DETRA NOTE

DETRA NOTE 2023-1

RISK MANAGEMENT WITH LOCAL LEAST SQUARES MONTE-CARLO

By Adnane Akbaraly & Donatien Hainaut

Detra²lytics

DISCLAIMER

Risk management with Local Least Squares Monte-Carlo

Donatien Hainaut *
UCLouvain & Detralytics
Adnane Akbaraly†
Detralytics

January 24, 2023

Abstract

The method of least squares Monte-Carlo (LSMC) has become a standard in the insurance and financial sectors for computing the exposure of a company to market risk. The sensitive point of this procedure is the non-linear regression of simulated responses on risk factors. This article proposes a novel approach for this step, based on an a-priori segmentation of responses. Using a K-means algorithm, we identify clusters of responses that are next locally regressed on corresponding risk factors. A global function of regression is obtained by combining local models and a logistic regression. The efficiency of the Local Least squares Monte-Carlo (LLSMC) is checked in two illustrations. The first one focuses on butterfly and bull trap options in a Heston stochastic volatility model. The second illustration analyzes the exposure to risks of a participating life insurance.

KEYWORDS: Least square Monte-Carlo, risk management, option valuation
JEL CODE : C5, G22

1 Introduction

The least squares Monte-Carlo method (LSMC) of Longstaff and Schwartz (2001) is a powerful and simple simulation method for pricing path dependent options. By its nature, simulation is an alternative to traditional finite difference and binomial techniques in particular when the value of the option depends on multiple factors. The LSMC method is based on the property that the conditional expectation of a random process minimizes the mean squared distance between a simulated sample of this process and an adapted Borel measurable function. This function is approximated by a regression in a subspace of basis functions.

Clement and al. (2002) prove under fairly general conditions, the almost sure convergence of LSMC. Glasserman and Yu (2004) investigate the behavior of this algorithm with the simultaneous grows of the number of the basis functions and the number of the Monte-Carlo simulations. Moreno and Navas (2003) and Stentoft (2004) consider the LSMC for different basis functions and deduce that the algorithm converges at least for American put options. This technique is also used for the valuation of insurance liabilities. For instance, Bacinello et al. (2009, 2010) price a unit linked contracts embedding American options.

*Postal address: Voie du Roman Pays 20, 1348 Louvain-la-Neuve (Belgium). E-mail to: donatien.hainaut(at)uclouvain.be
†Postal address: Rue Réaumur, 124, 75002 Paris. E-mail to: a.akbaraly(at)detralytics.eu

The LSMC is not only useful for pricing but also for managing risk. Bauer et al. (2012) adapt the LSMC method for computing the required risk capital in the Solvency II framework. Pelsser and Schweizer (2016) compare LSMC and portfolio replication for the modeling of life insurance liabilities. Floryszczak et al. (2016) confirm that using the LSMC method is relevant for Solvency 2 computations at the level of a company. In the insurance sector, the LSMC has become a standard. For instance, case studies from the industry are proposed in Hörig and Leitschkis (2012) or Hörig and al. (2014).

More recently, the standard least-squares regression has been replaced by a neural network approximation. Becker et al. (2020) use this for pricing and hedging American-style option with deep learning. Lapeyre and Lelong (2021) develop a similar approach for Bermudan option pricing. In insurance, Hejazi and Jackson (2017) propose a neural network approach to evaluate the capital requirement of a portfolio of variable annuities. Cheridito et al. (2020) benchmark this approach to results of Bauer (2012).

The LSMC requires a global regression model predicting the responses as a function of risk factors. In its classical version, the regression function is approximated either by a polynomial or by a combination of basis functions. In case of a non-linear relation between responses and factors, the order of the polynomial or the number of basis functions is high and this increases the risk of overfitting. An efficient alternative consists to work with a neural regression. Nevertheless, determining the optimal network architecture is a challenging task and the model lacks of interpretability. The main contribution of this article is to propose an alternative based on local regressions. The first stage consists to allocate responses to a few clusters with the K-means algorithm and next to locally regress them on corresponding risk factors. In a second stage, we fit a logistic regression model that a priori estimates the probability that a combination of risk factors belongs to each cluster. A global regression function is obtained by weighting local models by these probabilities. We show in two case studies that this local least squares Monte-Carlo (LLSMC) outperforms the LSMC and present a high level of interpretability. The first case study focuses on the risk analysis of butterfly and bull trap options in a Heston stochastic volatility model. In the second case study, we consider a participating life insurance and compare risk measures computed with local and non local LSMC.

The outline of the article is as follows. Section 2 reviews the least squares Monte-Carlo method applied to risk management. The next Section introduces the LLSMC method and motivates the reasons for segmenting the data-set based on responses instead of risk factors. Section 4 compares the capacity of LLSMC and LSMC to replicate butterfly and bull trap options in a stochastic volatility model. In Section 5, we compare risk measures computed by local and non-local LSMC for a participating life insurance.

2 Risk management with the LSMC method

This section briefly reviews the least squares Monte-Carlo method. We consider a probability space Ω, endowed with a probability measure $\mathbb{P}$, in which are defined m processes, noted $\boldsymbol{X}_t = \left(X_t^{(1)}, ..., X_t^{(m)}\right)_{t\geq 0}$. These processes are the risk factors driving the value of financial assets and derivatives, managed by a financial institution. Their natural filtration is denoted by $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$. If risk factors are Markov, the total asset value is a function of time and risk factors denoted by $A(t, \boldsymbol{X}_t)$. Under the assumption of absence of arbitrage, there exists at least one equivalent risk neutral measure, denoted by $\mathbb{Q}$, using the cash account $(B_t)_{t\geq 0}$ as numeraire. Random asset cash-flows are paid at time $(t_k)_{k=0,...,d}$ and

denoted by C_k^A. Therefore at time $t \leq t_d$, $A(t, \boldsymbol{X}_t)$ can be developped as follows

$$A(t, \boldsymbol{X}_t) = a(\boldsymbol{X}_t) + \mathbb{E}^{\mathbb{Q}} \left(\sum_{k=0}^{d} \frac{B_t}{B_{t_k}} C_k^A \mathbf{1}_{\{t_k \geq t\}} | \boldsymbol{X}_t \right), \quad (1)$$

where $a(\boldsymbol{X}_t)$ is directly determined by the value of underlying risk factors. Let us consider a risk measure denoted by $\rho(.)$. For risk management, we aim to calculate $\rho(A(t, \boldsymbol{X}_t))$. In applications, we mainly consider the value at risk (VaR) and the expected shortfall (ES) as risk measures. For a confidence level $\alpha \in (0, 1)$, the VaR and ES are defined as

$$VaR_\alpha = \max \{x \in \mathbb{R} : \mathbb{P}(A(t, \boldsymbol{X}_t) \leq x) \leq \alpha\}, \quad (2)$$

$$ES_\alpha = \frac{1}{\alpha} \int_0^\alpha VaR_\gamma \, d\gamma. \quad (3)$$

The ES also admits an equivalent representation that is used later for estimation:

$$\begin{aligned} ES_\alpha &= \mathbb{E}^{\mathbb{P}} (A(t, \boldsymbol{X}_t) \,|\, A(t, \boldsymbol{X}_t) \leq VaR_\alpha) \quad (4) \\ &= \frac{1}{\alpha} \mathbb{E}^{\mathbb{P}} \left(A(t, \boldsymbol{X}_t) \mathbf{1}_{\{A(t, \boldsymbol{X}_t) \leq VaR_\alpha\}} \right) \\ &\quad + VaR_\alpha \frac{1}{\alpha} \left(1 - \mathbb{P}(A(t, \boldsymbol{X}_t) \leq VaR_\alpha)\right). \end{aligned}$$

We draw the attention of the reader on the fact that VaR and ES are valued under the real measure.

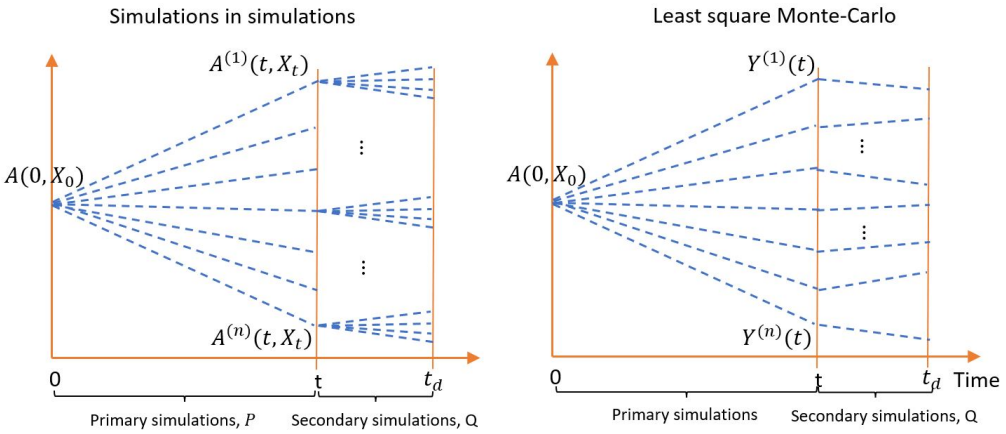

Figure 1: Simulations in simulations versus Least squares Monte-Carlo.

Computing the risk-neutral expectation (1) is a challenging task because closed-form expressions are usually not available. A solution consists to evaluate $A(t, \boldsymbol{X}_t)$ by simulations in simulations. This framework is illustrated in the left plot of Figure 1. For each primary simulated sample path of risk factors (under $\mathbb{P}$), we perform secondary simulations (under $\mathbb{Q}$). The value of $A(t, \boldsymbol{X}_t)$ is next obtained by averaging the sums of discounted cash-flows of secondary scenarios. This approach is nevertheless too computing intensive for being carried out with success. In practice, we rely on the method of Least squares Monte-Carlo (LSMC) to keep the computational time under control. We briefly recall what it consists in. For this purpose, let us denote by

$$Y(t) = \sum_{k=0}^{d} \frac{B_t}{B_{t_k}} C_k^A \mathbf{1}_{\{t_k \geq t\}},$$

the random variable that is $\mathcal{F}_{t_d}$-adapted and such that $A(t, \boldsymbol{X}_t) = a(\boldsymbol{X}_t) + \mathbb{E}^{\mathbb{Q}}(Y(t) \,|\, \boldsymbol{X}_t)$. This variable is called the "response" for a given set of risk factors et time t. The LSMC

method is based on property that the conditional expectation of a random variable $Y(t)$ given a random vector $\boldsymbol{X}_t$ minimizes the mean squared distance between $Y(t)$ and $h(\boldsymbol{X}_t)$ where $h(.)$ is a Borel measurable function. In practice, it means that we only need a single (or a few) secondary simulations under $\mathbb{Q}$, as illustrated on the right plot of Figure 1. The theoretical foundation of the LSMC approach is briefly recalled in the next proposition which uses the fact that $\boldsymbol{X}_t$ is also $\mathcal{F}_{t_d}$-adapted since $\mathcal{F}_t \subset \mathcal{F}_{t_d}$.

Proposition 2.1. *Let $Y(t)$ be a square-integrable random variable on $\mathbb{R}$ and $\boldsymbol{X}_t$ a m-dimensional random vector, both $\mathcal{F}_{t_d}$-adapted. The conditional expectation $\mathbb{E}^{\mathbb{Q}}\left(Y(t) \,|\, \boldsymbol{X}_t\right)$ is equal to a Borel measurable function $h(\boldsymbol{X}_t)$ such that*

$$h\left(\boldsymbol{X}_t\right) \;=\; \arg \min_{h \in \mathcal{B}(,\mathbb{R},\mathbb{R}^m)} \mathbb{E}^{\mathbb{Q}}\left(\left(h\left(\boldsymbol{X}_t\right) - Y(t)\right)^2\right) . \tag{5}$$

Proof Let us denote by $\nu_{\boldsymbol{X},Y}(\boldsymbol{x}, y)$ the joint probability density function (pdf) of $(\boldsymbol{X}_t, Y(t))$ and by $\nu_{\boldsymbol{X}}(\boldsymbol{x})$, $\nu_Y(y)$ the marginal pdf's. According to the Bayes rule, the conditional density of $Y(t)|\boldsymbol{X}_t$ is such that

$$\nu_{\boldsymbol{X},Y}(\boldsymbol{x}, y) = \nu_{Y|\boldsymbol{X}}(y|\boldsymbol{x})\nu_{\boldsymbol{X}}(\boldsymbol{x})$$

and the expectation in (5) may be rewritten as

$$\begin{aligned}&\mathbb{E}^{\mathbb{Q}}\left(\left(h(\boldsymbol{X}_t) - Y(t)\right)^2\right) = \\ &\int_{dom(\boldsymbol{X})} \int_{dom(Y)} \left(h(\boldsymbol{x}) - y\right)^2 \nu_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) dy \, \nu_{\boldsymbol{X}}(\boldsymbol{x}) \, dx \, .\end{aligned}$$

The function $h(\boldsymbol{X}_t)$ minimizes (5) if and only if

$$h(\boldsymbol{x}) = \arg\min \int_{dom(Y)} \left(h(\boldsymbol{x}) - y\right)^2 \nu_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) dy \, ,$$

which is achieved for $h(\boldsymbol{x}) = \mathbb{E}^{\mathbb{Q}}\left(Y(t)|\boldsymbol{X}_t = \boldsymbol{x}\right)$.

end

In many real-world applications, there is no closed form expression for the function $h(\boldsymbol{X_t})$ but risks factors can be simulated under the $\mathbb{P}$ measure. Longstaff and Schwartz (2001) assume that the unknown function $h(.)$ belongs to the L^2-space of square-integrable functions. Since L^2 is a Hilbert space, it admits a countable orthornormal basis. The function $h(.)$ may then be represented as a combination of basis functions. If $m = 1$, one common choice is the set of weighted Laguerre polynomials. In higher dimension, basis functions are usually replaced by polynomials of risk factors. In practice, the LSMC algorithm consists in simulating a sample denoted by

$$\mathcal{S} = \{(\boldsymbol{x}_1, y_1), ..., (\boldsymbol{x}_n, y_n)\} \, , \tag{6}$$

of n realizations of $(\boldsymbol{X}_t, Y(t))$ and in regressing responses on risk factors. We recall that $\boldsymbol{X}_t$ is simulated up to time t under the real measure $\mathbb{P}$ while the response $Y(t)$ is obtained by simulations from t up to t_d, under the risk neutral measure $\mathbb{Q}$. Let us denote by $\mathcal{P}_h$ the set of polynomials $\widehat{h}(\boldsymbol{x})$ of degree d_h approximating $h(\boldsymbol{x})$. It is estimated by least squares minimization:

$$\widehat{h}(.) \;=\; \arg \min_{\widehat{h} \in \mathcal{P}_h} \left(\sum_{(\boldsymbol{x}_i, y_i) \in \mathcal{S}} \left(y_i - \widehat{h}(\boldsymbol{x}_i)\right)^2 \right) . \tag{7}$$

Let us denote by

$$\boldsymbol{z} = \left(\left(x_{i_1}^{j_1} x_{i_2}^{j_2} ... x_{i_h}^{j_h} \right)_{i_1,...,i_h \in \{1,...,m\}}^{j_1,...,j_h \in \mathbb{N}, j_1+...+j_h \leq d_h} \right) , \tag{8}$$

the vector of powers of risk factors up to order $d_h \in \mathbb{N}$. We define $\widehat{h}(\boldsymbol{x}) = \boldsymbol{z}^\top \boldsymbol{\beta}$ as a polynomial of order d_h where $\boldsymbol{\beta}_k$ is a real vector of same dimension as $\boldsymbol{z}$. The sample of powers of risk factors is $\{\boldsymbol{z}_1, ..., \boldsymbol{z}_n\}$. Let us respectively denote by Z and $\boldsymbol{y}$, the matrix $Z = \left(\boldsymbol{z}_j^\top \right)_{j \in \mathcal{S}}$ and the vector $\boldsymbol{y} = (y_j)_{j \in \mathcal{S}}$. Using standard arguments, the polynomial coefficients minimizing (7) are $\widehat{\boldsymbol{\beta}} = \left(Z^\top Z \right)^{-1} Z^\top \boldsymbol{y}$.

