative policy, and we denote  $\mathbf{z}_{j,t}^k$  the vector containing the time- $t$  values of the risk factors along the  $j$ th inner trajectory corresponding to the  $k$ th outer scenario. Moreover, we label  $\mathbf{Y}$  a  $n \times s$  matrix where the element  $y_{ik}$  represents the value of the  $i$ th policy corresponding to the  $k$ th outer scenario obtained by averaging across the few inner simulations. Formally,

$$y_{ik} = \frac{1}{\bar{n}} \sum_{j=1}^{\bar{n}} \sum_{\tau < t \leq T_i} \Phi_t^i(\mathbf{z}_{j,t}^k) \quad i = 1, \dots, s, \text{ and } k = 1, \dots, n, \quad (4)$$

where  $\Phi_t^i(\cdot)$ s represent the discounted cash-flows at time  $t$  of the  $i$ th policy with maturity  $T_i$ .

In this way, we obtain a first (rough) estimate of each representative policy value distribution at the future time  $\tau$ . The next step is to obtain a more accurate estimate of the distribution of the time- $\tau$  value of each representative policy and then to infer the distribution of the time- $\tau$  value of the entire portfolio. We achieve this by applying two different approaches, an OLS as in the least-squares Monte Carlo method and a GB2 model.

### 3.1. The LSMC Method

The least-squares Monte Carlo method applied to the problem of computing the distribution of the insurer's liabilities at a certain future date is based on the idea that the bias deriving from the few inner simulations can be reduced by approximating the involved conditional expectations with a linear combination of basis functions depending on some covariates, whose coefficients are estimated through an ordinary least-squares procedure (see [Bauer et al. 2010](#) for further details).

A straightforward application of the LSMC approach would be to apply the method on each policy inside the insurance portfolio. However, this kind of strategy would be quite computationally expensive due to the big dimensions of an insurance portfolio. Due to this, we propose applying the LSMC method first on just a set of representative policies and then through an OLS regression extend it to the entire portfolio.

Hence, according to the LSMC method, we assume that the conditional  $i$ th representative policy value,  $\hat{y}_{ik}$ , can be expressed as a linear combination of basis functions depending on the covariate matrix  $\mathbf{z}_k^i$  as follows:

$$\hat{y}_{ik} = \sum_{j=1}^L \hat{\beta}_j^i e_j(\mathbf{z}_k^i) \quad i = 1, \dots, s \text{ and } k = 1, \dots, n, \quad (5)$$

where  $e_j(\cdot)$  is the  $j$ th basis function in the regression,  $L$  is the number of basis functions, and  $\hat{\beta}_j^i$ s represent the coefficients estimated through

$$(\hat{\beta}_1^i, \dots, \hat{\beta}_L^i) = \underset{\beta_1, \dots, \beta_L}{\operatorname{argmin}} \left[ \sum_{k=1}^n \left( y_{ik} - \sum_{j=1}^L \beta_j^i e_j(\mathbf{z}_k^i) \right)^2 \right].$$

In this way, we obtain an  $n \times s$  matrix  $\hat{\mathbf{Y}}$  where each row vector  $\hat{\mathbf{y}}_k$  contains the values of each representative policy corresponding to the  $k$ th outer scenario.

Now, in order to approximate the distribution of the value of the entire portfolio, we construct an OLS regression model for each outer scenario. In this regard, we denote with  $\mathbf{X}$  an  $M \times (w + 1)$  matrix, where the row vector  $\mathbf{x}_i$  contains the  $w$  covariates (*gender, product type, age, and maturity*) characterizing the  $i$ th contract in the portfolio plus an intercept term ( $M$  is the total number of contracts inside the insurance portfolio). Moreover, let  $\tilde{\mathbf{X}}$  be the  $s \times (w + 1)$  matrix describing the structure of the representative insurance portfolio. Hence,  $\tilde{\mathbf{x}}_i$  contains the  $w$  covariates characterizing the  $i$ th representative contract plus an intercept term.

Therefore, we regress each row vector  $\hat{\mathbf{y}}_k$  ( $k = 1, \dots, n$ ) on the covariate matrix  $\tilde{\mathbf{X}}$ , and once the coefficients are estimated, we extend them to the remaining policies by

exploiting the matrix  $\mathbf{X}$ . In this way, we obtain the value of the  $i$ th contract corresponding to the  $k$ th outer scenario, which is denoted by  $\hat{v}_{ik}$ . Formally,

$$\hat{v}_{ik} = \mathbf{x}_i \hat{\boldsymbol{\beta}}'_k \quad i = 1, \dots, M \text{ and } k = 1, \dots, n, \quad (6)$$

where

$$\hat{\boldsymbol{\beta}}'_k = (\bar{\mathbf{X}}' \bar{\mathbf{X}})^{-1} \bar{\mathbf{X}}' \hat{\mathbf{y}}'_k.$$

Finally, the entire portfolio value distribution is obtained by adding up all the policy values in Equation (6) corresponding to each outer scenario.

