



Dipartimento di scienze economiche,
aziendali, matematiche e statistiche
“Bruno de Finetti”

Research Paper Series, N. 3, 2021

Modelling the Future Value Distribution of a Life Insurance Portfolio

MASSIMO COSTABILE

*Department of Economics, Statistics and Finance
University of Calabria, Ponte Bucci Cubo 0 C, 87036 Rende (CS), Italy*

FABIO VIVIANO

*Department of Business, Economics, Mathematics and Statistics “B. de Finetti”
University of Trieste, Piazzale Europa 1, 34127 Trieste, Italy
Department of Economics and Statistics
University of Udine, Via Tomadini 30/A, 33100 Udine, Italy*



UNIVERSITÀ
DEGLI STUDI DI TRIESTE

Research Paper Series

Dipartimento di Scienze Economiche, Aziendali, Matematiche e Statistiche

“Bruno de Finetti”

Piazzale Europa 1

34127, Trieste

Tel. +390405587927

Fax +390405587033

<http://www.deams.units.it>

EUT Edizioni Università di Trieste

Via E.Weiss, 21 - 34128 Trieste

Tel. +390405586183

Fax +390405586185

<http://eut.units.it>

eut@units.it

ISBN: 978-88-5511-258-1



9 788855 112581 >

Modelling the Future Value Distribution of a Life Insurance Portfolio

MASSIMO COSTABILE

*Department of Economics, Statistics and Finance
University of Calabria, Ponte Bucci Cubo 0 C, 87036 Rende (CS), Italy*

FABIO VIVIANO

*Department of Business, Economics, Mathematics and Statistics "B. de Finetti"
University of Trieste, Piazzale Europa 1, 34127 Trieste, Italy
Department of Economics and Statistics
University of Udine, Via Tomadini 30/A, 33100 Udine, Italy*

ABSTRACT¹

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 Monte Carlo simulations we obtain a rough estimate of the policies' value 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, and a regression method based on the class of generalized beta of the second kind distributions. Extensive numerical experiments are provided to assess the performance of the proposed models.

KEYWORDS: GB2, LSMC, Meta-model, Regression models, Solvency II

¹**Corresponding author:** Fabio Viviano, email: viviano.fabio@spes.uniud.it

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 require to maintain solvency capital requirements in order to continue to conduct business, as stated in Solvency II directive or in the Swiss Solvency Test. As a consequence, insurers are obliged to assess the value of assets and liabilities at a future date, the so-called risk-horizon. To do 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. 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 literature to reduce the computational effort and to preserve the accuracy of the desired estimates, is the Least-Squares Monte Carlo, firstly introduced by Carrière (1996), Tilley (1993), and Longstaff and Schwartz (2001) in the context of American-type Option Pricing. An 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 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 10000 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 by 10000 contracts, the total number of cash-flow projections needed would be $10000 \times v \times 2500 \times 12 \times 20 \times 10000 = v \times 6 \times 10^{13}$, which is very hard to be managed.

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 been widely used also in actuarial literature to estimate the price and the Greeks of large portfolios of life insurance policies. For instance, Gan (2013) develops 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 the Latin hypercube sampling to select representative variable annuities. Finally, Gan and Valdez (2018) propose 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 modelling 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 underscribed 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 each other also in terms of maturity, policyholder's age, and sex. In particular, we consider unit-linked products, term life insurance and immediate life annuities. We assume that the unit-linked product pays at maturity, upon the survival of the insured, a maximum between a minimum guaranteed benefit and the value of a specific reference asset. The immediate life annuity is assumed to pay continuously the 10% of the level of a given reference asset until the insured is alive; and 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.

<i>Feature</i>	<i>Value</i>
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), as in Gan and Valdez (2018). More precisely, 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 longevity, interest rate and the reference asset are taken into account.

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 σ -algebra generated by $\mathcal{G}_t \cup \mathcal{H}_t$, where

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

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 ς is a stopping time.

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

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

where λ is the risk-premium, σ_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

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

Here $W^{2,\mathbb{P}}(t)$ is a standard Wiener process, and the coefficients α, θ, σ_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 correlation coefficient ρ .

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 deeper details. 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

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

Note that $W^{1,\mathbb{Q}}(t)$ and $W^{2,\mathbb{Q}}(t)$ are two correlated standard Wiener processes with coefficient of correlation ρ under \mathbb{Q} .

Concerning longevity risk, 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 $W^{2,\mathbb{P}}(t)$.

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 δ is the market price of the systematic mortality risk.