Let us next denote by $\widehat{a}_i = a(\boldsymbol{x}_i) + \widehat{h}(\boldsymbol{x}_i) \approx A(t, \boldsymbol{x}_i)$, the approximation of the value of total assets for a given vector of risk factors $\boldsymbol{X}_t = \boldsymbol{x}$. The ordered samples of $(\widehat{a}_i)_{i=1,...,n}$, $(\boldsymbol{x}_i)_{i=1,...,n}$ are denoted by $\left(\widehat{a}_{(i)}\right)_{i=1,...,n}$ and $\left(\boldsymbol{x}_{(i)}\right)_{i=1,...,n}$. They are such that

$$\widehat{a}_{(1)} \leq \widehat{a}_{(2)} \leq .... \leq \widehat{a}_{(n)}.$$

We define $j(\alpha)$ as the indice of the α-quantile of $\left(\widehat{a}_{(i)}\right)_{i=1,...,n}$:

$$j(\alpha) = \max \left\{ k \in \{1, ..., n\} \, : \, \frac{k}{n} \leq \alpha \right\} . \tag{9}$$

The estimate of VaR_α is the α- quantile of $\left(\widehat{a}_{(i)}\right)_{i=1,...,n}$:

$$\widehat{VaR}_\alpha = \widehat{e}_{(j(\alpha))} .$$

From Equation (4), the ES_α estimator is computed as follows:

$$\widehat{ES}_\alpha = \frac{1}{\alpha} \sum_{i=1}^{j(\alpha)-1} \frac{\widehat{e}_{(i)}}{n} + \widehat{e}_{(j(\alpha))} \left(1 - \frac{j(\alpha) - 1}{\alpha \, n} \right) .$$

A critical step in the LSMC procedure is the choice of the function $\widehat{h}(\boldsymbol{X}_t)$ that approximates the unknown conditional expectation, $h(\boldsymbol{X_t})$. This requires to test multiple candidate regressors and to carefully monitor potential overfit of the dataset. In the next section, we propose a new approach based on local regressions.

3 The local least squares Monte-Carlo (LLSMC)

As previously detailed, a common approach in LSMC consists to fit a global polynomial regression predicting responses $(y_i)_{i=1,...,n}$ as a function of risk factors $(\boldsymbol{x}_i)_{i=1,...n}$. The best model is selected by testing different sets of covariates. As alternative, we can consider machine learning regressions. The main advantage is their ability to capture non-linear dependencies between $\mathbb{E}^{\mathbb{Q}} \left(Y(t) \,|\, \boldsymbol{X}_t \right)$ and risk factors $\boldsymbol{X}_t$. Nevertheless, these methods are subject to overfitting and needs a careful tuning. On the other hand, risk management tools are audited by the regulator and must for this reason be easily explainable to authorities. Machine learning models being complex, this step requires to develop additional interpretability tools to understand the influence of risk factors on assets. In this article, we opt for an alternative approach based on local regressions. The method is based on a finite partitioning $(\mathcal{Y}_k)_{k=1,...,K}$ of the domain of Y (here $dom(Y) = \mathbb{R}$). Let us define

$$h_k(\boldsymbol{x}) \;\; = \;\; \mathbb{E}^{\mathbb{Q}} \left(Y(t) \,|\, \boldsymbol{X}_t = \boldsymbol{x} \, , \, Y(t) \in \mathcal{Y}_k \right) , k = 1, ..., K, \tag{10}$$

the conditional expectation of responses, knowing that $\boldsymbol{X}_t = \boldsymbol{x}$ and $Y(t) \in \mathcal{Y}_k$. Using standard properties of the conditional expectation, we can rewrite the function $h(\boldsymbol{x})$ as a weighted sum of $h_k(.)$:

$$\begin{aligned} h(\boldsymbol{x}) &= \mathbb{E}^{\mathbb{Q}}\left(Y(t) \,|\, \boldsymbol{X}_t = \boldsymbol{x}\right) \\ &= \sum_{k=1}^{K} \mathbb{Q}\left(Y(t) \in \mathcal{Y}_k \,|\, \boldsymbol{X}_t = \boldsymbol{x}\right) h_k(\boldsymbol{x}) \,. \end{aligned}$$

Based on this decomposition, we approximate the K unknown functions $h_k(.)$ by polynomial regressions of $Y(t) \in \mathcal{Y}_k$ on risk factors. In a second step, we use a multinomial logistic regression to estimate the probabilities $\mathbb{Q}\left(Y(t) \in \mathcal{Y}_k \,|\, \boldsymbol{X}_t\right)$ for $k = 1, ..., K$.

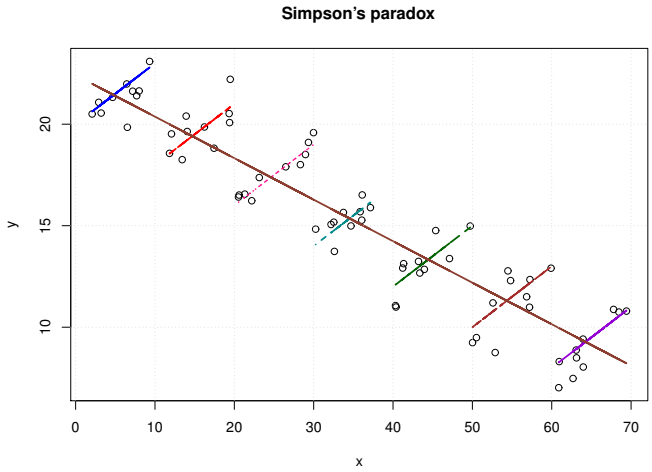

Figure 2: Illustration of the Simpson's paradox.

It may appear counterintuitive to partition the dataset using responses, $\mathcal{Y}$, instead of risk factors, $\mathcal{X}$. Two reasons motivate this choice. Firstly, local regressions based on hard clusters of risk factors generate discontinuities in predicted $\mathbb{E}^{\mathbb{Q}}\left(Y(t) \,|\, \boldsymbol{X}_t\right)$ on borders of clusters, even in a market in which all processes are continuous. This is clearly an undesirable feature for a model designed for risk management. Secondly, this prevents to observe the Simpson's paradox. This is a phenomenon in probability and statistics in which a trend appears in several groups of data but disappears or reverses when the groups are combined. This paradox is illustrated in Figure 2 which compares local versus global linear regressions. Regressions on clusters of $\boldsymbol{x}$ detect misleading local increasing trends whereas the slope of the global model is negative. We provide a financial illustration of the Simpson's paradox in the first case study.

In practice, the simulated sample, $\mathcal{S}$ defined in Equation (6), is the union of sampled risk factors, noted $\mathcal{X}$, and of corresponding responses $\mathcal{Y}$. In a first stage, we partition the sample dataset $\mathcal{S} = (\mathcal{X}, \mathcal{Y})$ into $K << n$ subsets, denoted by $(\mathcal{S}_k)_{k=1,...K}$:

$$\mathcal{S}_k \quad = \quad (\mathcal{X}_k, \mathcal{Y}_k) \,,\, k = 1, ..., K \,,$$

where $(\mathcal{Y}_k)_{k=1,...,k}$ is a partition of $\mathcal{Y}$ and $(\mathcal{X}_k)_{k=1,...,K}$ is the sample set of corresponding simulated risk factors. In this article, we use the K-means for partitioning $\mathcal{Y}$ in K clusters $(\mathcal{Y}_k)_{k=1,...,K}$. This heuristic algorithm computes a partition which reduces the within groups sum of squared errors (WGSS) or intraclass inertia. The K-means algorithm is

based on the concept of centroids that are the center of gravity of a cluster of objects. The coordinates of the u^{th} centroid is denoted $c_u \in \mathbb{R}$, $u = 1, ..., K$. If $d(.,.)$ is the Euclidian distance, we define the clusters $\mathcal{Y}_k$ for $k = 1, ..., K$ as follows:

$$\mathcal{Y}_k = \{y_i \ : \ d(y_i, c_k) \leq d(y_i, c_j) \ \forall j \in \{1, ..., K\}\} \quad k = 1, ..., K \, . \tag{11}$$

By extension, the joint cluster $\mathcal{S}_k$ of risk factors and responses is:

$$\mathcal{S}_k = \{(\boldsymbol{x}_i, y_i) \ : \ d(y_i, c_k) \leq d(y_i, c_j) \ \forall j \in \{1, ..., K\}\} \quad k = 1, ..., K \, . \tag{12}$$

The center of gravity of $\mathcal{Y}_k$, is denoted by $g_k = \frac{1}{|\mathcal{Y}_k|} \sum_{y_i \in \mathcal{Y}_k} y_i$ and the center of gravity of all responses is $g = \frac{1}{n} \sum_{i=1}^{n} g_i$. The global inertia is $I_{\mathcal{Y}} = \frac{1}{n} \sum_{i=1}^{n} d\left(y_i, g\right)^2$ and the interclass inertia I_c is the inertia of the cloud of centers of gravity:

$$I_c \ = \ \sum_{k=1}^{K} \frac{|\mathcal{Y}_k|}{n} d\left(g_k, g\right)^2 \, .$$

The intraclass inertia I_a is the sum of clusters inertiae, weighted by their size:

$$I_a \ = \ \frac{1}{n} \sum_{k=1}^{K} \sum_{y_i \in \mathcal{Y}_k} d\left(y_i, g_k\right)^2 \, .$$

According to the König-Huyghens theorem, the global inertia is the sum of the intraclass and interclass inertiae: $I_{\mathcal{Y}} = I_c + I_a$. We seek for a partition of $\mathcal{Y}$ minimizing the intraclass inertia I_a in order to have homogeneous clusters on average. This is equivalent to determine the partition maximizing the interclass inertia, I_c. Finding the partition that minimizes the intraclass inertia is computationally difficult (NP-hard) but efficient heuristic procedures exists. The most common method uses an iterative refinement technique called the K-means which is detailed in Algorithm 1, provided in appendix A. Given an initial set of K random centroids $y_1(0)$,...,$y_K(0)$, we construct a partition $\{\mathcal{Y}_1(0), \ldots, \mathcal{Y}_K(0)\}$ of the response dataset. Next, we replace the K random centroids by the K centers of gravity $(c_u(1))_{u=1:K} = (c_u(0))_{u=1:K}$ of these classes and we iterate till convergence. At each iteration, we can prove that the intraclass inertia is reduced. Nevertheless, we do not have any warranty that the partition found by this way is a global solution. In practice, this procedure is repeated several times and we keep the partition with the lowest intraclass inertia. Notice that any other partitioning procedure may be substituted to the K-means algorithm.

After having found a partition of $\mathcal{S}$ in $\mathcal{S}_k = (\mathcal{X}_k, \mathcal{Y}_k)$, $k = 1, ..., K$,we estimate functions $(h_k)_{k=1,...,K}$ by K polynomials of order d_h, denoted by $\left(\widehat{h}_k(.)\right)_{k=1,...,K}$. Let us recall that $\boldsymbol{z}$ as defined in Equation (8), is the vector of powers of risk factors up to $d_h \in \mathbb{N}$. We assume that $\widehat{h}_k(\boldsymbol{x}) = \boldsymbol{z}^\top \boldsymbol{\beta}_k$ is a polynomial of order d_h where $\boldsymbol{\beta}_k$ is a real vector of dimension equal to the one of $\boldsymbol{z}$. In a similar manner to LSMC, the $\left(\widehat{h}_k(.)\right)_{k=1,...,K}$ are estimated by least squares minimization over the set $\mathcal{P}_h$ of polynomials of degree d_h:

$$\boldsymbol{\beta}_k \ = \ \arg\min_{\widehat{h}_k \in \mathcal{P}_h} \left(\sum_{(\boldsymbol{x}_i, y_i) \in \mathcal{S}_k} \left(y_i - \widehat{h}_k(\boldsymbol{x}_i) \right)^2 \right) \, . \tag{13}$$

The sample of powers of risk factors is again $\{\boldsymbol{z}_1, ..., \boldsymbol{z}_n\}$ and we denote by Z_k and $\boldsymbol{y}_k$, the matrix $Z_k = \left(\boldsymbol{z}_j^\top\right)_{j \in \mathcal{S}_k}$ and the vector $\boldsymbol{y}_k = (y_j)_{j \in \mathcal{S}_k}$ for $k = 1, ..., K$. Using standard arguments, the polynomial coefficients minimizing (13) are $\widehat{\boldsymbol{\beta}}_k = \left(Z_k^\top Z_k\right)^{-1} Z_k^\top \boldsymbol{y}_k$.

Nevertheless, this model is useless for predicting the response for a vector $\boldsymbol{x}$ that is not in the training dataset. For $\boldsymbol{x} \notin \mathcal{S}$, the expected response should be

$$\widehat{h}(\boldsymbol{x}) = \sum_{k=1}^{K} \mathbb{Q}\left(Y(t) \in \mathcal{Y}_k \,|\, \boldsymbol{X}_t = \boldsymbol{x}\right) \widehat{h}_k(\boldsymbol{x}) \,, \tag{14}$$

where $\mathbb{Q}\left(Y(t) \in \mathcal{Y}_k \,|\, \boldsymbol{X}_t = \boldsymbol{x}\right)$ is the unknown probability that the response for $\boldsymbol{x}$ is in the k^{th} cluster. A solution consists to estimate these probabilities with a multinomial logistic regression. In this framework, we assume that conditional probabilities are the following functions

$$\mathbb{Q}\left(Y(t) \in \mathcal{Y}_k \,|\, \boldsymbol{X}_t = \boldsymbol{x}\right) \quad = \quad \begin{cases} \frac{e^{-\widehat{\boldsymbol{\gamma}}_k(\boldsymbol{x})}}{1+\sum_{j=2}^{K} e^{-\widehat{\boldsymbol{\gamma}}_j(\boldsymbol{x})}} & k = 2, ..., K \,, \\ \frac{1}{1+\sum_{j=2}^{K} e^{-\widehat{\boldsymbol{\gamma}}_j(\boldsymbol{x})}} & k = 1 \,, \end{cases} \tag{15}$$

where $\widehat{\boldsymbol{\gamma}}_k(\boldsymbol{x})$ is a polynomial of risk factors. If $\mathcal{P}_\gamma$ is the set of admissible polynomial functions of order $d_\gamma \in \mathbb{N}$, the $(\widehat{\gamma}_k(.))_{k=2,...,K}$ are estimated by log-likelihood maximization. We denote by

$$\boldsymbol{w} = \left(\left(x_{i_1}^{j_1} x_{i_2}^{j_2} ... x_{i_h}^{j_h} \right)_{i_1,...,i_h \in \{1,...,m\}}^{j_1,...,j_h \in \mathbb{N}, j_1+...+j_h \leq d_\gamma} \right)$$

the vector of powers of risk factors up to $d_\gamma \in \mathbb{N}$. We assume that $\widehat{\gamma}_k(\boldsymbol{x}) = \boldsymbol{w}^\top \boldsymbol{\zeta}_k$ is a polynomial of order d_γ where $\boldsymbol{\zeta}_k$ is a real vector of same dimension as $\boldsymbol{w}$. The log-likelihood is defined by

$$\mathcal{L}\left((\gamma_k)_{k=2,...,K}\right) \quad = \quad \sum_{i=1}^{n} \log\left(\sum_{k=2}^{K} \frac{\mathbf{1}_{\{y_i \in \mathcal{Y}_k\}} e^{-\boldsymbol{\gamma}_k(\boldsymbol{x}_i)}}{1+\sum_{j=2}^{K} e^{-\boldsymbol{\gamma}_j(\boldsymbol{x}_i)}} + \frac{\mathbf{1}_{\{y_i \in \mathcal{Y}_1\}}}{1+\sum_{j=2}^{K} e^{-\boldsymbol{\gamma}_j(\boldsymbol{x})}} \right) ,$$

and $(\boldsymbol{\zeta}_k)_{k=2,...,K} = \arg\max_{\gamma_k \in \mathcal{P}_\gamma} \mathcal{L}\left((\gamma_k)_{k=2,...,K}\right)$.