### 3.2. The GB2 Model

A GB2 model appears to provide a flexible family of distributions as it nests a range of standard distributions as special or limiting cases, such as the log-normal, the generalized-gamma, the Burr type III, the Burr type XII and many other (see McDonald 1984). Moreover, it has been used in several actuarial applications (e.g., see Gan and Valdez 2018) to model the fair market value of a portfolio made up of life insurance policies. A GB2 random variable can be constructed from a transformed ratio of two gamma random variables. The density function of a GB2 random variable,  $Y$ , is given by

$$f(y) = \frac{|a|}{bB(p, q)} \left(\frac{y}{b}\right)^{ap-1} \left[1 + \left(\frac{y}{b}\right)^a\right]^{-p-q}, \quad y > 0, \quad (7)$$

where  $a \neq 0, p > 0, q > 0$  are shape parameters,  $b > 0$  is the scale parameter,  $B(\cdot)$  is the Beta function, and its expectation equals:

$$\mathbb{E}[Y] = b \cdot \frac{B\left(p + \frac{1}{a}, q - \frac{1}{a}\right)}{B(p, q)}, \quad (8)$$

which exists if  $-p < \frac{1}{a} < q$ .

In order to approximate the value of the portfolio, at first we approximate the time- $\tau$  value of each representative policy, and then we use this information to approximate the distribution of the value of the entire insurance portfolio at the risk horizon. To achieve this, we construct two different GB2 regression models which exploit the generated information at the risk horizon (i.e.  $S(\tau)$ ,  $r(\tau)$ , and  $\mu(\tau)$ ), and then the features characterizing uniquely each policy, respectively.

Specifically, since the policy values  $y_{ik}$  obtained from Equation (4) are not accurate due to the few inner trajectories on which they are based on, we aim at reducing the bias by estimating the involved conditional expectation through a GB2 regression model. In this regard, we assume that the  $i$ th policy value at time  $\tau$  conditioned on a specific outer scenario is a GB2 random variable with parameters  $(a_i, p_i, q_i, b_i)$ . In particular, we make the  $b$ -parameter depend on some independent covariates (i.e., the value at time  $\tau$  of the risk factors which affect the policy of interest). Note that several approaches to incorporate covariates in the GB2 regression model exist as well as different re-parametrization (see Beirlant et al. 2004; Frees and Valdez 2008). However, as noticed by Sun et al. (2008) and Frees et al. (2016), incorporating them into the scale parameter,  $b$ , facilitates the interpretability of the model; indeed, as can be seen in Equation (8), the expectation will change proportionally with respect to  $b$ , allowing one to interpret the regression coefficients as proportional changes.

Hence,  $b(\mathbf{Z}^i) = \exp(\mathbf{Z}^i \boldsymbol{\beta}'_i)$ , where  $\boldsymbol{\beta}_i = (\beta_{i;0}, \beta_{i;1}, \dots, \beta_{i;v})$  are the corresponding coefficients attached to each risk-factor. Note that the matrix  $\mathbf{Z}^i$  now includes an intercept term.

We can use the maximum likelihood method to estimate the parameters. Since we incorporate covariates through the scale parameter, we can write the log-likelihood function of the model as



$$l(a_i, p_i, q_i, \beta_i) = n \ln \frac{|a_i|}{B(p_i, q_i)} - a_i p_i \sum_{k=1}^n \mathbf{z}_k^i \beta_i' + (a_i p_i - 1) \sum_{k=1}^n \ln(y_{ik}) - (p_i + q_i) \sum_{k=1}^n \ln \left[ 1 + \left( \frac{y_{ik}}{\exp(\mathbf{z}_k^i \beta_i')} \right)^{a_i} \right], \quad (9)$$

where  $i = 1, \dots, s$ ,  $n$  is the number of the generated outer scenarios and  $y_{ik}$  denotes the value of the  $i$ th policy corresponding to the  $k$ th outer scenario.

Once we estimate the parameters for the GB2 model, we use the expectation for predicting the value of the policy at time  $\tau$ . Since we incorporate covariates through the scale parameter, we can estimate it as

$$\hat{y}_{ik} = \frac{\exp(\mathbf{z}_k^i \hat{\beta}_i') B(\hat{p}_i + \frac{1}{\hat{a}_i}, \hat{q}_i - \frac{1}{\hat{a}_i})}{B(\hat{p}_i, \hat{q}_i)}, \quad i = 1, 2, \dots, s \text{ and } k = 1, \dots, n, \quad (10)$$

where  $\mathbf{z}_k^i$  is the vector containing the  $k$ th outer scenario of the risk factors affecting the  $i$ th representative policy.

Once we obtain an estimate of the distribution of each representative policy at time  $\tau$ , we extend this information to the remaining policies. As already carried out for the OLS model, we are going to exploit both the matrices  $\bar{\mathbf{X}}$  and  $\mathbf{X}$  on which we now construct a new GB2 regression model.

Therefore, let  $\hat{\mathbf{Y}}$  be the  $n \times s$  matrix whose elements  $\hat{y}_{ik}$  denote the value of the  $i$ th representative policy corresponding to the  $k$ th outer scenario obtained through Equation (10).

Now, we construct a GB2 regression model in order to infer, starting from the set of representative policies, the distribution of the entire portfolio. Hence, recalling the pdf defined in Equation (7), we define the following log-likelihood function as:

$$l(a_k, p_k, q_k, \beta_k) = s \ln \frac{|a_k|}{B(p_k, q_k)} - a_k p_k \sum_{i=1}^s \bar{\mathbf{x}}_i \beta_k' + (a_k p_k - 1) \sum_{i=1}^s \ln(\hat{y}_{ik}) - (p_k + q_k) \sum_{i=1}^s \ln \left[ 1 + \left( \frac{\hat{y}_{ik}}{\exp(\bar{\mathbf{x}}_i \beta_k')} \right)^{a_k} \right], \quad (11)$$

where  $s$  is the number of the representative policies and  $\bar{\mathbf{x}}_i$  is the row vector containing the information of the  $i$ th representative contract.