Note that the parameters entering in the stochastic mortality model are estimated by calibrating the implied survival curve to the one obtained from the population data of year 2016 (assumed to be $t = 0$) collected from the Human Mortality Database (see Fung et al., 2014). The calibration procedure is done for all policyholder ages and genders reported in Table 1.

3. Problem and methodology

Under the framework defined in Section 2, we need to evaluate the streams of payments embedded in each policies inside the insurance portfolio. For this purpose, we project the relevant risk factors affecting the policy (i.e. S, r , and μ) under the physical probability

measure from time $t = 0$ up to the risk-horizon τ , 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 a $n \times v$ 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,

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

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 τ . The next step is to obtain a more accurate estimate of the distribution of the time- τ value of each representative policy and then to infer the distribution of the time- τ value of the entire portfolio. We do this by applying two different approaches, an OLS as in the Least-Squares Monte Carlo method and a GB2 model.

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, also this kind of strategy would be quite computationally expensive due to the big dimension of an insurance portfolio. Because of that, we propose to apply firstly the LSMC method on just a set of representative policies, and then through an OLS regression we extend it to the entire portfolio.

Hence, accordingly 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:

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

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

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

In this way, we obtain a $n \times s$ 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} a $M \times (w+1)$ 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 $\bar{\mathbf{X}}$ be the $s \times (w+1)$ matrix describing the structure of the representative insurance portfolio. Hence, $\bar{\mathbf{x}}_i$ 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 $\bar{\mathbf{X}}$, and once the coefficients are estimated, we extend then 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,

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

where

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

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

The GB2 model

A GB2 model appears to provide a flexible family of distributions; indeed, 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, and $B(\cdot)$ is the Beta function.

In order to approximate the value of the portfolio, at first we approximate the time- τ 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 do this, we construct two different GB2 regression models which exploit respectively the

generated information at the risk-horizon (i.e. $S(\tau)$, $r(\tau)$, and $\mu(\tau)$), and then the features characterizing uniquely each policy.

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 τ 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 depending on some independent covariates (i.e. the value at time τ of the risk factors which affect the policy of interest).

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

$$l(a_i, p_i, q_i, \boldsymbol{\beta}_i) = n \ln \frac{|a_i|}{B(p_i, q_i)} - a_i p_i \sum_{k=1}^n \mathbf{z}_k^i \boldsymbol{\beta}'_i + (a_i p_i - 1) \sum_{k=1}^n \ln(y_{ik}) - (p_i + q_i) \sum_{k=1}^n \ln \left[1 + \left(\frac{y_{ik}}{\exp(\mathbf{z}_k^i \boldsymbol{\beta}'_i)} \right)^{a_i} \right], \quad (8)$$

where $i = 1, \dots, n$ 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 τ . Since we incorporate covariates through the scale parameter, we can estimate it as

$$\hat{y}_{ik} = \frac{\exp(\mathbf{z}_k^i \hat{\boldsymbol{\beta}}'_i) B\left(\hat{p}_i + \frac{1}{\hat{a}_i}, \hat{q}_i - \frac{1}{\hat{a}_i}\right)}{B(\hat{p}_i, \hat{q}_i)}, \quad i = 1, 2, \dots, s \text{ and } k = 1, \dots, n, \quad (9)$$

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 τ , we extend this information to the remaining policies. As already done 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 (9).

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

$$l(a_k, p_k, q_k, \boldsymbol{\beta}_k) = s \ln \frac{|a_k|}{B(p_k, q_k)} - a_k p_k \sum_{i=1}^s \bar{\mathbf{x}}_i \boldsymbol{\beta}'_k + (a_k p_k - 1) \sum_{i=1}^s \ln(\hat{y}_{ik}) - (p_k + q_k) \sum_{i=1}^s \ln \left[1 + \left(\frac{\hat{y}_{ik}}{\exp(\bar{\mathbf{x}}_i \boldsymbol{\beta}'_k)} \right)^{a_k} \right], \quad (10)$$

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 maximum likelihood approach, we can then derive the distribution at the risk-horizon for all the policies inside the insurance portfolio as

$$\widehat{v}_{ik} = \frac{\exp\left(\mathbf{x}_i \widehat{\beta}'_k\right) B\left(\widehat{p}_k + \frac{1}{\widehat{a}_k}, \widehat{q}_k - \frac{1}{\widehat{a}_k}\right)}{B(\widehat{p}_k, \widehat{q}_k)}, \quad i = 1, 2, \dots, M \text{ and } k = 1, \dots, n, \quad (11)$$

where \widehat{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 summing up all the policy values corresponding to each outer scenario.