We haven't discussed yet how to optimize the number of clusters K and the polynomial orders, d_h, d_γ. In practice, we check four indicators of goodness of fit. We first compare the R^2 for different settings. The R^2 is the percentage of variance of responses explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \widehat{h}(\boldsymbol{x}_i)\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2} \,, \tag{16}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. In LSMC regression, responses y, are by construction very noised estimates of $\mathbb{E}^{\mathbb{Q}}\left(Y(t) \,|\, \boldsymbol{x}\right)$ and therefore the R^2 is by nature small. We assess the fit of local regressions by

$$R^2_{loc} = 1 - \frac{\sum_{k=1}^{K} \sum_{(\boldsymbol{x}_i, y_i) \in \mathcal{S}_k} \left(y_i - \widehat{h}_k(\boldsymbol{x}_i)\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2} \,. \tag{17}$$

Contrary to R^2, we may expect a R^2_{loc} close to 1 and should exclude any models with a low R^2_{loc}. The R^2 and R^2_{loc} both increase with the complexity of the model, measured by the number of its parameters. For this reason, we also compute a second indicator of goodness of fit which is the mean squared error of residuals:

$$MSE = \frac{\sum_{i=1}^{n} \left(y_i - \widehat{h}(\boldsymbol{x})\right)^2}{n-p} \,, \tag{18}$$

where p is the number of regression parameters. This criterion tends to penalize models with a large number of parameters. To detect abnormal prices, we also calculate the sum of squared errors between exact values of $A(t, \boldsymbol{x})$ and their LLSMC estimates, $h(\boldsymbol{x})$ over a small sample of risk factors. We call this sample the validation set and denote it by $\mathcal{V}$. Depending upon the nature of assets, the exact values of $A(t, \boldsymbol{x}_t)$ is computed by performing a sufficient number of secondary simulations or by any other suitable numerical method. This step being computationally intensive, the size of the validation set, must be limited but should contain sufficiently diversified combinations of risk factors. A simple approach consists to combine quantiles of risk factors. Let us detail this approach. We denote by $\left(x_{(i)}^{(k)}\right)_{i=1,...,n}$, the ordered sample $\left(x_i^{(k)}\right)_{i=1,...,n}$, of the k^{th} risk factor:

$$x_{(1)}^{(k)} \leq x_{(2)}^{(k)} \leq ... \leq x_{(n)}^{(k)}.$$

We select a small number of $q \in \mathbb{N}$ quantiles $\left(x_{j(\alpha_1)}^{(k)}, ..., x_{j(\alpha_q)}^{(k)}\right)$ where $(\alpha_i)_{i=1,...,q}$ are probabilities and $j(\alpha_i)$ is the quantile index such as defined in Equation (9). We repeat this operation for each $k = 1, ..., m$. The validation set $\mathcal{V}$ contains all the combination of quantiles and its total size is $|\mathcal{V}| = q^m$. The MSE on the validation sample is

$$MSE(\mathcal{V}) = \frac{1}{|\mathcal{V}|} \sum_{\boldsymbol{x} \in \mathcal{V}} \left(A(t, \boldsymbol{x}) - \widehat{h}(\boldsymbol{x})\right)^2. \tag{19}$$

If the dimension of $|\mathcal{V}|$ is too large, we can select randomly a subset of $\mathcal{V}$ of appropriate size. Besides the analysis of these indicators of goodness of fit, it is recommanded to plot the function $\widehat{h}(\boldsymbol{x})$ in order to detect unexpected tail behaviour. This point is illustrated in the following sections.

4 Application to options management in the Heston model

In order to compare the LSMC method to its local version, we first consider a financial market made up of one stock with stochastic volatility and a cash account. We choose this market model, proposed by Heston (1993), because we can benchmark LSMC and LLSMC predictions to accurate option prices. computed by discrete Fourier transform, as briefly reviewed in the next subsection.

4.1 Heston model in a nutshell

We consider a financial market made up of two assets. The account earns a constant risk free rate r. The stock price, noted $(S_t)_{t\geq 0}$, is ruled by a geometric Brownian diffusion with a stochastic variance, $(V_t)_{t\geq 0}$:

$$\begin{cases} dS_t &= \mu\, S_t\, dt + S_t \sqrt{V_t} \left(\rho dW_t^v + \sqrt{1-\rho^2} dW_t^s\right), \\ dV_t &= \kappa\,(\gamma - V_t)\, dt + \sigma \sqrt{V_t} dW_t^v. \end{cases} \tag{20}$$

where $(W_t^s)_{t\geq 0}$ and $(W_t^v)_{t\geq 0}$ are independent Brownian motion defined on the real probability space $(\Omega, \mathcal{F}, \mathbb{P})$. $\mu \in \mathbb{R}$ is the expected instantaneous stock return and $\rho \in (-1, 1)$ is the correlation between the price and volatility. The variance reverts with a speed $\kappa > 0$ to a mean reversion level $\gamma > 0$. The volatility of the variance is a multiple $\sigma \in \mathbb{R}^+$ of the square root of variance.

For the sake of simplicity, we assume that the variance has the same dynamics under

$\mathbb{P}$ and $\mathbb{Q}$ (this assumption may be relaxed without any impact on our analysis) whereas the drift of the stock price is replaced by the risk free rate under $\mathbb{Q}$. As stated in the next proposition, the characteristic function of the log-return admits a closed-form expression.

Proposition 4.1. *The characteristic function of* $\ln(S_s/S_0)\,|\mathcal{F}_t$ *under the risk neutral* $\mathbb{Q}$, *for* $s \geq t$ *with* $\omega \in \mathbb{C}$, *is given by the following expression*

$$\mathbb{E}^{\mathbb{Q}}\left(e^{\omega \ln(S_s/S_0)}\,|\,\mathcal{F}_t\right) \;=\; \left(\frac{S_t}{S_0}\right)^{\omega} \exp\left(A(\omega,t,s) + B(\omega,t,s)V_t\right). \tag{21}$$

Let us define the following constants:

$$\begin{cases} d = \sqrt{(\rho\sigma\omega - \kappa)^2 + \sigma^2(\omega - \omega^2)}\,, \\ g = \frac{\kappa - \rho\sigma\omega + d}{\kappa - \rho\sigma\omega - d}\,. \end{cases}$$

The functions $A(\omega,t,s)$ *and* $B(\omega,t,s)$ *in Equation (21) are given by*

$$\begin{aligned} A(\omega,t,s) &= r\,\omega\,(s-t) + \\ &\frac{\kappa\gamma}{\sigma^2}\left((\kappa - \rho\sigma\omega + d)(s-t) - 2\ln\left(\frac{1 - g e^{d(s-t)}}{1-g}\right)\right), \end{aligned} \tag{22}$$

and

$$B(\omega,t,s) \;=\; \frac{\kappa - \rho\sigma\omega + d}{\sigma^2}\,\frac{1 - e^{d\,(s-t)}}{1 - g\,e^{d\,(s-t)}}\,. \tag{23}$$

For a proof, the reader can refer for instance, to Hainaut (2022), chapter 3, p. 65. European call or put options do not have analytical expressions. In order to evaluate these options, we can calculate numerically the probability density function of the log-return, $\ln(S_T/S_0)\,|\mathcal{F}_t$, by a discrete Fourier transform (DFT). The characteristic function of a random variable, here $\Upsilon_{t,T}(i\omega) = \mathbb{E}^{\mathbb{Q}}\left(e^{i\,\omega \ln(S_T/S_0)}\,|\,\mathcal{F}_t\right)$, is also the inverse Fourier transform of its probability density function (pdf):

$$\begin{aligned} f_{t,T}(u) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Upsilon_{t,T}(i\omega)\,e^{-i\,u\,\omega} d\omega \\ &= \frac{1}{\pi} Re\left(\int_0^{+\infty} \Upsilon_{t,T}(i\omega) e^{-i\,u\,\omega} d\omega\right) \end{aligned} \tag{24}$$

Therefore, we can retrieve the pdf by computing numerically its Fourier transform as stated in the next proposition.

Proposition 4.2. *Let* M *be the number of steps used in the Discrete Fourier Transform (DFT) and* $\Delta_u = \frac{2u_{max}}{M-1}$ *be this step of discretization. Let us denote* $\Delta_\omega = \frac{2\pi}{M\,\Delta_u}$ *and*

$$\omega_m \;=\; (m-1)\Delta_\omega,$$

for $m = 1...M$. *Let* $\Upsilon_{t,T}(\omega) = \mathbb{E}^{\mathbb{Q}}\left(e^{\omega \ln(S_T/S_0)}\,|\,\mathcal{F}_t\right)$ *be mgf of* $\ln(S_T/S_0)$. *The values of* $f_{t,T}(.)$ *the pdf of* $\ln(S_T/S_0)\,|\mathcal{F}_t$ *at points* $u_k = -\frac{M}{2}\Delta_u + (k-1)\Delta_u$ *are approached by the sum:*

$$f(u_k) \;\approx\; \frac{2}{M\,\Delta_u} Re\left(\sum_{m=1}^{M} \varrho_m \Upsilon_{t,T}(i\,\omega_m)(-1)^{m-1} e^{-i\frac{2\pi}{M}(m-1)(k-1)}\right). \tag{25}$$

where $\varrho_m = \frac{1}{2}1_{\{m=1\}} + 1_{\{m\neq 1\}}$.

This result is proven by discretizing the integral (24). The value of a European option of maturity T and payoff $H(S_T)$ is then approached by the following sum

$$\mathbb{E}^{\mathbb{Q}}\left(e^{-r(T-t)}H(S_T)|\mathcal{F}_t\right) \quad = \quad \sum_{k=1}^{M} f\left(u_k\right) H(S_0 e^{u_k})\,. \tag{26}$$

Prices obtained by this method are compared to LSMC and LLSMC prices in the next subsection.

4.2 LLSMC applied to butterfly options

In order to apply the LSMC to the Heston model, we consider as risk factors, the normed stock price and volatility:

$$\boldsymbol{X}_t := \left(\frac{S_t - \mathbb{E}_0^{\mathbb{P}}\left(S_t\right)}{\sqrt{\mathbb{V}_0^{\mathbb{P}}\left(S_t\right)}}, \frac{\sqrt{V_t} - \mathbb{E}_0^{\mathbb{P}}\left(\sqrt{V_t}\right)}{\sqrt{\mathbb{V}_0^{\mathbb{P}}\left(\sqrt{V_t}\right)}}\right).$$

In practice, expectations and variances of S_t and $\sqrt{V_t}$ are estimated by empirical averages and variances of the simulated sample. We consider a European butterfly option of maturity T and strikes E_1, E_2 and E_3. The payoff of this option is

$$H(S_T) = \left(S_T - E_1\right)_+ - 2\left(S_T - E_2\right)_+ + \left(S_T - E_3\right)_+,$$

and its price $A(t, \boldsymbol{X}_t)$, at time $t \leq T$, is equal to the $\mathbb{Q}-$expected discounted payoff, $A(t, \boldsymbol{X}_t) = \mathbb{E}^{\mathbb{Q}}\left(e^{-r(T-t)}H(S_T)\,|\,\mathcal{F}_t\right)$. We choose this derivative because its payoff presents three inflection points and is not an invertible function with respect to stock prices. As we will see, the price of such an option is difficult to replicate by LSMC. We will next consider a bull trap option that has a increasing payoff. Table 1 reports model and payoff parameters. The Heston model is fitted to the time series of the S&P 500 from 31/1/2001 to 31/1/2020 by Bayesian log-likelihood maximization (for details on the estimation procedure, see Hainaut 2022, p. 75). In a first stage, we perform 10000 primary simulations under $\mathbb{P}$ of responses $Y(t) = e^{-r(T-t)}H(S_T)$, with 350 steps of time per year. For each primary simulation, we simulate a single secondary sample path under $\mathbb{Q}$.

Parameters			
μ	0.1232	r	0.02
κ	0.7171	ρ	-0.5390
γ	0.1016	σ	0.4234
S_0	100	E_1	100
E_2	108	E_3	116
t	1 year	T	2 years

Table 1: Parameters of the Heston stochastic volatility model and of the payoff.

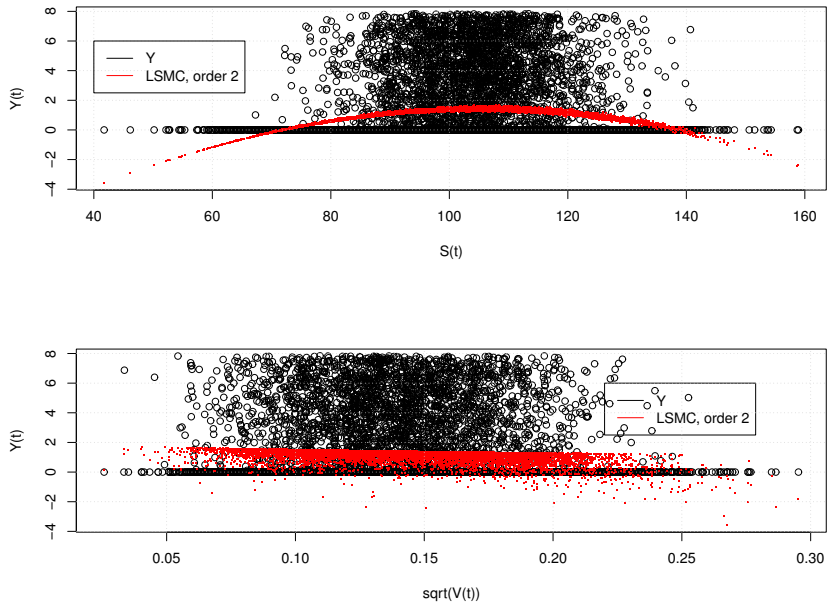

Figure 3: Simulated responses $Y(t) = e^{-r(T-t)}H(S_T)$ versus stock prices S_t, and volatilities, $\sqrt{V_t}$. The red curves are the predictions $h(\boldsymbol{X}_t)$ from the LSMC with a second order polynomials regression.

The upper and lower plots of Figure 3 show simulated responses versus stock prices and volatilities. The red curves correspond to LSMC estimates of butterfly option prices in one year with a second order polynomial of risk factors. These graphs reveal a weakness of the classical LSMC approach: it fails to predict a positive option price for extreme high and low values of stock prices. This point is particularly critical when LSMC is used for computing risk measures.

Table 2 reports the R^2, the MSE and MSE($\mathcal{V}$) of the LSMC, such as defined by Equations (16), (18) and (19). The validation set counts 100 pairs of risk factors. We consider $q = 10$ empirical quantiles of stock prices and volatilities for probabilities from 1% to 5% and from 95% to 99% by step of 1%. This choice is motivated by the fact that extreme values of risk factors are likely to generate extreme high and low option prices. The (nearly) exact prices of butterfly options in these 100 scenarios are computed by discrete Fourier transform (DFT) that provides the density $f_{t,T}(u)$ of $\ln(S_T/S_0)\,|\mathcal{F}_t$ on a grid $\{u_1, ..., u_M\}$ of size M for given risk factors at time t. The value of the butterfly option is next computed by formula (26).

d_h	R^2	$\sqrt{\text{MSE}(\mathcal{V})}$	$\sqrt{\text{MSE}}$	d.f.
2	0.0397	0.36	2.10	6
3	0.0451	0.57	2.10	10
4	0.0499	1.07	2.09	15
5	0.0522	2.39	2.09	21
6	0.0536	1.93	2.09	28

Table 2: R^2, MSE and MSE($\mathcal{V}$) of regressions of Y_t on $\boldsymbol{X}_t$ in the LSMC model. d.f. is the number of parameters.

K	d_γ	d_h	R^2	$\sqrt{\text{MSE}(\mathcal{V})}$	$\sqrt{\text{MSE}}$	d.f.	R^2_{loc}
3	2	3	0.0526	0.19	2.09	42	0.9513
5	2	3	0.0527	0.2	2.1	74	0.9850
6	2	4	0.0523	0.2	2.1	120	0.9900
2	2	3	0.0525	0.2	2.09	26	0.8693
5	2	4	0.0527	0.21	2.1	99	0.9851
5	2	2	0.0525	0.21	2.1	54	0.9849
3	2	1	0.0523	0.21	2.09	21	0.9511
4	2	2	0.0525	0.21	2.09	42	0.9755
4	2	1	0.0525	0.21	2.09	30	0.9754
2	2	2	0.0521	0.21	2.09	18	0.8691

Table 3: R^2, MSE, MSE($\mathcal{V}$) and R^2_{loc} for the LLSMC model. d.f. is the number of parameters.

In Table 2, butterfly prices are approached by polynomial regressions of order d_h from 2 to 6. As expected, R^2's are tiny since responses are noised estimates of $\mathbb{E}^{\mathbb{Q}}(Y(t) \,|\, \boldsymbol{x})$. The R^2's also increase with the complexity of the model. The MSE on the training set is inversely proportional to the polynomial order whereas the lowest MSE($\mathcal{V}$) on the validation set is achieved with an order 2 polynomial. A next step consists in analyzing the tail behaviour of these approximations. This is done by plotting the function $\widehat{h}(\boldsymbol{x})$. We will come back on this point later and before focus on the LLSMC.

Table 3 presents the statistics of goodness of fit for the LLSMC model. The number of clusters, K, varies from 2 to 6. We test polynomials of degrees d_h from 2 to 6 and d_γ equal to 1 and 2. Models are sorted by increasing MSE($\mathcal{V}$)'s and we report statistics of the 10 first best models according to this criterion. The best goodness of fit is achieved with 2 or 3 clusters, a cubic regression on each cluster and second order polynomial for the multinomial logistic regression. A comparison with LSMC figures of Table 2, reveals that the LLSMC reduces by half the MSE($\mathcal{V}$) on the validation dataset whereas MSE on the training set are comparable. This is a good indicator that the LLSMC model better replicates extremely high and low prices. We also notice that the LLSMC- and LSMC-R^2 are comparable.

We compare now the LSMC and the LLSMC with hyperparameters $K = 3$, $d_\gamma = 2$, $d_h = 3$ (denoted by LLSMC 3-2-3) as this setting leads to a low MSE($\mathcal{V}$) and a high R^2_{loc}. The upper plot of Figure 4 shows the segmentation of responses in 3 clusters with the K-means algorithm. The mid and lower plots shows the responses and local predictions $\widehat{h}_k(\boldsymbol{X}_t)$ (red lines) on clusters, with respect to stock prices and volatilities. Contrary to the LSMC, the local regressions do not yield large negative responses.