Once again, after we estimate the parameters through the maximum likelihood approach, we can then derive the distribution at the risk horizon for all the policies inside the insurance portfolio as

$$\hat{v}_{ik} = \frac{\exp(\mathbf{x}_i \hat{\beta}_k') B(\hat{p}_k + \frac{1}{\hat{a}_k}, \hat{q}_k - \frac{1}{\hat{a}_k})}{B(\hat{p}_k, \hat{q}_k)}, \quad i = 1, 2, \dots, M \text{ and } k = 1, \dots, n, \quad (12)$$

where  $\hat{v}_{ik}$  is the value of the  $i$ th contract corresponding to the  $k$ th outer scenario.

Finally, the entire portfolio value distribution is again obtained by adding up all the policy values corresponding to each outer scenario.

Note that the log-likelihood functions in Equations (9) and (11) may have multiple local maxima and since an analytic solution does not exist, we need to rely on a numerical procedure to estimate the involved parameters. We adopt the same multistage optimization algorithm described in [Gan and Valdez \(2018\)](#).

#### 4. Numerical Results

In this section, we present some numerical results obtained by exploiting the previously defined models. In particular, we consider a life insurance portfolio with  $M = 10,000$  contracts, and we focus on approximating its value distribution at the future time  $\tau = 1$  year. These policies can be of three different types: a unit-linked pure endowment contract with a minimum maturity guarantee  $G = 100$  payable upon the survival of the policyholder at the maturity date  $T$ , term life insurance policy which pays the value of a reference asset in case of death before maturity  $T$ , and an immediate life annuity contract

with continuous survival benefits equal to the 10% of a reference asset value up to the entire life of the insured person. We consider different policyholders, both males and females, with different ages  $x$  at time  $t = 0$ , which is also assumed to be the inception time of each policy. These characteristics are reported in Table 1. We assume that the insurance benefits depend upon a reference asset with the initial value  $S_0$ .

In Tables A1 and A2 given in Appendix A, we report the values of the involved parameters in Equations (1)–(3). In particular, concerning mortality, we have calibrated the survival curve implied by Equation (3) on the Italian males and females mortality data in the year 2016 obtained from the Human Mortality Database for each age  $x \in \{55, \dots, 65\}$ , and we assumed a longevity risk premium  $\delta = 0$ .

We conduct this numerical experiment by varying both the number of outer simulations,  $n$ , and the number of representative policies,  $s$ . In particular, we adopt a monthly Euler's discretization setting in order to project  $n \in \{1000, 5000, 10000\}$  outer trajectories of each risk factor under the  $\mathbb{P}$ -measure, and then for each outer scenario we further simulate  $\bar{n} = 2$  inner trajectories under the risk-neutral probability measure. With this simulation set, we are able to obtain a first rough estimate of  $Y$  on which we construct the LSMC and GB2 models discussed in Sections 3.1 and 3.2, respectively. Note that, concerning the LSMC method, we exploit as basis functions Hermite polynomials of orders 1 and 2, which are denoted, respectively, as LSMC\_1 and LSMC\_2 hereafter.

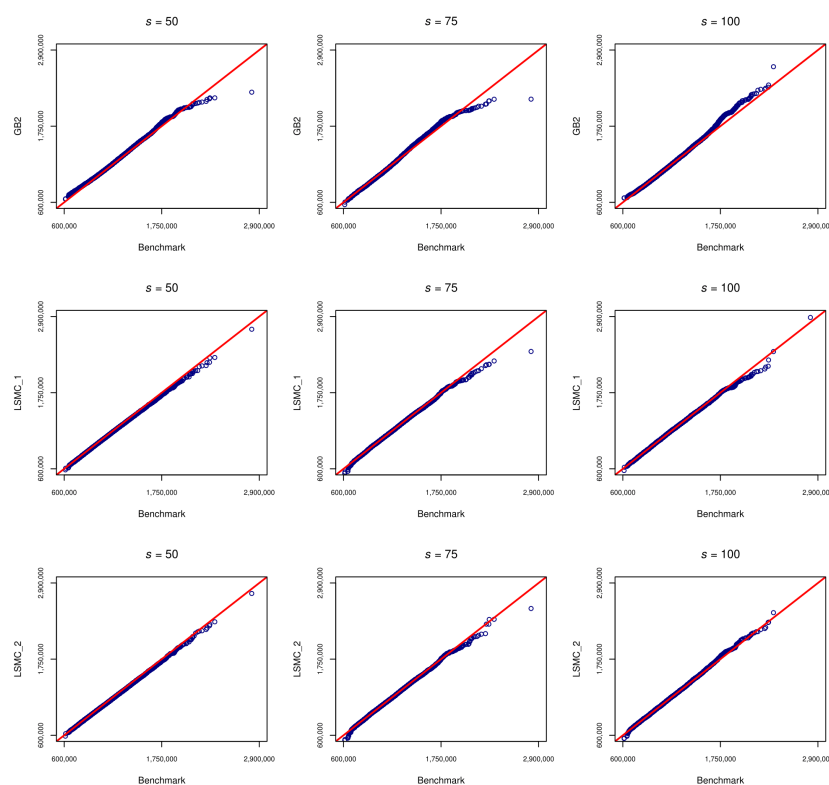
To determine the number of representative contracts  $s$ , we start from the informal rule proposed by Loeppky et al. (2009), which provide reasons and evidence supporting that the sample size should be about 10 times the input dimension. In our case, the dimension of covariates in the design matrix  $X$  is 5 (including the binary dummy variables converted from the categorical variables), and so we choose  $s = 50$  as the initial number of representative contracts. However, we investigate the models' performances by setting  $s = 75$  and  $s = 100$ .

