

RESEARCH ARTICLE

Risk management with local least squares Monte Carlo

Donatien Hainaut¹  and Adnane Akbaraly²

¹UCLouvain- LIDAM Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium and ²Detralytics Rue Réaumur, 124, 75002 Paris, France

Corresponding author: Donatien Hainaut; Email: donatien.hainaut@uclouvain.be

Received: 22 February 2023; **Revised:** 17 May 2023; **Accepted:** 6 June 2023; **First published online:** 14 July 2023

Keywords: Least square Monte Carlo; risk management; option valuation

JEL codes: C5; G22

Abstract

The least squares Monte Carlo method has become a standard approach in the insurance and financial industries for evaluating a company's exposure to market risk. However, the non-linear regression of simulated responses on risk factors poses a challenge in this procedure. This article presents a novel approach to address this issue by employing an a-priori segmentation of responses. Using a K-means algorithm, we identify clusters of responses that are then locally regressed on their corresponding risk factors. The global regression function is obtained by combining the local models with logistic regression. We demonstrate the effectiveness of the proposed local least squares Monte Carlo method through two case studies. The first case study investigates butterfly and bull trap options within a Heston stochastic volatility model, while the second case study examines the exposure to risks in a participating life insurance scenario.

1. Introduction

The least squares Monte Carlo (LSMC) method 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 relies 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 *et 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 Monte Carlo simulations. Moreno and Navas (2003) and Stentoft (2004) consider the LSMC for different basis functions and deduce that the algorithm converges 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 contract embedding American options.

The LSMC is useful not only for pricing but also for managing risks. 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 II 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 in Hörig and Leitschkis (2012) or Hörig *et 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 options 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 method relies on a global regression model that predicts responses based on risk factors. In its traditional form, the regression function is approximated using either a polynomial or a combination of basis functions, as discussed in Ha and Bauer (2022, Section 4). However, when there is a non-linear relationship between responses and factors, using a high-order polynomial or a large number of basis functions increases the risk of overfitting. For a discussion on the robustness of this approach, we refer to Moreno and Navas (2003). An efficient alternative is to employ a neural regression, as proposed by Cheridito *et al.* (2020). Nevertheless, determining the optimal network architecture is a challenging task, and the lack of interpretability of the model is highlighted by Molnar (2023, Section 3). The main contribution of this article is to propose an alternative approach based on local regressions. In the first stage, we partition the responses into clusters using the K-means algorithm and then perform local regressions on the corresponding risk factors within each cluster. In the second stage, we employ a logistic regression model to estimate the probability that a combination of risk factors results in a response within each cluster. By weighting the local models using these probabilities, we obtain a global regression function. Through two case studies, we demonstrate that this local least squares Monte Carlo (LLSMC) method outperforms LSMC and offers a higher level of interpretability. The first case study examines the risk analysis of butterfly and bull trap options within a Heston stochastic volatility model. In the second case study, we consider a participating life insurance scenario and compare the risk measures computed using LLSMC and the standard LSMC approach.

The outline of the article is as follows. Section 2 reviews the LSMC 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 pure endowment. Proofs and additional numerical illustrations are provided in the Online Supplementary Materials.

2. LSMC method for risk management

2.1. Model framework

We consider a probability space Ω , endowed with a probability measure \mathbb{P} , in which are defined m Markov square-integrable processes, noted $\mathbf{X}_t = (X_t^{(1)}, \dots, X_t^{(m)})_{t \geq 0}$, with bounded variances. 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 $A(t, \mathbf{X}_t)$. \mathbb{P} is called the real or historical measure in rest of this article. 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 $(t_k)_{k=0, \dots, d}$ and denoted by C_k^A . Therefore at time $t \leq t_d$, $A(t, \mathbf{X}_t)$ can be developed as follows:

$$A(t, \mathbf{X}_t) = a(\mathbf{X}_t) + \mathbb{E}^{\mathbb{Q}} \left(\sum_{k=0}^d \frac{B_t}{B_{t_k}} C_k^A \mathbf{1}_{\{t_k \geq t\}} | \mathbf{X}_t \right), \tag{2.1}$$

where $a(\mathbf{X}_t)$ is directly determined by the value of underlying risk factors. By construction, $A(t, \mathbf{X}_t)$ is \mathcal{F}_t -adapted. We assume that $A(t, \mathbf{X}_t)$ is square integrable and has therefore a finite variance.