Note that the log-likelihood functions in Equations (8) and (10) 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 $M = 10000$ contracts, and we focus on approximating its value distribution at the future time $\tau = 1$ year. These policies can be of three different types: 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 10% of a reference asset value 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 initial value S_0 .

Concerning the dynamics of the financial risk factors, we report in Table 2 the values of the involved parameters in Equations (1) and (2).

TABLE 2. Parameters of the reference asset value process, S , and interest rate stochastic process, r .

S_0	σ_S	λ	r_0	α	θ	σ_r	ρ
100	0.20	0.00	0.04	0.10	0.02	0.02	0.00

Concerning mortality, we have calibrated the survival curve implied by Equation (3) on the Italian males and females mortality data in year 2016 obtained from the Human Mortality Database for each age $x \in \{55, \dots, 65\}$, and we assumed a longevity risk-premium $\delta = 0$. The corresponding parameter estimates are reported in Table 3.

TABLE 3. 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

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 simulations set, we are able to obtain a first rough estimate of \mathbf{Y} on which we construct the LSMC and GB2 models discussed in Section 3. Note that, concerning the LSMC method, we exploit as basis functions Hermite polynomials of order 1 and 2 which are denoted, respectively, LSMC_1 and LSMC_2 hereafter.

To determine the number of representative contracts s , we start from the informal rule proposed by Loepky 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 \mathbf{X} is 5 (including the binary dummy variables converted from the categorical variables), and so we choose as the initial number of representative contracts $s = 50$. However, we investigate the models performance also by setting $s = 75$ and $s = 100$.

Finally, the results are compared with a solid benchmark obtained through a nested simulations approach based on 10000×2500 simulations. This allows us to conclude about the reliability of the proposed methodologies, and to compare them in terms of computational demand. For a comprehensive analysis we perform multiple runs of each proposed method; in particular, the following analysis is based on 50 runs.

In Tables 4, 5 and 6 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 ($n = 10000$) and by varying the number of representative contracts ($s \in \{50, 75, 100\}$).

TABLE 4. This table reports the MAPE of the estimates obtained by running 50 times the GB2 and LSMC methods with $n = 10000$ and $s = 50$. The benchmark values are based on a nested simulations algorithm with 10000×2500 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 5. This table reports the MAPE of the estimates obtained by running 50 times the GB2 and LSMC methods with $n = 10000$ and $s = 75$. The benchmark values are based on a nested simulations algorithm with 10000×2500 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 6. This table reports the MAPE of the estimates obtained by running 50 times the GB2 and LSMC methods with $n = 10000$ and $s = 100$. The benchmark values are based on a nested simulations algorithm with 10000×2500 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 4, 5 and 6, it is evident that increasing the number of representative contracts s induces 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 model adequately the right tail of the distribution.

In Table 7, we report the Mean Percentage Error (MPE) and MAPE relative to the mean estimates obtained by running 50 times the GB2 and LSMC methods with different numbers of outer simulations, n , and representative contracts, s .

TABLE 7. 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 10000×2500 trajectories applied to the entire portfolio.

Outer	Method	$s=50$		$s=75$		$s=100$	
		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%
10000	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 7, 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 7 ($s = 100$), 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 10000 paths), not sufficient to appreciate the improvement which is usually expected. In Figure 1, 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.