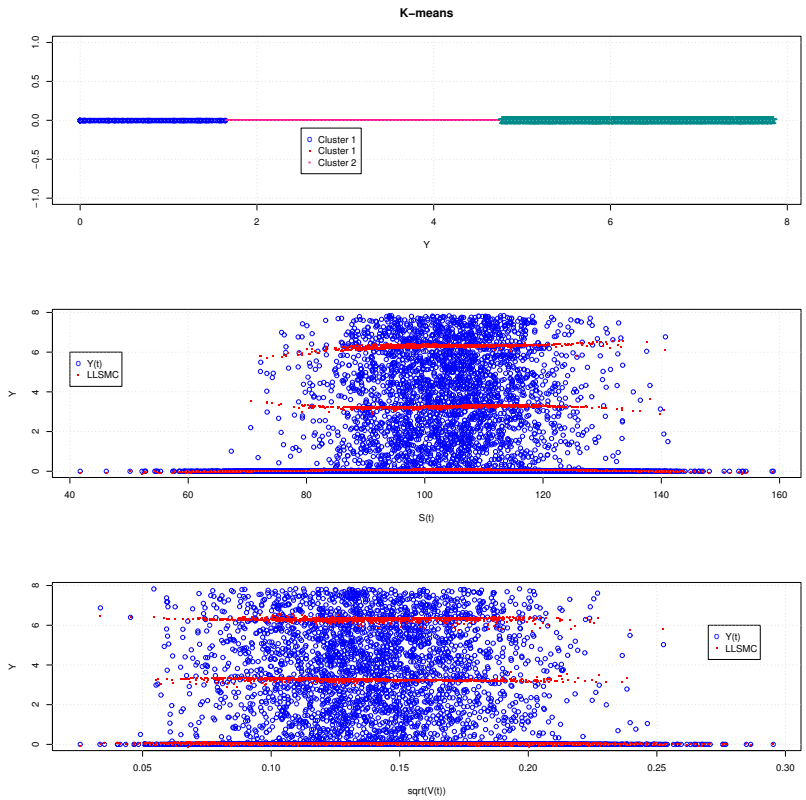

Figure 4: Upper plot: a-priori segmentation of responses in 3 clusters. Mid and lower plots: responses (blue points) and local regressions (red points) with respect to stock prices and volatilities.

Figure 5 compares LSMC and LLSMC butterfly options for stock prices S_t ranging from 68 to 139, the 1% and 99% percentiles of simulated stock prices and $\sqrt{V_t} \in \{7\%, 14\%, 23\%\}$, the 1%, 50% and 99% quantiles of simulated volatilities. Exact option prices are computed by DFT with $u_{max} = 2$ and $M = 2^8$ steps of discretization. The mid plot displays prices in standard market conditions. The right and left plots correspond to extreme volatility conditions. We observe that LSMC models of order 2 or 4 generates negative prices in the tails. To measure the overall accuracy of methods in these three scenarios of volatility, we report in Table 4 the average pricing errors. This table confirms that the LLSMC globally outperforms LSMC. Tables 5 and 6 present the VaR's and TVaR's of the butterfly option for various quantiles. The LSMC models yield negative values for the lowest percentiles whereas the LLSMC provides slightly lower VaR's and TVaR's than the LSMC of orders 3 to 6, for the highest percentiles. This would be interesting to compare these results to VaR's and TVaR's based on exact prices computed by DFT. Unfortunately, the valuation by DFT of 10000 butterfly options is computationally too intensive. This comparison is nevertheless possible in the second case study (Section 5).

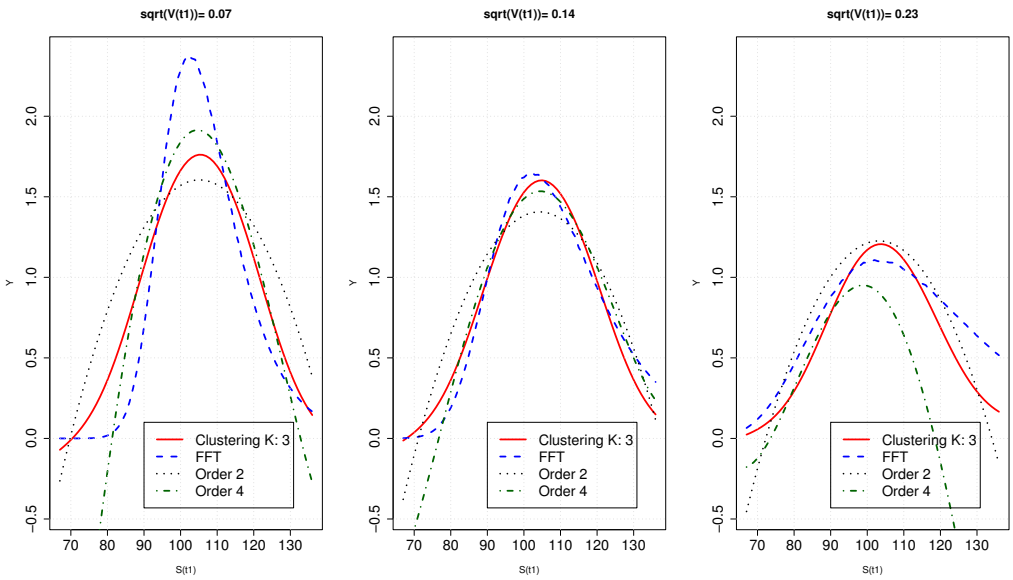

Figure 5: Butterfly option for S_t ranging from 40 to 180 and volatilities $\sqrt{V_t} \in \{7\%, 14\%, 23\%\}$.

	$\sqrt{V_t} = 7\%$	$\sqrt{V_t} = 14\%$	$\sqrt{V_t} = 23\%$
LLSMC, 3-2-3	0.32	0.11	0.18
LSMC, order 2	0.54	0.23	0.22
LSMC, order 3	0.61	0.32	0.33
LSMC, order 4	0.75	0.21	0.93
LSMC, order 5	2.34	0.1	1.32
LSMC, order 6	1.09	0.07	1.87

Table 4: Average pricing errors for the three cases presented in Figure 5.

	0.05%	0.1%	1%	5%	95%	99%	99.9%	99.95%
LLSMC, 3-2-3	-0.08	-0.06	0.01	0.16	1.66	1.71	1.76	1.76
LSMC, d_h =2	-2.20	-1.92	-0.58	0.22	1.47	1.54	1.60	1.61
LSMC, d_h =3	-2.00	-1.51	-0.57	0.10	1.56	1.70	1.86	1.90
LSMC, d_h =4	-0.85	-0.78	-0.33	0.01	1.61	1.75	1.93	2.01
LSMC, d_h =5	-0.60	-0.53	-0.24	0.00	1.66	1.83	2.02	2.07
LSMC, d_h =6	-0.15	-0.11	-0.02	0.17	1.71	1.87	2.07	2.13

Table 5: VaR 1 year, LSMC model and LLSMC.

	0.05%	0.1%	1%	5%	95%	99%	99.9%	99.95%
LLSMC, 3-2-3	-0.10	-0.09	-0.03	0.06	1.69	1.73	1.77	1.77
LSMC, d_h =2	-2.87	-2.51	-1.18	-0.29	1.51	1.57	1.62	1.64
LSMC, d_h =3	-2.60	-2.18	-1.05	-0.35	1.64	1.78	1.95	2.02
LSMC, d_h =4	-1.08	-0.95	-0.5	-0.23	1.7	1.85	2.17	2.37
LSMC, d_h =5	-0.71	-0.64	-0.35	-0.16	1.76	1.92	2.1	2.17
LSMC, d_h =6	-0.77	-0.45	-0.08	0.05	1.81	1.97	2.22	2.33

Table 6: Expected shortfall, 1 year, LSMC model and LLSMC.

We now detail how the LLSMC 3-2-3 operates with the help of Figure 6. The left plot shows the local regression functions $\widehat{h}_k(x)$, for various stock prices S_t and a volatility of 14%. The mid plot displays the probabilities that a pair of risk factors leads to a response belonging to the k^{th} cluster. We clearly see that the first cluster explains left and right tails of butterfly option prices. If S_t is below 80 or above 130, the response is in cluster 1 with a probability higher than 90% and the correspond function $\widehat{h}_3(\boldsymbol{x})$ is nearly flat and null. The probabilities of belonging to clusters 2 and 3 are quite similar and higher than 5% for $S_t \in [80, 130]$. The right plot shows the products of regression and probabilities functions. According to Equation (14), the estimated option price is the sum of these three terms.

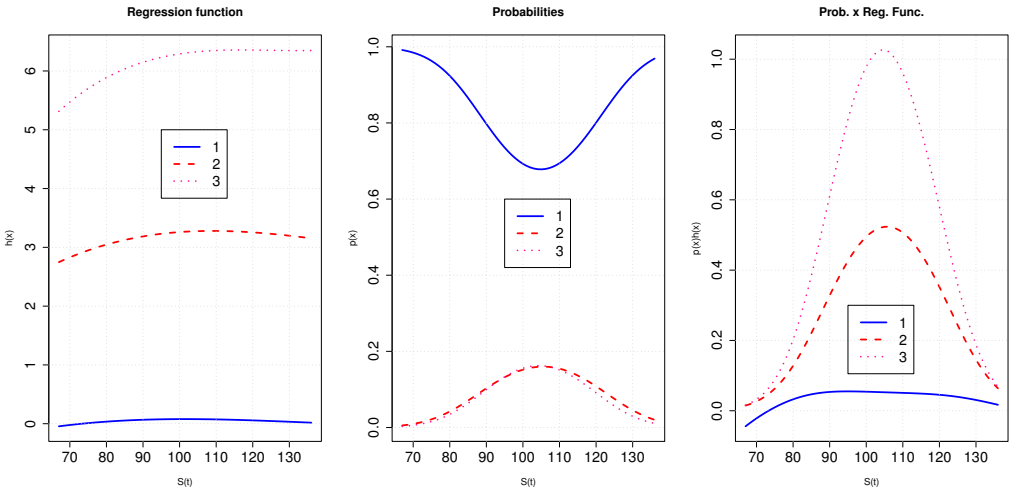

Figure 6: Plots of regression functions $h_k(\boldsymbol{x})$ for $k = 1, ..., K$, probabilities $\mathbb{Q}\left(Y(t) \in \mathcal{Y}_k \,|\, \boldsymbol{X}_t = \boldsymbol{x}\right)$ and their products, for $\sqrt{V_t} = 14\%$.

	Cluster, k			std. errors		
	1	2	3	1	2	3
Intercept	0.0765	3.2602	6.3226	0.0042	0.0408	0.038
$x^{(1)}$	-0.0015	0.0371	0.1091	0.0059	0.0688	0.0729
$x^{(2)}$	-0.0135	-0.0579	0.1003	0.0062	0.0633	0.0581
$\left(x^{(1)}\right)^2$	-0.0163	-0.0465	-0.0997	0.0032	0.0634	0.0584
$\left(x^{(2)}\right)^2$	-0.0007	-0.0287	-0.0063	0.0034	0.0364	0.0362
$x^{(1)}x^{(2)}$	-0.0009	-0.0871	-0.0629	0.0054	0.0732	0.0681
$\left(x^{(1)}\right)^3$	0.0026	0.0063	0.0192	0.0019	0.0424	0.0408
$\left(x^{(2)}\right)^3$	0.0006	0.0145	-0.0275	0.0025	0.0291	0.0269
$\left(x^{(1)}\right)^2 x^{(2)}$	0.0057	0.0874	-0.0897	0.0047	0.0604	0.0721
$x^{(1)}\left(x^{(2)}\right)^2$	0.0008	0.048	-0.0808	0.0052	0.0603	0.0655

Table 7: LLSMC, 3-2-3 : parameters of cubic regressions, $h_k(\boldsymbol{x})$ per cluster and standard errors (std. err.)

	Cluster, k		std. errors	
	2	3	2	3
Intercept	-1.4843	-1.4643	0.0467	0.0471
$x^{(1)}$	0.2552	0.2226	0.0525	0.0547
$x^{(2)}$	-0.0882	-0.1144	0.0428	0.043
$\left(x^{(1)}\right)^2$	-0.5778	-0.6956	0.0565	0.0622
$\left(x^{(2)}\right)^2$	-0.0197	-0.0128	0.039	0.0392
$x^{(1)}x^{(2)}$	0.0409	-0.0338	0.0717	0.076

Table 8: LLSMC, 3-2-3 : Parameters of logistic regressions, $\gamma_k(\boldsymbol{x})$ per cluster and standard errors (std. err.)

Tables 7 and 8 report the estimated coefficients of polynomials $\widehat{h}_k(.)$ and $\widehat{\gamma}_k(.)$ of the LLSMC, 3-2-3. Some standard errors seem too high at a first sight. As for low R^2, this is explained by the fact that responses are by construction very noised estimates of $\mathbb{E}^{\mathbb{Q}}\left(Y(t)\,|\,\boldsymbol{x}\right)$. Therefore, these statistics should be analyzed with a certain care.

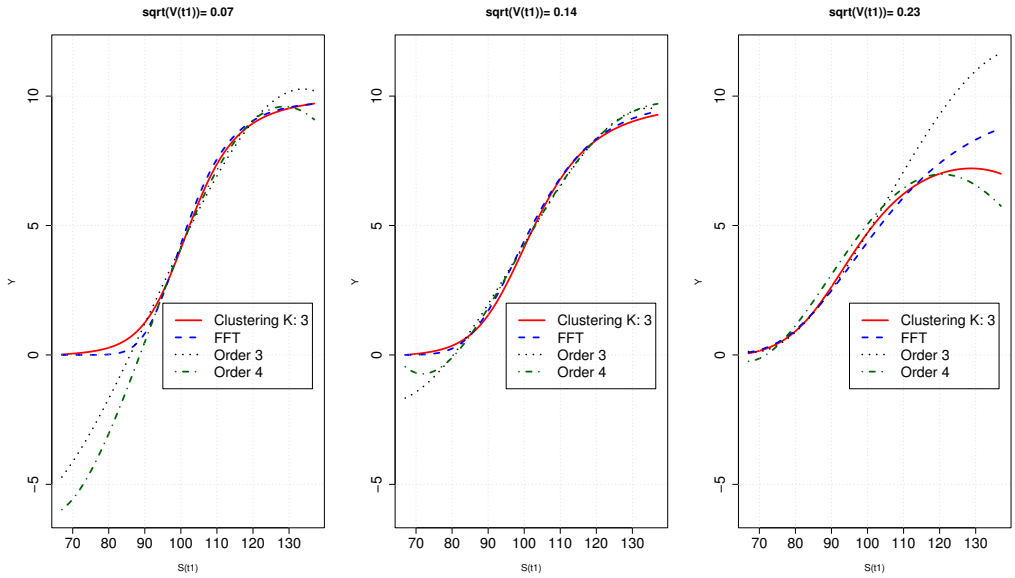

Figure 7: Bull trap option for S_t ranging from 40 to 180 and volatilities $\sqrt{V_t} \in \{7\%, 14\%, 23\%\}$.

Till now, we have focused on a butterfly option because its payoff and therefore its price is not strictly increasing or decreasing function of S_T. We now show that the LLSMC still outperforms the LSMC for increasing payoffs. We consider a long and a short position in call options of maturity T and strikes E_1, E_2. The total payoff of this bull trap option is equal to

$$H(S_T) = (S_T - E_1)_+ - (S_T - E_2)_+ \ .$$

We use again the Heston model with parameters of Table 1, except that we set strikes to $E_1 = 100$, $E_2 = 110$. We compare the LSMC and the LLSMC with hyperparameters $K = 3$,$d_\gamma = 2$, $d_h = 1$. Figure 7 compares LSMC and LLSMC bull trap option for stock prices S_t ranging from 68 to 139, and $\sqrt{V_t} \in \{7\%, 14\%, 23\%\}$. Table 9 reports the average pricing errors in these 3 scenarios. These results confirm that LLSMC achieves a better overall accuracy than the LSMC.

	$\sqrt{V_t} = 7\%$	$\sqrt{V_t} = 14\%$	$\sqrt{V_t} = 23\%$
LLSMC, 3-2-1	0.21	0.13	0.58
LSMC, order 2	1.59	0.94	0.53
LSMC, order 3	1.59	0.53	1.40
LSMC, order 4	2.24	0.32	0.98
LSMC, order 5	2.54	0.18	3.38
LSMC, order 6	1.16	0.25	3.00

Table 9: Average pricing errors, bull trap portfolio.