Finally, the results are compared with a solid benchmark obtained through a nested simulations approach based on  $10,000 \times 2500$  simulations. This allows us to conclude on the reliability of the proposed methodologies and to compare them in terms of computational demand.

Figure 1 shows the Quantile-Quantile (Q-Q) plots of the portfolio value at time  $\tau = 1$  obtained by the nested simulations algorithm (assumed to be the theoretical one) and those predicted by the GB2 regression model and the LSMC models based on  $n = 10,000$  outer simulations and by varying the number of representative contracts  $s \in \{50, 75, 100\}$ . In this regard, we can see from Figure 1 that the proposed methodologies provide a good approximation except for the right tail of the distribution. In particular, concerning the GB2 regression model, we can see that the higher the number of representative contracts, the better the approximation.

For a comprehensive analysis, we perform multiple runs of each proposed method; in particular, the following analysis is based on 50 runs.

In Tables 2–4, we report the Mean Absolute Percentage Error (MAPE) relative to different quantities obtained by performing 50 runs of the proposed methodologies with a fixed number of outer scenarios ( $n = 10,000$ ) and by varying the number of representative contracts ( $s \in \{50, 75, 100\}$ ).



**Figure 1.** Q-Q plots relative to the future value distribution of the insurance portfolio. The theoretical distribution is assumed to be the one obtained by nested simulations based on  $10,000 \times 2500$  trajectories. The first row refers to the GB2 regression model based on 10,000 outer scenarios and by varying the number of representative contracts,  $s \in \{50, 75, 100\}$ . The second and third rows refer to the LSMC method with Hermite polynomials of orders 1 and 2 based on 10,000 outer scenarios and by varying the number of representative contracts,  $s \in \{50, 75, 100\}$ .

**Table 2.** This table reports the MAPE of the estimates obtained by running 50 times the GB2 and LSMC methods with  $n = 10,000$  and  $s = 50$ . The benchmark values are based on a nested simulations algorithm with  $10,000 \times 2500$  trajectories applied to the entire portfolio.

	5th Perc.	10th Perc.	Median	Mean	90th Perc.	95th Perc.	99th Perc.	99.5th Perc.
GB2	2.812%	2.180%	1.798%	2.594%	3.832%	4.016%	6.154%	4.375%
LSMC_1	3.238%	3.000%	2.399%	2.557%	2.398%	2.174%	2.436%	2.722%
LSMC_2	2.762%	2.754%	2.567%	2.557%	2.436%	2.114%	2.356%	2.841%

**Table 3.** This table reports the MAPE of the estimates obtained by running 50 times the GB2 and LSMC methods with  $n = 10,000$  and  $s = 75$ . The benchmark values are based on a nested simulations algorithm with  $10,000 \times 2500$  trajectories applied to the entire portfolio.

	5th Perc.	10th Perc.	Median	Mean	90th Perc.	95th Perc.	99th Perc.	99.5th Perc.
GB2	1.971%	1.782%	0.806%	0.542%	3.605%	3.949%	6.094%	3.867%
LSMC_1	2.500%	1.338%	1.530%	1.392%	1.251%	1.657%	0.941%	1.678%
LSMC_2	1.828%	1.047%	1.756%	1.392%	1.307%	1.485%	1.842%	2.142%

**Table 4.** This table reports the MAPE of the estimates obtained by running 50 times the GB2 and LSMC methods with  $n = 10,000$  and  $s = 100$ . The benchmark values are based on a nested simulations algorithm with  $10,000 \times 2500$  trajectories applied to the entire portfolio.

	5th Perc.	10th Perc.	Median	Mean	90th Perc.	95th Perc.	99th Perc.	99.5th Perc.
GB2	1.986%	1.745%	0.519%	0.347%	1.129%	1.313%	2.856%	1.944%
LSMC_1	1.629%	1.504%	0.440%	0.627%	0.764%	0.824%	0.958%	2.561%
LSMC_2	1.148%	1.145%	0.578%	0.627%	0.762%	0.986%	2.101%	2.334%

If we compare Tables 2–4, it is evident that increasing the number of representative contracts  $s$  leads to a better approximation of the mean and of the other considered measures of position. Moreover, it seems that the GB2 model, at least for a low number of representative contracts, is not able to adequately model the right tail of the distribution.

In Table 5, we report the Mean Percentage Error (MPE) and MAPE relative to the mean estimates obtained by running the GB2 and LSMC methods 50 times with different numbers of outer simulations,  $n$ , and representative contracts,  $s$ .