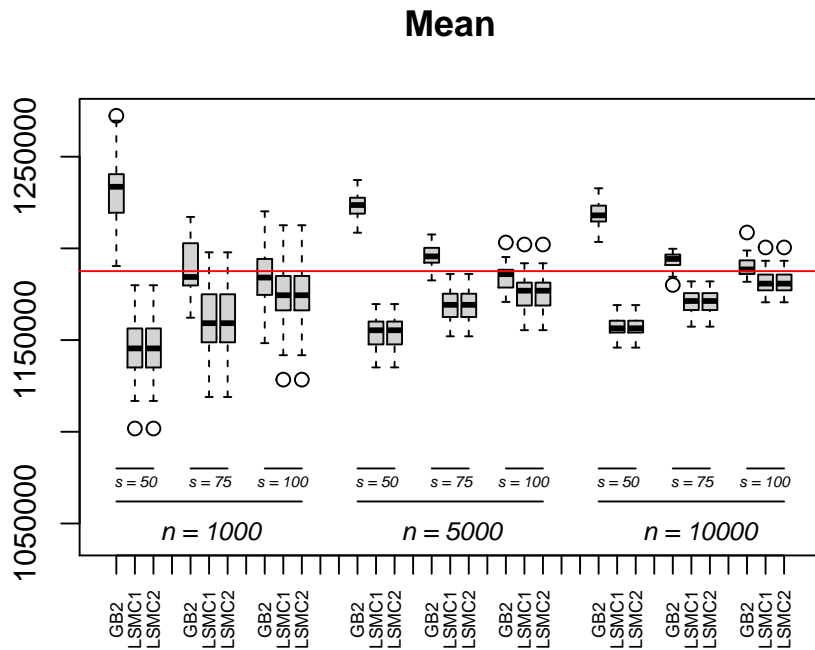


FIG. 1. Boxplots relative to the **mean** estimates obtained by running 50 times the GB2 and LSMC methods and varying the number of outer simulations n and that of representative contracts s . The red line refers to the benchmark value based on a nested simulations algorithm with 10000×2500 trajectories applied to the entire portfolio.

Concerning the estimate of the 99.5-th percentile of the distribution, which would be of interest for valuing solvency capital requirements, Table 8 reports the MPE and MAPE relative to 50 estimates obtained by varying both the number of simulations and the number of representative contracts.

TABLE 8. This table reports the MPE and MAPE of the **99.5-th percentile** 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 10000×2500 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.1	-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%
10000	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 8, 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 fixed n) does not lead to a clear improvement of 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. As in the case of the mean estimate this could be due to the small number of outer simulations. Once again, if we look at the case of $n = 10000$ and $s = 100$, the GB2 model outperforms the LSMC approach. Figure 2 shows the box-plots relative to the 99.5-th percentile estimates.

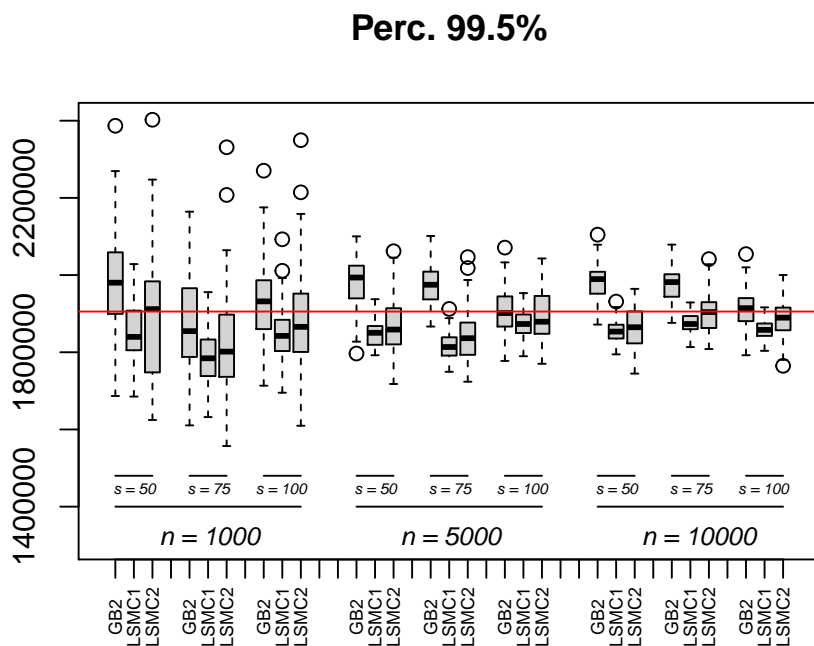


FIG. 2. Boxplots relative to the **99.5-th percentile** estimates obtained by running 50 times the GB2 and LSMC methods and varying the number of outer simulations n and that of representative contracts s . The red line refers to the benchmark value based on a nested simulations algorithm with 10000×2500 trajectories applied to the entire portfolio.

From Figure 2, we can see that the variability of the estimates decreases as the number of outer scenarios and the number of representative contracts increases. Moreover, there is a clear pattern in the GB2 model results since an increase in both number of simulations and representative policies helps in reaching the convergence to the benchmark value. Regarding LSMC, we can see that in each of the considered configurations, the method built on Hermite polynomials of order 1 (LSMC.1) produces less variable estimates compared to the second order Hermite polynomials (LSMC.2), even if we increase the number of outer simulations. We may conclude that passing from 1000 to 10000 trajectories is still not sufficient to exploit more basis functions.