To conclude this section, we illustrate the Simpson's paradox when we price a butterfly option with parameters of Table 1. For this purpose, we divide the sample $\mathcal{S}$ into $K << n$ subsets $(\mathcal{S}_k)_{k=1,...K}$:

$$\mathcal{S}_k \quad = \quad (\mathcal{X}_k, \mathcal{Y}_k)\,,\, k = 1, ..., K\,,$$

where partition is this time, based on a partitioning of risk factors. Each cluster is defined by a centroid $\boldsymbol{c}_k \in \mathbb{R}^m$ of dimension m such that

$$\mathcal{S}_k = \{(\boldsymbol{x}_i, y_i) \,:\, d(\boldsymbol{x}_i, \boldsymbol{c}_k) \leq d(\boldsymbol{x}_i, \boldsymbol{c}_j)\ \forall j \in \{1, ..., K\}\} \quad k = 1, ..., K\,.$$

We use the K-means algorithm to find the partition of $\mathcal{S}$ in $\mathcal{S}_k = (\mathcal{X}_k, \mathcal{Y}_k)\,,\, k = 1, ..., K$. The conditional expectation of responses is approached by a piecewise function

$$\widehat{h}(\boldsymbol{x}) = \sum_{k=1}^{K} \mathbf{1}_{\{\boldsymbol{x} \in \mathcal{S}_k\}} \widehat{h}_k(\boldsymbol{x})\,,$$

where $\widehat{h}_k \in \mathcal{P}_h$, the set of polynomials of order d_h. As for a LLSMC regression, the $\widehat{h}_k$ are estimated by least squares minimization, as in Equation (7). This variant of local model is denoted by $\mathcal{X}$-LLSMC.

K	d_h	R^2	$\sqrt{\text{MSE}(\mathcal{V})}$	$\sqrt{\text{MSE}}$	d.f.
4	1	0.0501	0.27	2.09	12
5	1	0.0523	0.39	2.09	15
6	1	0.052	0.46	2.09	18
4	2	0.0524	0.55	2.09	24
5	2	0.0549	0.63	2.09	30
4	3	0.0546	0.64	2.09	40
2	2	0.0479	0.73	2.1	12
3	1	0.0429	0.75	2.1	9
6	2	0.0558	0.94	2.09	36
3	4	0.0565	0.97	2.09	45

Table 10: R^2, MSE, MSE($\mathcal{V}$) and R^2_{loc} for the $\mathcal{X}$-LLSMC model. d.f. is the number of parameters.

Table 10 reports statistics of goodness of fit for models with K from 2 to 6 and d_h from 1 to 4. Models are sorted by increasing MSE($\mathcal{V}$)'s and we only report statistics of the 10 best models according to this criterion. In view of LLSMC figures of Table 3, the $\mathcal{X}$-LLSMC 4-1 achieves a similar accuracy with even less parameters. If we limit our analysis to compare statistics of goodness of fit, the $\mathcal{X}$-LLSMC and LLSMC seems both eligible for computing VaR or TVaR. Plotting the $\mathcal{X}$-LLSMC regression function leads to

another conclusion. Figure 7 compares $\mathcal{X}$-LLSMC 4-1 and FFT Butterfly prices for S_t ranging from 68 to 139, and $\sqrt{V_t} \in \{7\%, 14\%, 23\%\}$. We observe that local regressions based on clusters of risk factors generate discontinuities in predicted responses on borders of clusters. Secondly, we identify local trends not relevant with the global slope of price curves. These two elements disqualify the $\mathcal{X}$-LLSMC for risk management purposes.

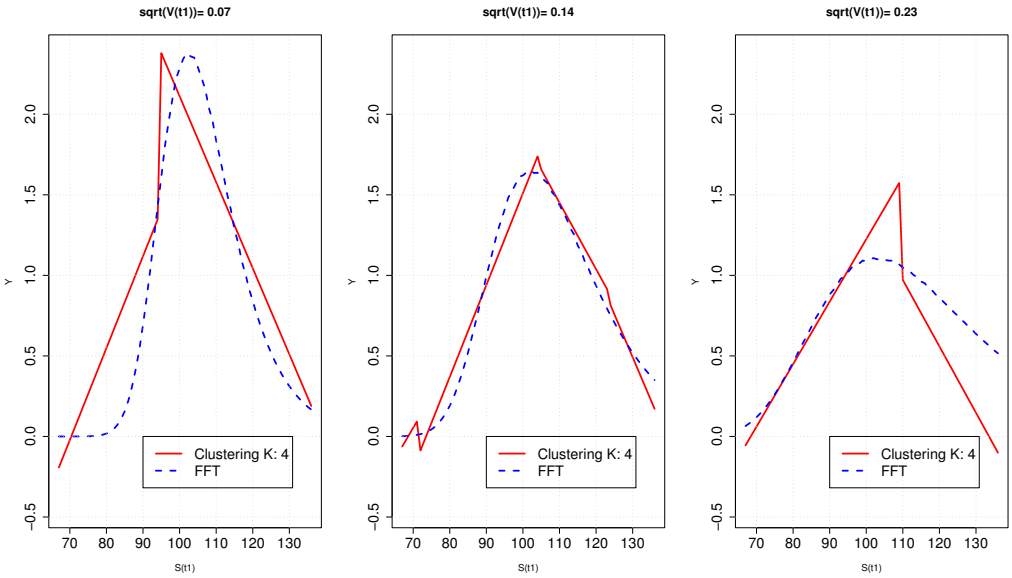

Figure 8: Butterfly option for S_t ranging from 40 to 180 and volatilities $\sqrt{V_t} \in \{7\%, 14\%, 23\%\}$.

5 Application to life insurance management

In this second example, we compare the performance of LSMC and LLSMC for assessing the risk of a participating pure endowment. The next subsection briefly presents the features of this product in a market with 3 risk factors. As the contract admits a closed-form valuation formula, we will compare approached and exact VaR's and TVaR's.

5.1 A participating pure endowment

We consider a combined life insurance and financial market. The stock price indice, the interest rate and the force of mortality are respectively denoted by $(S_t)_{t\geq 0}$, $(r_t)_{t\geq 0}$ and $(\mu_{x+t})_{t\geq 0}$. These processes are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ by the following dynamics:

$$\begin{pmatrix} dS_t \\ dr_t \\ d\mu_{x+t} \end{pmatrix} = \begin{pmatrix} \mu\, S_t \\ \kappa_r\left(\gamma_r(t) - r_t\right) \\ \kappa_\mu\left(\gamma_x(t) - \mu_{x+t}\right) \end{pmatrix} dt + \begin{pmatrix} S_t\sigma_S & 0 & 0 \\ 0 & \sigma_r & 0 \\ 0 & 0 & \sigma_x(t) \end{pmatrix} \Sigma \begin{pmatrix} dW_t^{(1)} \\ dW_t^{(2)} \\ dW_t^{(3)} \end{pmatrix} \quad (27)$$

$W_t^{(1)}, W_t^{(2)}, W_t^{(3)}$ are independant Brownian motions. μ, κ_r, κ_μ, σ_S and σ_r belong to $\mathbb{R}^+$ whereas $\gamma_r(t)$, $\gamma_x(t)$ and $\sigma_x(t)$ are positive functions of time. $\gamma_r(t)$ and $\gamma_x(t)$ are respectively fitted to term structures of interest and mortality rates. The initial age of the insured is denoted by $x \in [0, x_{max})$. Furthermore, we assume that the standard deviation of mortality is related to age through the relation $\sigma_x(t) = \alpha e^{\beta(x+t)}$. Details about $\gamma_x(t)$ and $\sigma_x(t)$ are provided in Appendix D. The matrix Σ is the (upper) Choleski decomposition of

the correlation matrix and is such that

$$\Sigma = \begin{pmatrix} \epsilon_{SS} & \epsilon_{Sr} & \epsilon_{S\mu} \\ 0 & \epsilon_{rr} & \epsilon_{r\mu} \\ 0 & 0 & 1 \end{pmatrix} \quad , \quad \begin{pmatrix} 1 & \rho_{Sr} & \rho_{S\mu} \\ \rho_{Sr} & 1 & \rho_{r\mu} \\ \rho_{S\mu} & \rho_{r\mu} & 1 \end{pmatrix} = \Sigma\Sigma^{\top},$$

where $\rho_{Sr}, \rho_{S\mu}$, ρ_{Sr} and $\rho_{r\mu} \in (-1, 1)$. This model allows for correlation between financial and mortality shocks caused for instance by a pandemic such as the of one of Covid 19. Notice that Ha and Bauer (2019) use a similar framework, excepted that in our framework the mortality is mean reverting, with an age-dependent volatility. Another difference is that we consider a participating pure endowment contract for benchmarking the LLSMC algorithm. The main motivation being that we derive a closed-form expression for its price. This contract subscribed by a x- years old individual promises at expiry (date T) the maximum between a capital C_T and the value of the stock indice S_T, in case of survival. The benefit is nevertheless upper bounded by C_M. If we denote by $\tau \in \mathbb{R}^+$, the random time of insured's death, the value of such a policy is equal to the expected discounted cash-flows under the chosen risk neutral measure:

$$\begin{aligned} V_t &= \mathbb{E}^{\mathbb{Q}}\left(e^{-\int_t^T r_s ds}\mathbf{1}_{\{\tau \geq T\}}\left(C_T + (S_T - C_T)_+ - (S_T - C_M)_+\right) \mid \mathcal{F}_t\right) \qquad (28) \\ &= \mathbf{1}_{\{\tau \geq t\}} C_T \mathbb{E}^{\mathbb{Q}}\left(e^{-\int_t^T (r_s + \mu_{x+s})ds} \mid \mathcal{F}_t\right) \\ &+ \mathbf{1}_{\{\tau \geq t\}} \mathbb{E}^{\mathbb{Q}}\left(e^{-\int_t^T (r_s + \mu_{x+s})ds}\left((S_T - C_T)_+\right) \mid \mathcal{F}_t\right) \\ &- \mathbf{1}_{\{\tau \geq t\}} \mathbb{E}^{\mathbb{Q}}\left(e^{-\int_t^T (r_s + \mu_{x+s})ds}\left((S_T - C_M)_+\right) \mid \mathcal{F}_t\right). \end{aligned}$$

For the sake of simplicity, we assume that the dynamics of r_t and μ_{x+t} are similar under $\mathbb{P}$ and $\mathbb{Q}$ (this assumption may be relaxed without impacting our results). The instantaneous return of the stock indice is r_t under $\mathbb{Q}$. The zero-coupon bond, the survival probabilities and the pure endowent[1] are respectively defined by the following expectations:

$$\begin{cases} P(t,T) &= \mathbb{E}^{\mathbb{Q}}\left(e^{-\int_t^T r_s ds} | \mathcal{F}_t\right), \\ {}_Tp_{x+t} &= \mathbb{E}^{\mathbb{Q}}\left(e^{-\int_t^T \mu_{x+s} ds} \mid \mathcal{F}_t\right), \\ {}_TE_t &= \mathbf{1}_{\{\tau \geq t\}} \mathbb{E}^{\mathbb{Q}}\left(e^{-\int_t^T (r_s + \mu_{x+s})ds} \mid \mathcal{F}_t\right). \end{cases}$$

The model being affine, we can easily derive the closed-form expressions of these products. In the remainder of this article, we adopt following the notation

$$B_y(t,T) = \frac{1 - e^{-y(T-t)}}{y},$$

where $y \in \mathbb{R}^+$ is a positive parameter. We also need the following integrals of $B_y(.,.)$:

$$\begin{cases} \int_t^T B_{\kappa_r}(u,T)du &= \frac{1}{\kappa_r}\left((T-t) - B_{\kappa_r}(t,T)\right), \\ \int_t^T \sigma_x(u) B_{\kappa_\mu}(u,T)du &= \frac{\alpha e^{\beta(x+T)}}{\kappa_u}\left(B_\beta(t,T) - B_{\beta+\kappa_\mu}(t,T)\right), \end{cases}$$

and the integrals of cross-product of $B_{\kappa_r}(.,T)$ and $\sigma_x(.)B_\mu(.,T)$:

$$\begin{cases} \int_t^T B_{\kappa_r}(u,T)^2 du &= \frac{1}{\kappa_r^2}\left((T-t) - B_{\kappa_r}(t,T) - \frac{1}{2}\kappa_r B_{\kappa_r}(t,T)^2\right), \\ \int_t^T \left(\sigma_x(u) B_{\kappa_\mu}(u,T)\right)^2 du &= \frac{\alpha^2 e^{2\beta(x+T)}}{\kappa_u^2}\left(B_{2\beta}(t,T) - 2B_{2\beta+\kappa_\mu}(t,T)\right. \\ & \quad \left. + B_{2(\beta+\kappa_\mu)}(t,T)\right), \\ \int_t^T \sigma_x(u) B_{\kappa_\mu}(u,T) B_{\kappa_r}(u,T)du &= \frac{\alpha e^{\beta(x+T)}}{\kappa_\mu \kappa_r}\left(B_\beta(t,T) - B_{\kappa_\mu+\beta}(t,T)\right. \\ & \quad \left. - B_{\kappa_r+\beta}(t,T) + B_{\kappa_\mu+\kappa_r+\beta}(t,T)\right). \end{cases}$$

[1] The pure endowment pays one monetary unit at time T if the individual is still alive at maturity.

The dynamics of interest and mortality rates are affine and therefore zero-coupon and survival probabilities admit closed-form expressions detailed in the next proposition.

Proposition 5.1. *At time $0 \leq t \leq T$, the value of the discount bond of maturity T is equal to*

$$\begin{aligned} P(t,T) &= \exp\left(-r_t B_{\kappa_r}(t,T) - \int_t^T \gamma_r(u)\left(1-e^{-\kappa_r(T-u)}\right)du\right) \\ &\quad \times \exp\left(\frac{\sigma_r^2}{2}\int_t^T B_{B_{\kappa_r}}(u,T)^2 du\right), \end{aligned} \tag{29}$$

The survival probability up to time T, is given by

$$\begin{aligned} {}_Tp_{x+t} &= \exp\left(-\mu_{x+t} B_{\kappa_\mu}(t,T) - \int_t^T \gamma_x(u)\left(1-e^{-\kappa_\mu(T-u)}\right)du\right) \\ &\quad \times \exp\left(\frac{1}{2}\int_t^T \left(\sigma_x(u) B_{\kappa_\mu}(u,T)\right)^2 du\right), \end{aligned} \tag{30}$$

The pure endowment, ${}_TE_t$, admits the following expression:

$$\begin{aligned} {}_TE_t &= \mathbf{1}_{\{\tau \geq t\}} \exp\left(-r_t B_{\kappa_r}(t,T) - \mu_{x+t} B_{\kappa_\mu}(t,T) + \frac{\sigma_r^2}{2}\int_t^T B_{\kappa_r}(u,T)^2 du\right) \times \\ &\exp\left(-\int_t^T \gamma_r(u)\left(1-e^{-\kappa_r(T-u)}\right)du - \int_t^T \gamma_x(u)\left(1-e^{-\kappa_\mu(T-u)}\right)du\right) \times \\ &\exp\left(\sigma_r \epsilon_{r\mu}\int_t^T \sigma_x(u) B_{\kappa_\mu}(u,T) B_{\kappa_r}(u,T) du + \frac{1}{2}\int_t^T \left(\sigma_x(u) B_{\kappa_\mu}(u,T)\right)^2 du\right), \end{aligned} \tag{31}$$

The sketch of the proof is provided in Appendix D. In order to match the initial yield curve of zero-coupon bonds, the function $\gamma_r(u)$ satisfies the relation

$$\int_0^T \gamma_r(u)\left(1-e^{-\kappa_r(T-u)}\right)du = -\ln P(0,T) - r_0 B_{\kappa_r}(0,T) + \frac{\sigma_r^2}{2}\int_0^T B_{\kappa_r}(u,T)^2 du \tag{32}$$

Deriving twice this expression leads to the following useful reformulation of $\gamma_r(T)$:

$$\begin{aligned} \gamma_r(T) &= -\frac{1}{\kappa_r}\partial_T^2 \ln P(0,T) - \partial_T \ln P(0,T) \\ &\quad + \frac{\sigma_r^2}{2\kappa_r^2}\left(1-e^{-2\kappa_r T}\right), \end{aligned} \tag{33}$$

where $-\partial_T \ln P(0,T)$ is the instantaneous forward rate. For a given initial mortality curve ${}_Tp_x$, we show in a similar manner that the function $\gamma_x(u)$ satisfies the relation

$$\begin{aligned} \gamma_x(T) &= -\frac{1}{\kappa_\mu}\partial_T^2 \ln {}_Tp_x - \partial_T \ln {}_Tp_x + \frac{1}{\kappa_\mu}\int_0^T \sigma_x(u)^2\left(e^{-2\kappa_\mu(T-u)}\right)du \\ &= -\frac{1}{\kappa_\mu}\partial_T^2 \ln {}_Tp_x - \partial_T \ln {}_Tp_x + \frac{\alpha^2 e^{2\beta x}}{2\kappa_\mu(\kappa_\mu+\beta)}\left(e^{2\beta T} - e^{-2\kappa_\mu T}\right). \end{aligned} \tag{34}$$

Equations (33) and (34) allows us to rewrite bond prices, survival probabilities and endowment as function of initial term structure of mortality and interest rates.