We consider a risk measure denoted by $\rho(\cdot)$. For risk management, we aim to calculate $\rho(A(t, \mathbf{X}_t))$. In applications, we mainly consider the value at risk (VaR) and the expected shortfall (ES). For a confidence level $\alpha \in (0, 1)$, the VaR and ES for a continuous distribution of $A(t, \cdot)$ are defined as

$$VaR_\alpha = \max \{x \in \mathbb{R} : \mathbb{P}(A(t, \mathbf{X}_t) \leq x) \leq \alpha\}, \tag{2.2}$$

$$ES_\alpha = \frac{1}{\alpha} \int_0^\alpha VaR_\gamma d\gamma. \tag{2.3}$$

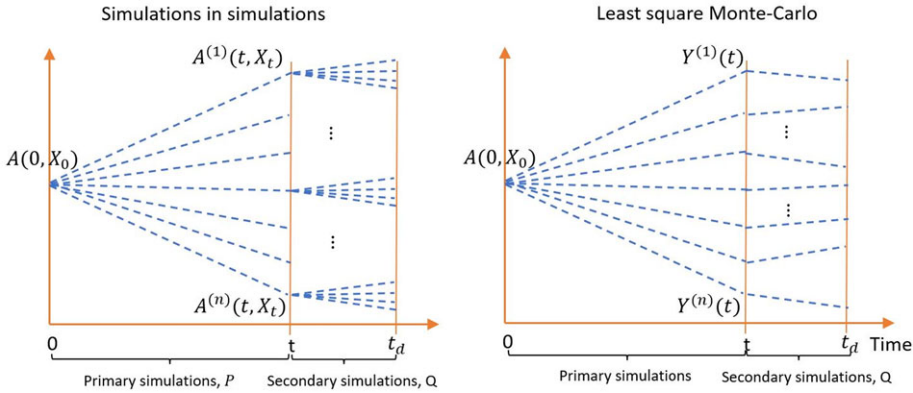


Figure 1. Simulations in simulations versus least squares Monte Carlo.

The ES also admits an equivalent representation (see McNeil *et al.*, 2015, Proposition 8.13, p. 283) that is used later for estimation:

$$\begin{aligned}
 ES_\alpha &= \mathbb{E}^{\mathbb{P}} (A(t, \mathbf{X}_t) \mid A(t, \mathbf{X}_t) \leq VaR_\alpha) \\
 &= \frac{1}{\alpha} \mathbb{E}^{\mathbb{P}} (A(t, \mathbf{X}_t) \mathbf{1}_{\{A(t, \mathbf{X}_t) \leq VaR_\alpha\}}) \\
 &\quad + VaR_\alpha \frac{1}{\alpha} (1 - \mathbb{P} (A(t, \mathbf{X}_t) \leq VaR_\alpha)).
 \end{aligned}
 \tag{2.4}$$

We draw the attention of the reader on the fact that VaR and ES are valued under the real measure. Computing the risk neutral expectation (2.1) is a challenging task because closed-form expressions are usually not available. A solution consists in evaluating $A(t, \mathbf{X}_t)$ 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 $A(t, \mathbf{X}_t)$ 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 LSMC to keep the computational time under control.

2.2. LSMC algorithm

We briefly review the LSMC method. We denote by

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

the random variable that is \mathcal{F}_{t_d} -adapted and such that $A(t, \mathbf{X}_t) = a(\mathbf{X}_t) + \mathbb{E}^{\mathbb{Q}} (Y(t) \mid \mathbf{X}_t)$. This variable is called the “response” for a given set of risk factors at time t . The LSMC method is based on property that the conditional expectation of a random variable $Y(t)$ given a random vector \mathbf{X}_t minimizes the mean squared distance between $Y(t)$ and $h(\mathbf{X}_t)$, where $h(\cdot)$ 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 \mathbf{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 \mathbf{X}_t an m -dimensional random vector, both \mathcal{F}_{t_d} -adapted. The conditional expectation $\mathbb{E}^{\mathbb{Q}} (Y(t) \mid \mathbf{X}_t)$ is equal to $h(\mathbf{X}_t) \in \mathcal{B}(\mathbb{R}^m, \mathbb{R})$, where $\mathcal{B}(\mathbb{R}^m, \mathbb{R})$ is the set of Borel measurable functions from $\mathbb{R}^m \rightarrow \mathbb{R}$, such that*

$$h(\mathbf{X}_t) = \arg \min_{h \in \mathcal{B}(\mathbb{R}^m, \mathbb{R})} \mathbb{E}^{\mathbb{Q}} ((h(\mathbf{X}_t) - Y(t))^2).
 \tag{2.5}$$