Now let us examine the speed of the proposed algorithms with respect to the benchmark. Table 9 shows the runtime of GB2 and LSMC expressed as a percentage of the time required by the nested simulation method based on 10000 outers and 2500 inners. 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 9. 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 10000×2500 scenarios applied to the entire portfolio is about 187200 seconds.

Method	$n = 1000$			$n = 5000$			$n = 10000$		
	$s = 50$	$s = 75$	$s = 100$	$s = 50$	$s = 75$	$s = 100$	$s = 50$	$s = 75$	$s = 100$
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 9, 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 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, the estimation procedure for the GB2 model is based on a multi-stage optimization algorithm due to the complexity of the likelihood functions which may have multiple local maxima. Anyway, if compared with the simulations within simulations method, the GB2 model proved to be an accurate and efficient alternative.

Full LSMC

In this Section, 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 10000×2 simulations (LSMC_Full). Table 10 reports the MPE and MAPE relative to the 5th-percentile, the mean, and the 99.5-th 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 10000×2 simulations and $s = 100$ representative policies.

TABLE 10. This table reports the MPE and MAPE relative to the 5-th percentile, the mean, and the 99.5-th percentile estimates obtained by applying different methodologies. GB2 stands for the GB2 regression model based on $n = 10000$ outer scenarios and $s = 100$ representative policies; LSMC_1 refers to the LSMC method based on $n = 10000$ outer scenarios and $s = 100$ representative policies with Hermite polynomials of order 1; LSMC_2 refers to the LSMC method based on $n = 10000$ outer scenarios and $s = 100$ representative policies with Hermite polynomials of order 2; LSMC_Full refers to the LSMC method based on $n = 10000$ 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 10000×2500 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 it is shown in Table 10, 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 3 reports the box-plots on which the quantities in Table 10 are based on.

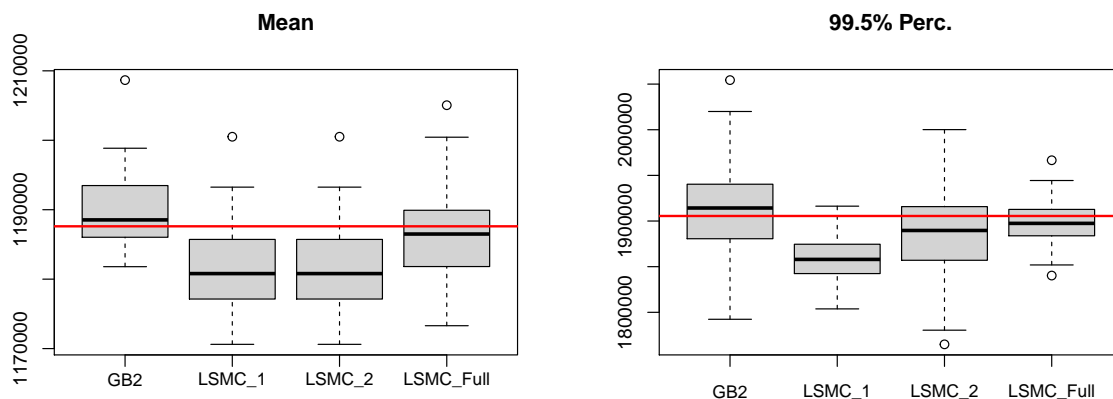


FIG. 3. Boxplots relative to the **Mean** and the **99.5-th percentile** estimates obtained by running 50 times the proposed methodologies. GB2 stands for the GB2 regression model based on 10000 outer scenarios and $s = 100$ representative policies; LSMC_1 refers to the LSMC method based on 10000 outer scenarios and $s = 100$ representative policies with Hermite polynomials of order 1; LSMC_2 refers to the LSMC method based on 10000 outer scenarios and $s = 100$ representative policies with Hermite polynomials of order 2; LSMC_Full refers to the LSMC method based on 10000 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 10000×2500 trajectories applied to the entire portfolio.

Finally, we compare these methods in terms of time. In Table 11, 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 11. Runtime, in seconds, of GB2 Model and LSMC methods based on 10000×2 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.

<i>Method</i>	Time
GB2	1911.445
LSMC_1	87.8236
LSMC_2	88.29
LSMC_Full	7847.96

5. Conclusion

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, widely used to approximate the fair value of portfolios of life insurance policies. Extensive numerical results have shown that the proposed methods represent viable alternatives to the full nested Monte Carlo model.