Corollary 5.2. *The price at time $t \leq T$ of a discount bond of maturity T is linked to the initial interest rate curve at time $t = 0$ by the relation*

$$\begin{aligned} P(t,T) &= \exp\left(-r_t B_{\kappa_r}(t,T) - \left(\partial_t \ln P(0,t)\right) B_{\kappa_r}(t,T) + \ln\frac{P(0,T)}{P(0,t)}\right) \\ &\quad \times \exp\left(-\frac{\sigma_r^2}{4\kappa_r}\left(\left(1-e^{-2\kappa_r t}\right) B_{\kappa_r}(t,T)^2\right)\right). \end{aligned} \tag{35}$$

In a similar manner, we can show that, being alive at age $x+t$, the survival probability up to time T depends on the initial survival term structure as follows:

$$
\begin{aligned}
{}_Tp_{x+t} = \exp&\left(-\mu_{x+t}B_{\kappa_\mu}(t,T) - \left(\partial_t \ln {}_tp_x\right) B_{\kappa_\mu}(t,T) + \ln \frac{{}_Tp_x}{{}_tp_x}\right) \times \\
&\exp\left(\frac{\alpha^2 e^{2\beta(x+T)}}{2\kappa_\mu^2}\left(B_{2\beta}(t,T) - 2B_{2\beta+\kappa_\mu}(t,T) + B_{2\beta+2\kappa_\mu}(t,T)\right)\right) \times \\
&\exp\left(\frac{\alpha^2 e^{2\beta x}}{2\kappa_\mu(\kappa_\mu+\beta)}\left(e^{2\beta T}B_{2\beta+\kappa_\mu}(t,T) - e^{2\beta T}B_{2\beta}(t,T)\right)\right) \times \\
&\exp\left(\frac{\alpha^2 e^{2\beta x}}{2\kappa_\mu(\kappa_\mu+\beta)}\left(e^{-2\kappa_\mu t}B_{2\kappa_\mu}(t,T) - e^{-\kappa_\mu(T+t)}B_{\kappa_\mu}(t,T)\right)\right),
\end{aligned}
\tag{36}
$$

whereas the pure endowment contract becomes:

$$
\begin{aligned}
{}_TE_t = \mathbf{1}_{\{\tau\geq t\}}\, {}_Tp_{x+t}\, P(t,T)\ \times \\
\exp\left(\frac{\sigma_r\epsilon_{r\mu}\alpha e^{\beta(x+T)}}{\kappa_\mu\kappa_r}\left(B_\beta(t,T) - B_{\kappa_\mu+\beta}(t,T) - B_{\kappa_r+\beta}(t,T) + B_{\kappa_\mu+\kappa_r+\beta}(t,T)\right)\right).
\end{aligned}
\tag{37}
$$

The next result presents the dynamics of the discount bond and endowment under the pricing measure. This is a direct consequence of the Itô's lemma applied to Proposition 5.1.

Corollary 5.3. *Under the risk neutral measure $\mathbb{Q}$, the dynamics of the zero-coupon bond and of the pure endowment at time $t \leq T$ are given by*

$$
\begin{cases}
dP(t,T) &= r_t P(t,T)dt - P(t,T)B_{\kappa_r}(t,T)\sigma_r\left(\epsilon_{rr}dW_t^{(2)} + \epsilon_{r\mu}dW_t^{(3)}\right), \\
d\,{}_TE_t &= {}_TE_t\left(r_t + \mu_{x+t}\right)dt - {}_TE_t\sigma_r\epsilon_{rr}B_{\kappa_r}(t,T)dW_t^{(2)} \\
&\quad -{}_TE_t\left(B_{\kappa_\mu}(t,T)\sigma_x(t) + \sigma_r\epsilon_{r\mu}B_{\kappa_r}(t,T)\right)dW_t^{(3)} + {}_T\,E_t\,d\mathbf{1}_{\{\tau\geq t\}}.
\end{cases}
\tag{38}
$$

As $\mathbb{E}^{\mathbb{Q}}\left(d\mathbf{1}_{\{\tau\geq t\}}\right) = -\mu_{x+t}dt$, we check that the pure endowment has a return equal to the risk free rate: $\mathbb{E}^{\mathbb{Q}}\left(d\,{}_TE_t\right) = {}_TE_t\,r_t\,dt$. In order to obtain a closed form expression of the saving contract (28), we perform a change of measure using as Radon-Nykodym derivative :

$$
\left.\frac{d\mathbb{F}}{d\mathbb{Q}}\right|_T = \mathbb{E}^{\mathbb{Q}}\left(\frac{d\mathbb{F}}{d\mathbb{Q}}|\mathcal{F}_T\right) = \frac{e^{-\int_0^T(r_s+\mu_{x+s})ds}}{\mathbb{E}^{\mathbb{Q}}\left(e^{-\int_0^T(r_s+\mu_{x+s})ds}|\mathcal{F}_0\right)}.
\tag{39}
$$

From Equations (46) of Appendix D, this change of measure may be rewritten as follows:

$$
\begin{aligned}
\left.\frac{d\mathbb{F}}{d\mathbb{Q}}\right|_T = \exp&\left(-\frac{\sigma_r^2\epsilon_{rr}^2}{2}\int_0^T B_{\kappa_r}(u,T)^2du - \sigma_r\epsilon_{rr}\int_0^T B_{\kappa_r}(u,T)\,dW_u^{(2)}\right) \\
&\times\exp\left(-\int_0^T\left(\sigma_r\epsilon_{r\mu}B_{\kappa_r}(u,T) + \sigma_x(u)B_{\kappa_\mu}(u,T)\right)dW_u^{(3)}\right) \\
&\times\exp\left(-\frac{1}{2}\int_0^T\left(\sigma_r\epsilon_{r\mu}B_{\kappa_r}(u,T) + \sigma_x(u)B_{\kappa_\mu}(u,T)\right)^2du\right).
\end{aligned}
$$

We recognize a Doleans-Dade exponential and then under the measure $\mathbb{F}$, $W_t^{(2)\mathbb{F}}$ and $W_t^{(3)\mathbb{F}}$ defined by

$$
\begin{cases}
dW_t^{(2)\mathbb{F}} &= dW_t^{(2)} + \sigma_r\epsilon_{rr}B_{\kappa_r}(t,T)dt\,, \\
dW_t^{(3)\mathbb{F}} &= dW_t^{(3)} + \sigma_r\epsilon_{r\mu}B_{\kappa_r}(t,T)dt + \sigma_x(t)B_{\kappa_\mu}(t,T)dt\,,
\end{cases}
\tag{40}
$$

are Brownian motions. The dynamic of the stock indice is modified as follows under $\mathbb{F}$,

$$\begin{aligned}\frac{dS_t}{S_t} &= \left(r_t - \sigma_S\left(\epsilon_{Sr}\sigma_r\epsilon_{rr} + \epsilon_{S\mu}\sigma_r\epsilon_{r\mu}\right)B_{\kappa_r}(t,T) - \sigma_S\epsilon_{S\mu}\sigma_x(t)B_{\kappa_\mu}(t,T)\right)dt \\ &\quad +\sigma_S\epsilon_{SS}dW_t^{(1)} + \sigma_S\epsilon_{Sr}dW_t^{(2)\mathbb{F}} + \sigma_S\epsilon_{S\mu}dW_t^{(3)\mathbb{F}}\,.\end{aligned} \tag{41}$$

If we remember that $\epsilon_{SS}^2 + \epsilon_{Sr}^2 + \epsilon_{S\mu}^2 = 1$, applying the Itô's lemma to $\ln S_t$ leads to the following expression for the stock indice under $\mathbb{F}$:

$$\begin{aligned}S_t &= S_0\exp\left(\int_0^t r_u du - \frac{\sigma_S^2}{2}t - \sigma_S\epsilon_{S\mu}\int_0^t \sigma_x(u)B_{\kappa_\mu}(u,T)du\right) \\ &\quad\times\exp\left(-\sigma_S\left(\epsilon_{Sr}\sigma_r\epsilon_{rr} + \epsilon_{S\mu}\sigma_r\epsilon_{r\mu}\right)\int_0^t B_{\kappa_r}(u,T)du\right) \\ &\quad\times\exp\left(\sigma_S\epsilon_{SS}W_t^{(1)} + \sigma_S\epsilon_{Sr}W_t^{(2)\mathbb{F}} + \sigma_S\epsilon_{S\mu}W_t^{(3)\mathbb{F}}\right)\,,\end{aligned} \tag{42}$$

Taking advantage the log-normality of S_T under the $\mathbb{F}$-measure, we can deduce a closed-form expression of call options embedded in the benefits, such as defined in Equation (28).

Proposition 5.4. *The log-return* $\ln\left(S_T/S_t\right) \sim N(\left(\mu_{\mathbb{F}}(t,T), v_{\mathbb{F}}(t,T)^2\right)$ *is log-normal with a mean and variance respectively given by*

$$\begin{cases}\mu_{\mathbb{F}}(t,T) &= -\frac{\sigma_r^2}{2}\int_t^T B_{\kappa_r}(u,T)^2du - \sigma_r\epsilon_{r\mu}\int_t^T \sigma_x(u)B_{\kappa_r}(u,T)B_{\kappa_\mu}(u,T)du \\ &\quad -\frac{\sigma_S^2(T-t)}{2} - \sigma_S\sigma_r\left(\epsilon_{Sr}\epsilon_{rr} + \epsilon_{S\mu}\epsilon_{r\mu}\right)\int_t^T B_{\kappa_r}(u,T)du \\ &\quad -\sigma_S\epsilon_{S\mu}\int_t^T \sigma_x(u)B_{\kappa_\mu}(u,T)du\,, \\ v_{\mathbb{F}}(t,T)^2 &= \sigma_S^2(T-t) + \sigma_r^2\int_t^T B_{\kappa_r}(u,T)^2du \\ &\quad +2\sigma_S\sigma_r\left(\epsilon_{Sr}\epsilon_{rr} + \epsilon_{S\mu}\epsilon_{r\mu}\right)\int_t^T B_{\kappa_r}(u,T)du\,.\end{cases} \tag{43}$$

If we adopt the following notations,

$$\begin{cases}d_2(t,T) &= \frac{\ln\left(\frac{C}{S_t/P(t,T)}\right) - \mu_{\mathbb{F}}(t,T)}{v_{\mathbb{F}}(t,T)}\,, \\ d_1(t,T) &= d_2 - v_{\mathbb{F}}(t,T),\end{cases}$$

The embedded call options in the participating pure endowment contract (28) are valued by:

$$\begin{aligned}&\mathbf{1}_{\{\tau\geq t\}}\mathbb{E}^{\mathbb{Q}}\left(e^{-\int_t^T(r_s+\mu_{x+s})ds}\left((S_T - C)_+\right)\mid\mathcal{F}_t\right) \\ &=_T E_t\left[\frac{S_t e^{\mu_{\mathbb{F}}(t,T)+\frac{v_{\mathbb{F}}(t,T)^2}{2}}}{P(t,T)}\Phi\left(-d_1(t,T)\right) - C_T\,\Phi\left(-d_2(t,T)\right)\right]\,.\end{aligned} \tag{44}$$

The sketch of the proof is provided in Appendix B. The exact value of the pure endowment is obtained by combining Equations (37) and (44). This allow us to compare LSMC and LLSMC approximated value to exact price of the participating pure endowment in the next subsection..

5.2 Numerical illustration

We fit a Nelson-Siegel model to the Belgian state yield curve on the 23/11/22. Initial survival probabilities are described by a Makeham's model adjusted to male Belgian mortality rates. Details are provided in Appendix C and D. Other market parameters are estimated from time series of the Belgian stock index BEL 20 and of the 1 year Belgian state yield

from the 26/11/10 to the 23/11/22. As we do not have enough data, the correlations $\rho_{S\,\mu}$ and $\rho_{r\,\mu}$ are set to -5% and 0%. Parameter estimates are reported in Table 1.

Parameters			
μ	0.04642	σ_S	0.18470
κ_r	0.20482	σ_r	0.00774
ρ_{Sr}	-0.03957	r_0	0.0235
α	8.5277e-7	β	0.11094
κ_μ	0.83925	μ_0	3.325e-03
$\rho_{S\,\mu}$	-0.05000	$\rho_{r\mu}$	0.00000
t	5 years	T	10 years
S_0	100	C_T	100
x	50	C_M	$100(1+3\%)^{10}$

Table 11: Model parameters and features of the contract.

The three risk factors are the normed stock price, normed short rate and normed mortality rate at the end of the time horizon of primary simulations, noted t:

$$\boldsymbol{X}_t := \left(\frac{S_t - \mathbb{E}_0^{\mathbb{P}}(S_t)}{\sqrt{\mathbb{V}_0^{\mathbb{P}}(S_t)}}, \frac{\sqrt{r_t} - \mathbb{E}_0^{\mathbb{P}}(r_t)}{\sqrt{\mathbb{V}_0^{\mathbb{P}}(\sqrt{r_t})}}, \frac{\sqrt{\mu_{x+t}} - \mathbb{E}_0^{\mathbb{P}}(\mu_{x+t})}{\sqrt{\mathbb{V}_0^{\mathbb{P}}(\sqrt{\mu_{x+t}})}} \right).$$

Expectations and variances are approached by empirical averages and variances of the simulated sample. The features of the contract are reported in Table 1. We simulate 10000 primary scenarios and a single secondary response per scenario,

$$Y(t) \quad = \quad e^{-\int_t^T r_s + \mu_s ds} \left(C_T + (S_T - C_T)_+ - (S_T - C_M)_+ \right).$$

We use 350 steps of time per year. We also calculate the exact value of the contract in each scenario using analytical formulas of the previous section.

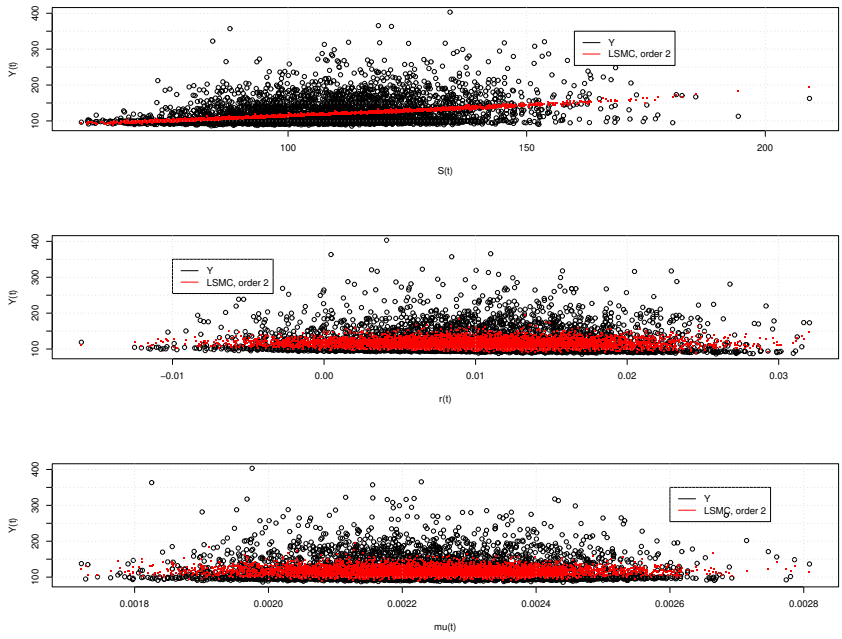

Figure 9: Simulated responses $Y(t) =$ versus stock prices S_t, r_t and $\mu_x + t$. The red curves are the predictions $h(\boldsymbol{X}_t)$ from the LSMC with a second order polynomials regression.

The three plots of Figure 9 respectively show responses versus stock prices, interest and mortality rates. The red curves correspond to LSMC estimates of the endowment in one year with a second order polynomial of risk factors.

Table 12 reports the R^2, the MSE and MSE($\mathcal{V}$) of the LSMC, such as defined by Equations (16), (18) and (19). The validation set counts 1000 triplets of risk factors. We consider combinations of $q = 10$ empirical quantiles of risk factors for probabilities from 1% to 5% and from 95% to 99% by step of 1%. We also calculate the exact MSE between model and analytical prices of the endowment, denoted by EMSE.

d_h	R^2	$\sqrt{\text{MSE}(\mathcal{V})}$	$\sqrt{\text{MSE}}$	$\sqrt{\text{EMSE}}$	d.f.
2	0.3767	3.02	11.34	1.59	10
3	0.3855	2.47	11.27	1.04	20
4	0.3868	3.94	11.28	1.12	35
5	0.3912	4.85	11.26	1.02	56
6	0.3947	6.47	11.26	1.10	84

Table 12: R^2, MSE, MSE($\mathcal{V}$) of regressions of Y_t on $\boldsymbol{X}_t$ in the LSMC model. $\sqrt{\text{EMSE}}$ is the MSE valued with analytical prices. d.f. is the number of parameters.