**Table 5.** This table reports the MPE and MAPE of the mean estimates obtained by running 50 times the GB2 and LSMC methods and varying the number of outer simulations (Outer) and that of representative contracts  $s$ . The benchmark value is based on a nested simulations algorithm with  $10,000 \times 2500$  trajectories applied to the entire portfolio.

		$s = 50$		$s = 75$		$s = 100$	
Outer	Method	MPE	MAPE	MPE	MAPE	MPE	MAPE
1000	GB2	3.612%	3.612%	0.163%	0.983%	−0.240%	0.923%
	LSMC_1	−3.475%	3.475%	−2.104%	2.221%	−1.017%	1.364%
	LSMC_2	−3.475%	3.475%	−2.104%	2.221%	−1.017%	1.364%
5000	GB2	2.981%	2.981%	0.715%	0.747%	−0.301%	0.474%
	LSMC_1	−2.840%	2.840%	−1.533%	1.533%	−1.029%	1.092%
	LSMC_2	−2.840%	2.840%	−1.533%	1.533%	−1.029%	1.092%
10,000	GB2	2.594%	2.594%	0.491%	0.542%	0.179%	0.347%
	LSMC_1	−2.557%	2.557%	−1.392%	1.392%	−0.490%	0.627%
	LSMC_2	−2.557%	2.557%	−1.392%	1.392%	−0.490%	0.627%

Looking at Table 5, we can see that for a fixed number of outer scenarios and for each applied method, the accuracy of the mean estimates increases with the number of representative contracts  $s$ . Moreover, it is evident that in most of the considered configurations, the GB2 model outperforms the LSMC methods. Furthermore, if we look at the last column of Table 5 ( $s = 100$ ), for instance, we can see that the higher the number of outer scenarios, the better the approximation. Finally, we can see that increasing the number of basis functions up to degree two in the LSMC method does not improve the accuracy of the mean estimates. This is probably due to the few outer simulated trajectories (at most 10,000 paths), which is not sufficient to appreciate the improvement which is usually expected. In the left-hand side of Figure A1 given in Appendix B, we report the corresponding box-plots from which it is possible to see that, in each of the considered configurations, the LSMC method systematically underestimates the quantity of interest.

Concerning the estimate of the 99.5th percentile of the distribution, which would be of interest for valuing solvency capital requirements, Table 6 reports the MPE and MAPE relative to 50 estimates obtained by varying both the number of simulations and the number of representative contracts.

**Table 6.** This table reports the MPE and MAPE of the 99.5th percentile estimates obtained by running the GB2 and LSMC methods 50 times and varying the number of outer simulations (Outer) and that of representative contracts  $s$ . The benchmark value is based on a nested simulations algorithm with  $10,000 \times 2500$  trajectories applied to the entire portfolio.

Outer	Method	$s = 50$		$s = 75$		$s = 100$	
		MPE	MAPE	MPE	MAPE	MPE	MAPE
1000	GB2	3.936%	6.570%	−1.512%	5.453%	1.410%	4.494%
	LSMC_1	−2.664%	3.715%	−6.308%	6.478%	−2.961%	4.253%
	LSMC_2	−0.252%	6.487%	−4.211%	7.150%	−1.438%	5.517%
5000	GB2	4.110%	4.723%	3.813%	4.018%	0.081%	2.653%
	LSMC	−2.908%	3.001%	−4.708%	4.722%	−1.659%	2.006%
	LSMC_2	−1.787%	3.484%	−3.118%	4.017%	−0.462%	3.110%
10,000	GB2	4.157%	4.375%	3.737%	3.867%	0.421%	1.944%
	LSMC_1	−2.643%	2.722%	−1.560%	1.678%	−2.522%	2.561%
	LSMC_2	−2.259%	2.841%	−0.131%	2.142%	−1.007%	2.334%

From Table 6, we can detect a similar behaviour as the one previously discussed. Specifically, we can see that, concerning the GB2 model, an increase in the number of representative contracts (for fixed  $n$ ) leads to an improvement of the resulting estimates. On the contrary, for the LSMC method, there is no clear pattern. Indeed, as we can see, increasing the number of representative contracts (for a fixed  $n$ ) does not lead to a clear improvement in the results. Moreover, increasing the number of basis functions as well as the number of outer simulations does not increase the accuracy of the estimates (see also the right side of Figure A1 in Appendix B). As in the case of the mean estimate, this could be due to the small number of outer simulations, and so we may conclude that passing from 1000 to 10000 trajectories is still not sufficient to exploit more basis functions. Once again, if we look at the case of  $n = 10,000$  and  $s = 100$ , the GB2 model outperforms the LSMC approach.

Now, let us examine the speed of the proposed algorithms with respect to the benchmark. Table 7 shows the runtime of GB2 and LSMC expressed as a percentage of the time required by the nested simulation method based on 10,000 outsiders and 2500 insiders. Note that we conducted all experiments using R on a computer equipped with an Intel® Core(TM) i7-1065G7 CPU 1.50 GHz processor with 12 GB of RAM and Windows 10 Home operating system.

**Table 7.** Percentage of the runtime required by the GB2 and LSMC methods with respect to the nested simulations approach. Note that the computational demand to construct the benchmark with a nested simulations approach based on  $10,000 \times 2500$  scenarios applied to the entire portfolio is about 187,200 s.