REFERENCES

- Dean Allemang and James Hendler. *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2 edition, 2011.
- R. R. Barton. Tutorial: Simulation metamodeling. In *2015 Winter Simulation Conference (WSC)*, pages 1765–1779, 2015. doi: 10.1109/WSC.2015.7408294.
- Daniel Bauer, J Mack, Daniela Bergmann, and Andreas Reuss. Solvency II and nested simulations: A least-squares monte carlo approach. 01 2010.
- Enrico Biffis. Affine processes for dynamic mortality and actuarial valuations. *Insurance: Mathematics and Economics*, 37(3):443–468, 2005. ISSN 0167-6687. doi: <https://doi.org/10.1016/j.insmatheco.2005.05.003>. URL <https://www.sciencedirect.com/science/article/pii/S0167668705000600>.
- Jacques F. Carrière. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics*, 19(1):19–30, 1996. ISSN 0167-6687. doi: [https://doi.org/10.1016/S0167-6687\(96\)00004-2](https://doi.org/10.1016/S0167-6687(96)00004-2). URL <https://www.sciencedirect.com/science/article/pii/S0167668796000042>.
- M. Cathcart and S. Morrison. Variable annuity economic capital: The least-squares monte carlo approach. *Life & Pensions*, pages 44–48, 2009.
- Anthony Floryszczak, Olivier Le Courtois, and Mohamed Majri. Inside the solvency 2 black box: Net asset values and solvency capital requirements with a least-squares monte-carlo approach. *Insurance: Mathematics and Economics*, 71:15–26, 2016. ISSN 0167-6687. doi: <https://doi.org/10.1016/j.insmatheco.2016.07.005>. URL <https://www.sciencedirect.com/science/article/pii/S0167668716303080>.
- Man Chung Fung, Katja Ignatieva, and Michael Sherris. Systematic mortality risk: An analysis of guaranteed lifetime withdrawal benefits in variable annuities. *Insurance: Mathematics and Economics*, 58:103–115, 2014. ISSN 0167-6687. doi: <https://doi.org/10.1016/j.insmatheco.2014.06.010>. URL <https://www.sciencedirect.com/science/article/pii/S0167668714000821>.

- G. Gan. Application of metamodeling to the valuation of large variable annuity portfolios. In *2015 Winter Simulation Conference (WSC)*, pages 1103–1114, 2015. doi: 10.1109/WSC.2015.7408237.
- Guojun Gan. Application of data clustering and machine learning in variable annuity valuation. *Insurance: Mathematics and Economics*, 53(3):795–801, 2013. ISSN 0167-6687. doi: <https://doi.org/10.1016/j.insmatheco.2013.09.021>. URL <https://www.sciencedirect.com/science/article/pii/S0167668713001546>.
- Guojun Gan and X. Sheldon Lin. Valuation of large variable annuity portfolios under nested simulation: A functional data approach. *Insurance: Mathematics and Economics*, 62:138–150, 2015. ISSN 0167-6687. doi: <https://doi.org/10.1016/j.insmatheco.2015.02.007>. URL <https://www.sciencedirect.com/science/article/pii/S0167668715000219>.
- Guojun Gan and Emiliano A. Valdez. Regression modeling for the valuation of large variable annuity portfolios. *North American Actuarial Journal*, 22(1):40–54, 2018. doi: 10.1080/10920277.2017.1366863. URL <https://doi.org/10.1080/10920277.2017.1366863>.
- Anne-Sophie Krah, Zoran Nikolić, and Ralf Korn. A least-squares monte carlo framework in proxy modeling of life insurance companies. *Risks*, 6(2), 2018. ISSN 2227-9091. doi: 10.3390/risks6020062. URL <https://www.mdpi.com/2227-9091/6/2/62>.
- Jason L. Loeppky, Jerome Sacks, and William J. Welch. Choosing the sample size of a computer experiment: A practical guide. *Technometrics*, 51(4):366–376, 2009. doi: 10.1198/TECH.2009.08040. URL <https://doi.org/10.1198/TECH.2009.08040>.
- Francis A. Longstaff and Eduardo S. Schwartz. Valuing american options by simulation: A simple least-squares approach. *The Review of Financial Studies*, 14(1):113–147, 06 2001. ISSN 0893-9454. doi: 10.1093/rfs/14.1.113. URL <https://doi.org/10.1093/rfs/14.1.113>.
- B. Minasny and A. B. McBratney. A conditioned latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences*, page 1378–1388, 2006.
- J. A. Tilley. Valuing american options in a path simulation model. *Transactions of Society of Actuaries*, 45:499–520, 1993.