Table 12 reports statistics about LSMC polynomial regressions of order d_h from 2 to 6. The R^2's increase with the complexity of the model. The MSE($\mathcal{V}$) on the validation set is minimized by a polynomial of third degree. Table 3 presents the statistics of goodness of fit for the LLSMC model. The number of clusters, K, varies from 2 to 5. We test polynomials of degrees d_h from 1 to 3 and d_γ equal to 1 and 3. Models are sorted by increasing MSE($\mathcal{V}$)'s and we report statistics of the 10 first best models according to this criterion. The best goodness of fit is achieved with 2 or 3 clusters, a square regression on each cluster and a cubic multinomial logistic regression. Compared to the LSMC, the LLSMC clearly reduces by more than half the MSE($\mathcal{V}$) and the EMSE whereas MSE on the training set are comparable. This is a good indicator that the LLSMC model offers a better fit.

K	d_γ	d_h	R^2	$\sqrt{\text{MSE}(\mathcal{V})}$	$\sqrt{\text{MSE}}$	$\sqrt{\text{EMSE}}$	d.f.	R^2_{loc}
2	3	2	0.3910	0.69	11.24	0.67	40	0.8781
3	3	2	0.3912	0.79	11.27	0.55	70	0.9322
4	2	2	0.3865	0.87	11.32	0.77	70	0.9563
5	2	2	0.387	0.91	11.34	0.75	90	0.9687
3	2	2	0.3869	0.93	11.29	0.76	50	0.9322
2	2	2	0.3881	0.95	11.26	0.65	30	0.8781
4	3	2	0.3918	0.97	11.3	0.58	100	0.9563
4	2	3	0.3843	0.98	11.38	1.08	110	0.9568
2	2	3	0.3592	1.01	11.54	2.60	50	0.8789
2	3	3	0.3915	1.03	11.26	0.63	60	0.8789

Table 13: R^2, MSE, MSE($\mathcal{V}$) and R^2_{loc} for the LLSMC model. $\sqrt{\text{MSE}}$, exact is the MSE valued with analytical prices. d.f. is the number of parameters.

We next compare the LSMC of order 3 and the LLSMC with hyperparameters $K = 3$, $d_\gamma = 3$, $d_h = 2$ as this setting leads to a low MSE($\mathcal{V}$) and a high R^2_{loc}. Figure 5 compares LSMC and LLSMC endowment values for stock prices S_t ranging from 43 to 302, the 1% and 99% percentiles of simulated stock prices over 5 years and $r_t \in \{-0.16\%, 2.47\%, 5.13\%\}$, the 1%, 50% and 99% quantiles of simulated interest rates. The

mortality rate is set to its average $\mu_{x+t} = 0.0017$. LSMC and LLSMC both achieve a good accuracy in these three cases. Nevertheless, pricing errors of the LLSMC, reported in Table 14, are slightly lower on average than those of the LSMC when $r_t \in \{-0.16\%, 2.47\%\}$. In particular, the LLSMC better fits extreme low values. This will be confirmed by the comparison of VaR's and TVaR's.

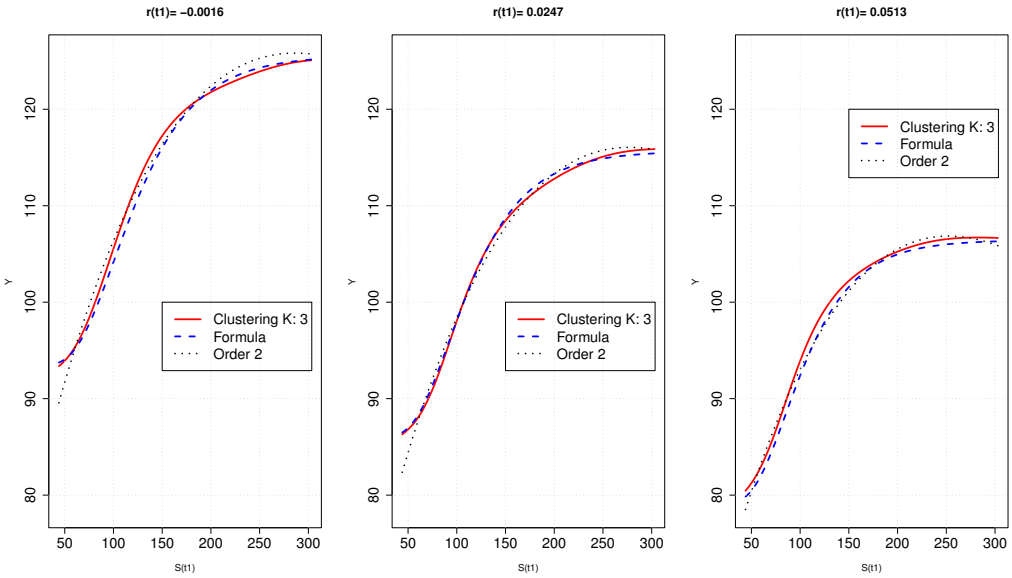

Figure 10: Endowment price for S_t ranging from 43 to 302, $r_t \in \{-0.16\%, 2.47\%, 5.13\%\}$ and $\mu_{x+t} = 0.0017$.

	r_t =-0.16%	r_t =2.47%	r_t =5.13%
LLSMC, 3-3-2	0.8459	0.3496	0.8198
LSMC, order 3	1.2602	0.8872	0.7606

Table 14: Average pricing errors for the three cases presented in Figure 10.

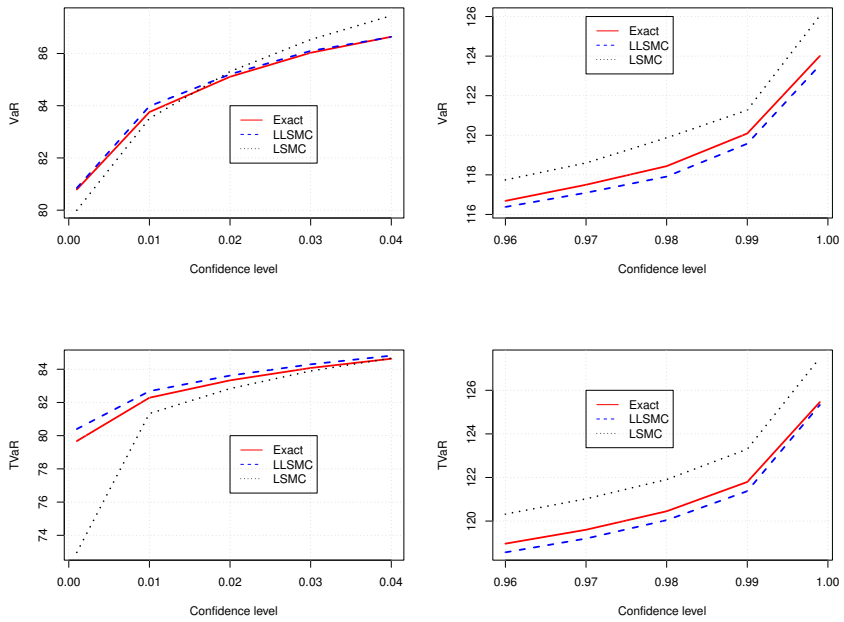

Figure 11: Lower/upper VaR's and TVaR's, LLSMC 3-3-2, LSMC of order 3.

Upper parts of Tables 15 and 16 respectively present VaR's and TVaR's computed with the LLSMC and LSMC models and with analytical prices. Figure 11 displays these statistics for the LLSMC 3-3-2 and for the LSMC of order 3. The lower Upper parts of Tables 15 and 16 report the relative spread between VaR/TVaR's computed with approximated and exact analytical prices. These results emphasizes that LLSMC yields a more accurate estimate of VaR/TVaR's. In particular, the failure of the LSMC to closely replicate extreme values leads to a significant divergence of TVaR's for very low or high confidence levels.

VaR	0.1%	1%	2%	3%	97%	98%	99%	99.9%
Exact prices	80.79	83.76	85.11	86.03	117.5	118.44	120.09	124.01
LLSMC 2-3-2	80.95	83.84	85.01	86	116.99	118.07	119.55	124.01
LLSMC 3-3-2	80.85	83.98	85.2	86.1	117.09	117.91	119.58	123.54
LLSMC 4-2-2	82.14	84.83	85.9	86.71	117.49	118.45	120.16	123.74
LSMC d_h =2	80	83.51	85.31	86.53	118.6	119.87	121.27	126.04
LSMC, d_h =3	78.16	81.39	84.02	85.16	117.49	118.41	120.06	123.46
Relative errors in %								
LLSMC 2-3-2	0.19	0.1	-0.11	-0.03	-0.43	-0.31	-0.45	0.00
LLSMC 3-3-2	0.07	0.27	0.11	0.08	-0.34	-0.44	-0.43	-0.38
LLSMC 4-2-2	1.67	1.28	0.93	0.79	-0.01	0.01	0.06	-0.22
LSMC d_h =2	-0.99	-0.29	0.23	0.59	0.94	1.21	0.98	1.64
LSMC, d_h =3	-3.26	-2.82	-1.28	-1.01	-0.01	-0.02	-0.03	-0.44

Table 15: VaR 5 years, LSMC model and LLSMC.

Tail VaR	0.1%	1%	2%	3%	97%	98%	99%	99.9%
Exact prices	79.68	82.28	83.33	84.08	119.6	120.45	121.79	125.46
LLSMC 2-3-2	80.42	82.5	83.45	84.14	119.2	120.06	121.42	126.11
LLSMC 3-3-2	80.40	82.68	83.62	84.3	119.2	120.04	121.37	125.33
LLSMC 4-2-2	81.59	83.63	84.50	85.12	119.63	120.45	121.71	125.04
LSMC d_h =2	72.97	81.33	82.84	83.9	121.02	121.91	123.3	127.49
LSMC, d_h =3	76.86	79.91	81.31	82.43	119.66	120.51	121.93	125.53
Relative errors in %								
LLSMC 2-3-2	0.93	0.26	0.15	0.08	-0.33	-0.33	-0.31	0.52
LLSMC 3-3-2	0.91	0.48	0.34	0.26	-0.34	-0.34	-0.35	-0.10
LLSMC 4-2-2	2.40	1.63	1.40	1.24	0.02	0.00	-0.07	-0.33
LSMC d_h =2	-8.41	-1.16	-0.60	-0.21	1.19	1.21	1.23	1.62
LSMC, d_h =3	-3.54	-2.88	-2.42	-1.96	0.05	0.05	0.11	0.06

Table 16: Expected shortfall, 5 years, LSMC model and LLSMC.

	Cluster, k		std. errors	
	1	2	1	2
Intercept	113.6213	88.9504	0.1533	0.1747
$x^{(1)}$	1.3713	1.7745	0.1523	0.1835
$\left(x^{(1)}\right)^2$	-0.2007	-0.2551	0.0452	0.1697
$x^{(2)}$	-2.951	-1.7939	0.1116	0.1326
$x^{(1)}x^{(2)}$	-0.4279	0.8160	0.0973	0.1560
$\left(x^{(2)}\right)^2$	-0.0175	0.1425	0.0733	0.0696
$x^{(3)}$	0.0639	0.1082	0.1123	0.1368
$x^{(1)}x^{(3)}$	-0.0894	0.2613	0.0999	0.1643
$x^{(2)}x^{(3)}$	-0.0260	0.0329	0.1016	0.1001
$\left(x^{(3)}\right)^2$	0.0114	-0.0497	0.0710	0.0699

Table 17: Parameters of cubic regressions, $h_k(\boldsymbol{x})$ per cluster and standard errors (std. err.)

Tables 7 and 8 report estimated coefficients of polynomials $\widehat{h}_k(.)$ and $\widehat{\gamma}_k(.)$ of the LLSMC, 3-2-2. As already mentioned, standard errors are high because responses are by construction very noised estimates of $\mathbb{E}^{\mathbb{Q}}\left(Y(t) \,|\, \boldsymbol{x}\right)$.

	Cluster, k	std. errors		Cluster, k	std. errors
	2	2		2	2
Intercept	-0.558	0.0573	$x^{(3)}$	-0.0048	0.0685
$x^{(1)}$	-1.6895	0.1037	$x^{(1)}x^{(3)}$	0.0230	0.0597
$\left(x^{(1)}\right)^2$	0.6223	0.0772	$\left(x^{(1)}\right)^2 x^{(3)}$	-0.0451	0.0639
$\left(x^{(1)}\right)^3$	-0.2098	0.0562	$x^{(2)}x^{(3)}$	0.0141	0.0370
$x^{(2)}$	0.01	0.0666	$x^{(1)}x^{(2)}x^{(3)}$	-0.0587	0.0568
$x^{(1)}x^{(2)}$	0.026	0.0551	$\left(x^{(2)}\right)^2 x^{(3)}$	0.0108	0.0251
$\left(x^{(1)}\right)^2 x^{(2)}$	-0.0069	0.0574	$\left(x^{(3)}\right)^2$	-0.0330	0.0264
$\left(x^{(2)}\right)^2$	-0.0244	0.0255	$x^{(1)}\left(x^{(3)}\right)^2$	-0.0435	0.0428
$x^{(1)}\left(x^{(2)}\right)^2$	0.0619	0.0369	$x^{(2)}\left(x^{(3)}\right)^2$	-0.0484	0.0265
$\left(x^{(2)}\right)^3$	0.0105	0.0152	$\left(x^{(3)}\right)^3$	-0.0007	0.0150

Table 18: Parameters of logistic regressions, $\gamma_k(\boldsymbol{x})$ per cluster and standard errors (std. err.)

6 Conclusions

This article proposes a powerful and simple extension of the least squares Monte-Carlo method for risk management. This combines local and logistic regressions. The novelty of our approach consists to segment the data set into clusters obtained by applying the K-means algorithm to responses instead of risk factors. We next fit polynomial regressions for each cluster. They are combined with probabilities of cluster membership estimated by a multinomial logistic regression.

We validate the LLSMC in two case studies. In both cases, numerical experiments emphasize that the LLMSC achieves a better accuracy than the LSMC in a wider range of scenarios. We also observe that the LLMSC yields fewer erratic prices for lower and upper quantiles of risk factors. This confirms that the LLMSC is better suited for computing risk measures such as the VaR and TailVaR, than the LSMC. Furthermore, the LLMSC has a high level of interpretability. We also compare the LLSMC to a local method based on a partition of risk factors ($\mathcal{X}$-LLSMC). We show that such an approach suffers from the Simpson's paradox, i.e. $\mathcal{X}$-LLSMC prices display local trends not relevant with global ones.

This work paves the way for further research. Firstly, the LLSMC algorithm is probably more efficient than the LSMC for estimating the solvency capital requirement in the Solvency II framework. Secondly, we can think to replace local polynomial approximations by local machine learning regressions. This hybrid procedure would probably be best suited for managing a high number of risk factors.

Appendix A

Algorithm 1 Algorithm for K-means clustering.

Initialization:

Randomly set up initial positions of centroids $c_1(0)$,...,$c_K(0)$.

Main procedure:

For $e = 0$ to maximum epoch, e_{max}

Assignment step:

For $i = 1$ to n

1) Assign $(\boldsymbol{x}_i, y_i)$ to a cluster $\mathcal{S}_k(e)$ and y_i to $\mathcal{Y}_k(e)$ where $k \in \{1, ..., K\}$

$$\mathcal{S}_k(e) = \{(\boldsymbol{x}_i, y_i) \ : \ d(y_i, c_k(e)) \leq d(y_i, c_j(e)) \ \forall j \in \{1, ..., K\}\},$$
$$\mathcal{Y}_k(e) = \{y_i \ : \ d(y_i, c_k(e)) \leq d(y_i, c_j(e)) \ \forall j \in \{1, ..., K\}\},$$

End loop on data set, i.

Update step:

For $k = 1$ to K

2) set the new centroids $c_k(e+1)$ to the center of gravity of $\mathcal{Y}_k(e)$

$$c_k(e+1) = \frac{1}{|\mathcal{Y}_k(e)|} \sum_{y_i \in \mathcal{Y}_k(e)} y_i \, .$$

End loop on centroids, k.