Method	$n = 1000$			$n = 5000$			$n = 10,000$		
	$s = 50$	$s = 75$	$s = 100$	$s = 50$	$s = 75$	$s = 100$	$s = 50$	$s = 75$	$s = 100$
GB2	0.069%	0.078%	0.098%	0.337%	0.380%	0.501%	0.660%	0.832%	1.021%
LSMC_1	0.005%	0.006%	0.007%	0.012%	0.018%	0.019%	0.036%	0.045%	0.047%
LSMC_2	0.005%	0.006%	0.007%	0.013%	0.019%	0.020%	0.037%	0.046%	0.047%

As we can see from Table 7, by applying the proposed methodologies, we have drastically reduced the computational time required instead by a nested simulations approach. Moreover, as expected, the LSMC method presented in Section 3.1 outperforms the GB2 model in terms of time in each of the proposed configurations. However, this is due to the existence of a closed form formula for the estimation of the involved parameters. Indeed, as stated in Section 3.2, the estimation procedure for the GB2 model is based on a multistage optimization algorithm due to the complexity of the likelihood functions, which

may have multiple local maxima. Regardless, if compared with the simulations within simulations method, the GB2 model proved to be an accurate and efficient alternative.

#### Full LSMC

To provide an exhaustive analysis, we consider a straightforward application of the LSMC method. Hence, we apply the LSMC method on each contract composing the insurance portfolio without considering any set of representative policies. The results are then compared with those already shown in the previous section both in terms of accuracy and computational demand. Just as an example, we construct the LSMC model by exploiting as set of basis functions Hermite polynomials with order 1 based on  $10,000 \times 2$  simulations (LSMC\_Full). Table 8 reports the MPE and MAPE relative to the 5th-percentile, the mean, and the 99.5th percentile estimates obtained by performing 50 runs of the proposed methods. Further, we report the results relative to the GB2 model (GB2) and LSMC method with Hermite polynomials of order 1 (LSMC\_1) and order 2 (LSMC\_2) based on  $10,000 \times 2$  simulations and  $s = 100$  representative policies.

**Table 8.** This table reports the MPE and MAPE relative to the 5th percentile, the mean, and the 99.5th percentile estimates obtained by applying different methodologies. GB2 stands for the GB2 regression model based on  $n = 10,000$  outer scenarios and  $s = 100$  representative policies; LSMC\_1 refers to the LSMC method based on  $n = 10,000$  outer scenarios and  $s = 100$  representative policies with Hermite polynomials of order 1; LSMC\_2 refers to the LSMC method based on  $n = 10,000$  outer scenarios and  $s = 100$  representative policies with Hermite polynomials of order 2; LSMC\_Full refers to the LSMC method based on  $n = 10,000$  outer scenarios and constructed on each contract in the insurance portfolio. The results are compared with the corresponding benchmark value based on nested simulations with  $10000 \times 2500$  trajectories applied to the entire portfolio.

Method	5th Perc.		Mean		99.5th Perc.	
	MPE	MAPE	MPE	MAPE	MPE	MAPE
GB2	−1.986%	1.986%	0.179%	0.347%	0.421%	1.944%
LSMC_1	−1.472%	1.629%	−0.490%	0.627%	−2.522%	2.561%
LSMC_2	−0.742%	1.148%	−0.490%	0.627%	−1.007%	2.334%
LSMC_Full	−0.501%	1.032%	−0.084%	0.461%	−0.420%	1.070%

As is shown in Table 8, the errors relative to the LSMC\_Full approach are lower than those of the other proposed methods since the estimates are based on the entire insurance portfolio, i.e., this approach does not suffer of any uncertainty related to the missingness of policies in its estimation procedure. Figure A2 given in Appendix B reports the box-plots on which the quantities in Table 8 are based on.

Finally, we compare these methods in terms of time. In Table 9, we report the computational time required by the algorithms. We can see that the naive application of the LSMC approach is more computationally expensive with respect to the GB2 and LSMC models based on a set of representative policies.

**Table 9.** Runtime, in seconds, of GB2 model and LSMC methods based on  $10,000 \times 2$  simulations and  $s = 100$  representative contracts (GB2, LSMC\_1, LSMC\_2). LSMC\_Full refers to the LSMC method applied to each contract in the insurance portfolio.

Method	Time
GB2	1911.445
LSMC_1	87.824
LSMC_2	88.290
LSMC_Full	7847.960



## 5. Conclusions

In this paper, we addressed the problem of approximating the value of a life insurance portfolio at a future time by proposing two different methodologies able to avoid the time-consuming nested simulations approach. The first approach can be thought of as extension of the well-known LSMC method, while the second is based on the GB2 distribution, which is widely used to approximate the fair value of portfolios of life insurance policies. To validate the proposal, we have considered a solid benchmark obtained by nested simulations, and we compared the two proposed methodologies both in terms of accuracy and efficiency. The analysis has been carried out by considering an ever increasing number of simulations and representative policies, from which it turned out that, generally, both the methodologies are able to provide increasingly accurate results. Moreover, the LSMC method proved to be faster in computational terms but also less accurate than the GB2 model. Furthermore, the proposed methodologies have been compared with a straightforward application of the LSMC method (i.e., without considering any subset of representative policies), which turned out to be more accurate but computationally more expensive.

Extensive numerical results have shown that the proposed methods represent viable alternatives to the full nested Monte Carlo model. Therefore, the proposed metamodeling approach may help insurance and reinsurance undertakings to reduce the computational budget needed, for instance, in the context of evaluating solvency capital requirements. In this regard, it can be used to evaluate the future cash-flows (inflows and outflows) generated by the entire portfolio by considering at first only a subset of policies, and then extend to the remaining ones. Indeed, this represents the main issue for deriving the full loss distribution on which the Value-at-Risk measure should be obtained, as prescribed by the European Solvency II directive.

**Author Contributions:** Both authors contributed equally to this manuscript. Both authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Parameter Values

Table A1 shows the parameter values assumed for the dynamics of the reference asset and interest rates defined in Equations (1) and (2).

**Table A1.** Parameters of the reference asset value process,  $S$ , and interest rate stochastic process,  $r$ .

$S_0$	$\sigma_S$	$\lambda$	$r_0$	$\alpha$	$\theta$	$\sigma_r$	$\gamma$	$\rho$
100	0.20	0.00	0.04	0.10	0.02	0.02	0.00	0.00

Table A2 shows the estimated parameters of the mortality model defined in Equation (3) obtained by fitting the corresponding survival curve on that implied by the Italian males and females mortality data in year 2016 obtained from the Human Mortality Database for each age  $x \in \{55, \dots, 65\}$ .

**Table A2.** Estimated parameters of the stochastic mortality model for Italian male (left) and female (right) aged  $x \in \{55, \dots, 65\}$  in 2016.

Age	Male			Female		
	$\hat{a}$	$\hat{b}$	$\hat{\sigma}_\mu$	$\hat{a}$	$\hat{b}$	$\hat{\sigma}_\mu$
55	0.00040	0.0881	0.00157	0.00010	0.10017	0.00100
56	0.00700	0.0705	0.00262	0.00001	0.11110	0.00100
57	0.00001	0.1051	0.00100	0.00001	0.11060	0.00100
58	0.00001	0.1045	0.00390	0.00009	0.10740	0.00850
59	0.00040	0.0832	0.00100	0.00001	0.11570	0.00100
60	0.00060	0.0743	0.00100	0.00042	0.08362	0.00669
61	0.00030	0.0907	0.00100	0.00044	0.08505	0.00100
62	0.00010	0.1033	0.00710	0.00001	0.11990	0.00100
63	0.00012	0.1063	0.00750	0.00040	0.09704	0.00182
64	0.00008	0.1112	0.00810	0.00039	0.09860	0.00376
65	0.00020	0.1075	0.00123	0.00049	0.09558	0.00720

## Appendix B. Further Results

Figure A1 reports the boxplot relative to the mean (left) and the 99.5th percentile (right) estimates obtained by running 50 times the GB2 and LSMC methods varying both the number of outer scenarios,  $n$ , and that of the representative policies,  $s$ . In this regard, we can see that the variability of the estimates decreases as the number of outer scenarios and the number of representative contracts increases.

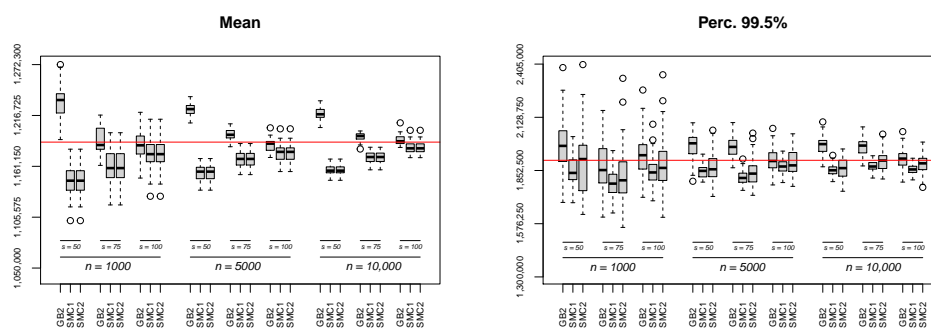
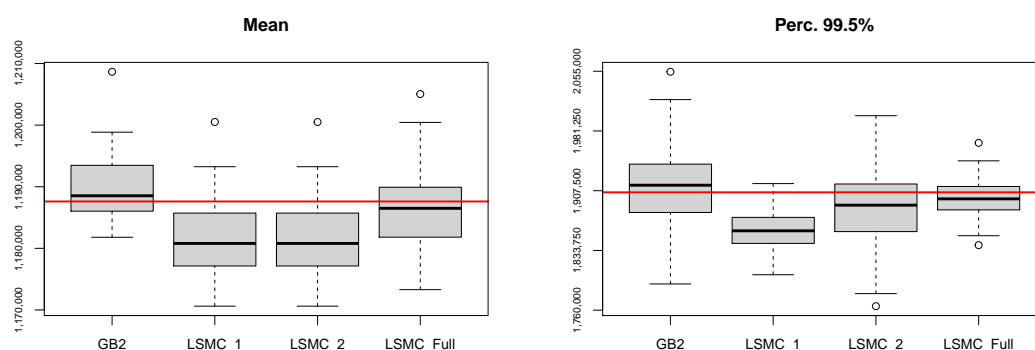
**Figure A1.** Boxplots relative to the mean (left) and the 99.5th percentile (right) estimates obtained by running the GB2 and LSMC methods 50 times and varying the number of outer simulations  $n$  and that of representative contracts  $s$ . The red line refers to the benchmark value based on a nested simulations algorithm with  $10,000 \times 2500$  trajectories applied to the entire portfolio.