3) Update the intraclass inertia :

$$I_a(e+1) = \frac{1}{n} \sum_{k=1}^{K} \sum_{y_i \in \mathcal{Y}_k(e)} d(x_i, c_k(e+1)) \, .$$

End loop on epochs e

Appendix B

Proposition 5.1, sketch of the proof. We can show by direct differentation that

$$\begin{pmatrix} r_s \\ \mu_{x+s} \end{pmatrix} = \begin{pmatrix} e^{-\kappa_r(s-t)} r_t \\ e^{-\kappa_\mu(s-t)} \mu_{x+t} \end{pmatrix} + \begin{pmatrix} \kappa_r \int_t^s \gamma_r(u) e^{-\kappa_r(s-u)} du \\ \kappa_\mu \int_t^s \gamma_x(u) e^{-\kappa_\mu(s-u)} du \end{pmatrix} + \begin{pmatrix} \int_t^s \sigma_r e^{-\kappa_r(s-u)} \epsilon_{rr} dW_u^{(2)} + \int_t^s \sigma_r e^{-\kappa_r(s-u)} \epsilon_{r\mu} dW_u^{(3)} \\ \int_t^s e^{-\kappa_\mu(s-u)} \sigma_x(u) dW_u^{(3)} \end{pmatrix} \tag{45}$$

The integrals are

$$\begin{pmatrix} \int_t^T r_s ds \\ \int_t^T \mu_{x+s} ds \end{pmatrix} = \begin{pmatrix} r_t B_{\kappa_r}(t,T) \\ \mu_{x+t} B_{\kappa_\mu}(t,T) \end{pmatrix} + \begin{pmatrix} \int_t^T \gamma_r(u) \left(1 - e^{-\kappa_r(T-u)}\right) du \\ \int_t^T \gamma_x(u) \left(1 - e^{-\kappa_\mu(T-u)}\right) du \end{pmatrix} + \begin{pmatrix} \sigma_r \epsilon_{rr} \int_t^T B_{\kappa_r}(u,T) \, dW_u^{(2)} + \sigma_r \epsilon_{r\mu} \int_t^T B_{\kappa_r}(u,T) dW_u^{(3)} \\ \int_t^T \sigma_x(u) B_{\kappa_\mu}(u,T) \, dW_u^{(3)} \end{pmatrix} \tag{46}$$

The results follows from the log-normality of $e^{-\int_t^T r_s ds}$ or $e^{-\int_t^T r_s + \mu_{x+s} ds}$ and of $\epsilon_{rr}^2 + \epsilon_{r\mu}^2 = 1$.

end

Corollary 5.2, sketch of the proof. By direct integration of Equations (33) and (34), we obtain

$$\begin{aligned}\int_t^T \gamma_r(u)\left(1-e^{-\kappa_r(T-u)}\right)du &= \left(\partial_t \ln P(0,t)\right)B_{\kappa_r}(t,T) - \ln\frac{P(0,T)}{P(0,t)} + \frac{\sigma_r^2}{2\kappa_r^2}(T-t)\\ &\quad -\frac{\sigma_r^2}{2\kappa_r^2}B_{\kappa_r}(t,T) - \frac{\sigma_r^2}{4\kappa_r}e^{-2\kappa_r t}B_{\kappa_r}(t,T)^2,\end{aligned}$$

and

$$\begin{aligned}\int_t^T \gamma_x(u)\left(1-e^{-\kappa_\mu(T-u)}\right)du &= \left(\partial_t \ln {}_tp_x\right)B_{\kappa_\mu}(t,T) - \ln\frac{{}_Tp_x}{{}_tp_x} + \frac{\alpha^2e^{2\beta x}}{2\kappa_\mu(\kappa_\mu+\beta)}\\ &\times\left(e^{2\beta T}B_{2\beta}(t,T) - e^{-2\kappa_\mu t}B_{2\kappa_\mu}(t,T) - e^{2\beta T}B_{2\beta+\kappa_\mu}(t,T) + e^{-\kappa_\mu(T+t)}B_{\kappa_\mu}(t,T)\right)\end{aligned}$$

end

Proposition 5.4, sketch of the proof.

As $S_T = \frac{S_T}{P(T,T)}$, we focus on on the dynamics of $d\frac{S_t}{P(t,T)}$. As

$$d\frac{1}{P(t,T)} = -\frac{r_t}{P(t,T)}dt + \frac{B_{\kappa_r}(t,T)^2\sigma_r^2}{P(t,T)}dt + \frac{B_{\kappa_r}(t,T)}{P(t,T)}\sigma_r\left(\epsilon_{rr}dW_u^{(2)} + \epsilon_{r\mu}dW_u^{(3)}\right)$$

We infer that

$$\begin{aligned}d\left(\frac{S_t}{P(t,T)}\right) &= \frac{S_t}{P(t,T)}\left[B_{\kappa_r}(t,T)^2\sigma_r^2 + B_{\kappa_r}(t,T)\left(\sigma_r\sigma_S\epsilon_{rr}\epsilon_{Sr} + \sigma_r\sigma_S\epsilon_{r\mu}\epsilon_{S\mu}\right)\right]dt\\ &\quad + \frac{S_t}{P(t,T)}\sigma_S\epsilon_{SS}dW_t^{(1)} + \frac{S_t}{P(t,T)}\left(\sigma_S\epsilon_{Sr} + \epsilon_{rr}\sigma_rB_{\kappa_r}(t,T)\right)dW_t^{(2)}\\ &\quad + \frac{S_t}{P(t,T)}\left(\sigma_S\epsilon_{S\mu} + \epsilon_{r\mu}\sigma_rB_{\kappa_r}(t,T)\right)dW_t^{(3)}\end{aligned}$$

Using Itô's lemma we can find the dynamics of $d\ln\left(\frac{S_t}{P(t,T)}\right)$ under $\mathbb{Q}$and from Equation (40), retrieve that

$$\begin{aligned}d\ln\left(\frac{S_t}{P(t,T)}\right) &= -\left[\frac{\sigma_r^2}{2}B_{\kappa_r}(t,T)^2 + \sigma_r\epsilon_{r\mu}\sigma_x(t)B_{\kappa_r}(t,T)B_{\kappa_\mu}(t,T)\right]dt\\ &\quad -\left[\frac{1}{2}\sigma_S^2 + \sigma_S\sigma_r\left(\epsilon_{rr}\epsilon_{Sr} + \epsilon_{S\mu}\epsilon_{r\mu}\right)B_{\kappa_r}(t,T) + \sigma_S\epsilon_{S\mu}\sigma_x(t)B_{\kappa_\mu}(t,T)\right]dt\\ &\quad +\sigma_S\epsilon_{SS}dW_t^{(1)} + \left(\sigma_S\epsilon_{Sr} + \sigma_r\epsilon_{rr}B_{\kappa_r}(t,T)\right)dW_t^{(2)\mathbb{F}}\\ &\quad +\left(\sigma_S\epsilon_{S\mu} + \sigma_r\epsilon_{r\mu}B_{\kappa_r}(t,T)\right)dW_t^{(3)\mathbb{F}}.\end{aligned}$$

By direct integration, we reformulate S_T as follows:

$$\begin{aligned}S_T = \frac{S_t}{P(t,T)}&\exp\left(-\frac{\sigma_r^2}{2}\int_t^T B_{\kappa_r}(u,T)^2du - \sigma_r\epsilon_{r\mu}\int_t^T\sigma_x(u)B_{\kappa_r}(u,T)B_{\kappa_\mu}(u,T)du\right)\\ &\times\exp\left(-\frac{\sigma_S^2(T-t)}{2} - \sigma_S\sigma_r\left(\epsilon_{Sr}\epsilon_{rr} + \epsilon_{S\mu}\epsilon_{r\mu}\right)\int_t^T B_{\kappa_r}(u,T)du\right)\\ &\times\exp\left(-\sigma_S\epsilon_{S\mu}\int_t^T\sigma_x(u)B_{\kappa_\mu}(u,T)du + \int_t^T\sigma_S\epsilon_{SS}dW_u^{(1)}\right)\\ &\times\exp\left(\int_t^T\left(\sigma_S\epsilon_{Sr} + \sigma_r\epsilon_{rr}B_{\kappa_r}(u,T)\right)dW_u^{(2)\mathbb{F}}\right)\\ &\times\exp\left(\int_t^T\left(\sigma_S\epsilon_{S\mu} + \sigma_r\epsilon_{r\mu}B_{\kappa_r}(u,T)\right)dW_u^{(3)\mathbb{F}}\right)\end{aligned} \tag{47}$$

Equation (47) emphasizes that $\ln \frac{S_T/P(T,T)}{S_t/P(t,T)} \sim N\left(\mu_{\mathbb{F}}, v_{\mathbb{F}}\right)$ is log-normal with a mean and variance given by equations (43). Using standard calculations, we can show that if $\Phi(.)$ is the cdf of a $N(0,1)$ and

$$\begin{aligned} d_2 &= \frac{\ln\left(\frac{C_T}{S_t/P(t,T)}\right) - \mu_{\mathbb{F}}}{v_{\mathbb{F}}}, \\ d_1 &= d_2 - v_{\mathbb{F}}, \end{aligned}$$

then the expected postive diffence between S_T and C_T under the forward measure is given by

$$\mathbb{E}^{\mathbb{F}}\left((S_T - C_T)_+ | \mathcal{F}_t\right) = \frac{S_t}{P(t,T)} e^{\mu_{\mathbb{F}} + \frac{v_{\mathbb{F}}^2}{2}} \Phi\left(-d_1\right) - C_T \Phi\left(-d_2\right). \tag{48}$$

This last result allows us to infer Equation (44).

end

Appendix C, interest rate assumptions

We model the initial yield curve with the Nelson-Siegel (NS) model. In this framework, initial instantaneous forward rates are provided by the following function:

$$f(0,t) := -\partial_t \ln P(0,t) = b_0^{(r)} + \left(b_{10}^{(r)} + b_{11}^{(r)} t\right) \exp\left(-c_1^{(r)} t\right).$$

Parameters $\{b_0, b_{10}, b_{11}, c_1\}$ are estimated by minimizing the quadratic spread between market and model zero-coupon yields:

$$P(0,t) = \exp\left(b_0^{(r)} + \frac{1}{t}\frac{b_{10}^{(r)}}{c_1^{(r)}}\left(1 - e^{-c_1^{(r)} t}\right) + \frac{1}{t}\frac{b_{11}^{(r)}}{\left(c_1^{(r)}\right)^2}\left(1 - \left(c_1^{(r)} t + 1\right) e^{-c_1^{(r)} t}\right)\right).$$

We fit the NS model to the yield curve of Belgian state bonds observed on the 23^{th} of November 22 and obtain estimates reported in Table 19.

Parameter	Value
$b_0^{(r)}$	0.0308
$b_{10}^{(r)}$	-0.0008
$b_{11}^{(r)}$	-0.0212
$c_1^{(r)}$	0.6594

Table 19: Nelson-Siegel parameters, Belgian state bonds, 23/11/22.

Appendix D, mortality rate assumptions

The volatility of mortality rates is fitted by least square minimization of spreads between $\sigma_x(.)$ and empirical deviations of variations of mortality rates by cohort (ages between 20 and 90 years from 1950 to 2020). If $\mu_x^{(y)}$ is the observed mortality rates at age x during the calendar year y, we denote by $\Delta\mu_x^{(y)} = \mu_x^{(y)} - \mu_{x-1}^{(y-1)}$ and by S_x the standard deviation of $\Delta\mu_x^{(y)}$ for y=1950 to 2020. The α and β are obtained by minimizing the sum

$$\alpha, \beta = \arg\min \sum_{x=20}^{90} \left(S_x - \alpha e^{\beta x}\right)^2.$$

On the other hand, the initial curve of survival probabilities is described by a Makeham's model, i.e.

$$\begin{aligned} {}_tp_x &= \exp - \int_x^{x+t} \left(a^{(\mu)} + b^{(\mu)} \left(c^{(\mu)} \right)^s \right) ds \\ &= \exp(-a^{(\mu)} t) \exp\left(-\frac{b^{(\mu)}}{\ln c^{(\mu)}} \left(\left(c^{(\mu)} \right)^{x+t} - \left(c^{(\mu)} \right)^x \right) \right). \end{aligned}$$

where $a^{(\mu)}, b^{(\mu)}, c^{(\mu)} \in \mathbb{R}^+$. These parameters and the reversion speed κ_u are obtained by least square minimization of spreads between prospective and model survival probabilities. Prospective survival probabilities are computed with a Lee-Carter model fitted to Belgian mortality rates from 1950 to 2020 for 0 to 105 years, male population. Model ${}_tp_x$ are computed with Equation (36) for $x = 20$ years old. Estimated parameters are provided in Table 20.

Parameters			
$a^{(\mu)}$	1.006349e-03	κ_μ	0.83925
$b^{(\mu)}$	2.790903e-07	α	8.5277e-7
κ_μ	0.83925	β	0.11094

Table 20: Mortality parameters, Belgian male mortality rates, year 2020.

Acknowledgement

The first author thanks the FNRS (Fonds de la recherche scientifique) for the financial support through the EOS project Asterisk (research grant 40007517). We also thank Comlan Rodrigue Tossou for his help on coding.

References

[1] Bacinello A.R., Biffis E., Millossovich P. 2009. Pricing life insurance contract with early exercisse features. Journal of Computational and Applied Mathematics, 233 (1), 27-35.

[2] Bacinello A.R., Biffis E., Millossovich P. 2010. Regression-based algorithms for life insurance contracts with surrender guarantees. Quantitative Finance, 10 (9), 1077-1090.

[3] Bauer D., Reuss A., Singer D., 2012. On the calculation of the solvency capital requirement based on nested simulations. ASTIN Bulletin 42, 453–99.

[4] Becker S., Cheridito P., Jentzen A. 2020. Pricing and hedging American-Style options with deep learning. Journal of Risk and Financial Management 13 (7), 158.

[5] Cheridito P., Ery J., Wüthrich M.V. 2020. Assessing asset-liability risk with neural networks. Risks 2020 8 (1).

[6] Clement E., Lamberton D., Protter P. 2002. An analysis of a least squares regression method for American option pricing. Finance Stoch. 6, 449–471.

[7] Floryszczak A., Le Courtois O., Majri M., 2016. 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.

[8] Hejazi S.A., Jackson K.R., 2017. Efficient valuation of SCR via a neural network approach. Journal of Computational and Applied Mathematics, 313 (15), 427–39.

[9] Glasserman P., Yu B. 2004. Number of paths versus number of basis functions in American option pricing. Ann. Appl. Probab. 14(4), 2090–2119.

[10] Hainaut D. 2022. Continuous time processes for finance. Switching, self-exciting fractional and other recent dynamics. Springer, Bocconi university press.

[11] Heston S. 1993. A closed-form solution for options with stochastic volatility with applications to bond andcurrency options, Review of Financial Studies, 6, 327–343

[12] Hörig M., Leitschkis M. 2012. Solvency II proxy modelling via least squares Monte Carlo. Milliman working paper. http://www.milliman.com/insight/insurance/Solvency-II-proxy-modelling-via-Least-Squares-Monte-Carlo/

[13] Hörig M, Leitschkis M, Murray K, Phelan E (2014) An application of Monte Carlo proxy techniques to variable annuity business: a case study. https://www.milliman.com/en/insight/an-application-of-monte-carlo-proxy-techniques-to-variable-annuity-business-a-case-study

[14] Lapeyre B., Lelong J. 2021. Neural network regression for Bermudan option pricing. Monte Carlo Methods and Applications 27 (3), 227-247.

[15] Longstaff F.A., Schwartz E.S. 2001. Valuing American options by simulation: a simple least square approach. The review of Financial Studies 14 (1), 113-147.

[16] Moreno M., Navas J.F. 2003. On the robustness of Least-Squares Monte-Carlo (LSM) for pricing american options. Review of derivatives research 6 (2), 107-128.

[17] Pelsser A., Schweizer J. 2016. The difference between LSMC and replicating portfolio in insurance liability modeling. European Actuarial Journal 6, 441-494.

[18] Stentoft L. 2004. Assessing the least squares Monte-Carlo approach to American option valuation. Review of derivatives research 7 (2), 107-128.

The DetraNotes

The Detra Notes are a series of educational papers dedicated to the insurance sector. Those notes are published by members of the Detralytics team and written in a clear and accessible language. The team combines academic expertise and business knowledge.

Detralytics was founded to support companies in the advancement of actuarial science and the solving of the profession's future challenges. It is within the scope of this mission that we make our work available through our DetraNotes and FAQctuary's series.

Authors' biographies

Donatien Hainaut

Donatien Hainaut is Scientific Director at Detralytics and professor at UCLouvain where he is Director of the new Master program in Data Science, statistical orientation. Prior to this he held several positions as associate professor at Rennes School of Business and the ENSAE in Paris. He also has several field experiences having worked as Risk Officer, Quantitative Analyst and ALM Officer.

Donatien is a Qualified Actuary and holds a PhD in the area of Assets and Liability Management. His current research focuses on contagion mechanism in stochastic processes and applications of neural networks to insurance.

Adnane Akbaraly

Adnane is part of the Talent Accelerator Program (TAP) at Detralytics. Prior to joining Detralytics, Adnane worked at CNP Assurances in retirement department and developed skills in Market Consistency Embedded Value (MCEV) assessment. He graduated with a Master's degree in Actuarial Sciences from ISFA in Lyon, France. His Master's thesis deals with the construction of a prospective mortality law by an exogenous method.

People drive actuarial innovation

www.detralytics.com - info@detralytics.eu