Figure A2 compares the straightforward application of the LSMC approach with respect to the proposed methodologies providing the boxplots relative to the mean and the 99.5th percentile estimates.



**Figure A2.** Boxplots relative to the mean and the 99.5th percentile estimates obtained by running the proposed methodologies 50 times. GB2 stands for the GB2 regression model based on 10,000 outer scenarios and  $s = 100$  representative policies; LSMC\_1 refers to the LSMC method based on 10,000 outer scenarios and  $s = 100$  representative policies with Hermite polynomials of order 1; LSMC\_2 refers to the LSMC method based on 10,000 outer scenarios and  $s = 100$  representative policies with Hermite polynomials of order 2; LSMC\_Full refers to the LSMC method based on 10,000 outer scenarios and constructed on each contract in the insurance portfolio. The red line refers to the benchmark value based on a nested simulations algorithm with  $10,000 \times 2500$  trajectories applied to the entire portfolio.

## References

- Allemang, Dean, and Jim Hendler. 2011. *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*, 2nd ed. San Francisco: Morgan Kaufmann Publishers Inc.
- Barton, Russel R. 2015. Tutorial: Simulation metamodeling. Paper presented at 2015 Winter Simulation Conference (WSC), Huntington Beach, CA, USA, December 6–9.
- Bauer, Daniel, Daniela Bergmann, and Andreas Reuss. 2010. Solvency II and Nested Simulations—a Least-Squares Monte Carlo Approach. Working Paper. Georgia State University and Ulm University. Available online: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.466.1983> (accessed on 17 August 2021).
- Beirlant, Jan, Yuri Goegebeur, Johan Segers, and Jozef L. Teugels. 2004. *Statistics of Extremes: Theory and Applications*. Chichester: Wiley.
- Biffis, Enrico. 2005. Affine processes for dynamic mortality and actuarial valuations. *Insurance: Mathematics and Economics* 37: 443–68.
- Boyer, M. Martin, and Lars Stentoft. 2013. If we can simulate it, we can insure it: An application to longevity risk management. *Insurance: Mathematics and Economics* 52: 35–45.
- Carrière, Jacques F. 1996. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics* 19: 19–30.
- Cathcart, Mark, and Steven Morrison. 2009. Variable annuity economic capital: The least-squares Monte Carlo approach. *Life & Pensions* 2: 44–48.
- European Parliament, and European Council. 2009. *Directive 2009/138/EC on the Taking-Up and Pursuit of the Business of Insurance and Reinsurance (Solvency II)*. Brussels: European Council.
- Floryszczak, Anthony, Olivier Le Courtois, and Mohamed Majri. 2016. Inside the Solvency II black box: Net Asset Values and Solvency Capital Requirements with a least-squares Monte-Carlo approach. *Insurance: Mathematics and Economics* 71: 15–26.
- Frees, Edward W., and Emiliano A. Valdez. 2008. Hierarchical Insurance Claims Modeling. *Journal of the American Statistical Association* 103: 1457–69.
- Frees, Edward W., Gee Lee, and Lu Yang. 2016. Multivariate Frequency-Severity Regression Models in Insurance. *Risks* 4: 4.
- Fung, Man Chung, Katja Ignatieva, and Michael Sherris. 2014. Systematic mortality risk: An analysis of guaranteed lifetime withdrawal benefits in variable annuities. *Insurance: Mathematics and Economics* 58: 103–15.
- Gan, Guojun, and Emiliano A. Valdez. 2018. Regression modeling for the valuation of large variable annuity portfolios. *North American Actuarial Journal* 22: 40–54.
- Gan, Guojun, and X. Sheldon Lin. 2015. Valuation of large variable annuity portfolios under nested simulation: A functional data approach. *Insurance: Mathematics and Economics* 62: 138–50.
- Gan, Guojun. 2013. Application of data clustering and machine learning in variable annuity valuation. *Insurance: Mathematics and Economics* 53: 795–801.
- Gan, Guojun. 2015. Application of metamodeling to the valuation of large variable annuity portfolios. Paper presented at 2015 Winter Simulation Conference (WSC), Huntington Beach, CA, USA, December 6–9.
- Krah, Anne-Sophie, Zoran Nikolić, and Ralf Korn. 2018. A least-squares Monte Carlo framework in proxy modeling of life insurance companies. *Risks* 6: 2–26.

- 
- Loeppky, Jason L., Jerome Sacks, and William J. Welch. 2009. Choosing the sample size of a computer experiment: A practical guide. *Technometrics* 51: 366–76.
- Longstaff, Francis A., and Eduardo S. Schwartz. 2001. Valuing American options by simulations: A simple least-squares approach. *The Review of Financial Studies* 1: 113–47.
- McDonald, James B. 1984. Some generalized functions for the size distribution of income. *Econometrica* 52: 647–63.
- Minasny, Budiman, and Alex B. McBratney. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences* 32: 1378–88.
- Sun, Jiafeng, Edward W. Frees, and Marjorie A. Rosenberg. 2008. Heavy-tailed longitudinal data modeling using copulas. *Insurance: Mathematics and Economics* 42: 817–30.
- Tilley, James A. 1993. Valuing American options in a path simulation model. *Transactions of Society of Actuaries* 45: 499–520.