The proof of this result is reminded in Appendix A. In many real-world applications, there is no closed-form expression for the function $h(\mathbf{X}_t)$, but risk factors can be simulated under the \mathbb{P} measure. Longstaff and Schwartz (2001) assume that the unknown function $h(\cdot)$ belongs to the L^2 -space of square-integrable functions. Since L^2 is a Hilbert space, it admits a countable orthonormal basis. The function $h(\cdot)$ 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

$$\mathcal{S} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}, \tag{2.6}$$

of n realizations of $(\mathbf{X}_t, Y(t))$ and in regressing responses on risk factors. We recall that \mathbf{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 $\widehat{h}(\mathbf{x})$ of degree d_h approximating $h(\mathbf{x})$. It is estimated by least squares minimization:

$$\widehat{h}(\cdot) = \arg \min_{\widehat{h} \in \mathcal{P}_h} \left(\sum_{(\mathbf{x}_i, y_i) \in \mathcal{S}} (y_i - \widehat{h}(\mathbf{x}_i))^2 \right). \tag{2.7}$$

Let us denote by

$$\mathbf{z} = \left((x_{i_1}^{j_1} x_{i_2}^{j_2} \dots x_{i_h}^{j_h})_{\substack{j_1, \dots, j_h \in \mathbb{N} \\ j_1 + \dots + j_h \leq d_h \\ i_1, \dots, i_h \in \{1, \dots, m\}}} \right), \tag{2.8}$$

the vector of powers of risk factors up to order $d_h \in \mathbb{N}$. We define $\widehat{h}(\mathbf{x}) = \mathbf{z}^\top \boldsymbol{\beta}$ as a polynomial of order d_h where $\boldsymbol{\beta}_k$ is a real vector of same dimension as \mathbf{z} . The sample of powers of risk factors is $\{\mathbf{z}_1, \dots, \mathbf{z}_n\}$. Let us respectively denote by Z and \mathbf{y} the matrix $Z = (\mathbf{z}_j^\top)_{j \in \mathcal{S}}$ and the vector $\mathbf{y} = (y_j)_{j \in \mathcal{S}}$. Using standard arguments, the polynomial coefficients minimizing (2.7) are $\widehat{\boldsymbol{\beta}} = (Z^\top Z)^{-1} Z^\top \mathbf{y}$.

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

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

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

$$j(\alpha) = \max \left\{ k \in \{1, \dots, n\} : \frac{k}{n} \leq \alpha \right\}. \tag{2.9}$$

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

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

From Equation (2.4), the ES_α estimator is computed as follows:

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

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

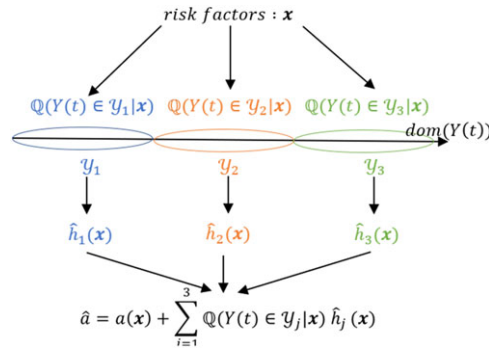


Figure 2. Graphical illustration of the LLSMC algorithm.

3. The LLSMC

3.1. General principles

As described above, it is a common practice to fit a global polynomial regression predicting responses $(y_i)_{i=1,\dots,n}$ as a function of risk factors $(x_i)_{i=1,\dots,n}$. In this article, we opt for an alternative approach based on local regressions. The method is based on a finite partitioning $(\mathcal{Y}_k)_{k=1,\dots,K}$ of the domain of $Y(t)$ (here $dom(Y(t)) = \mathbb{R}$). Let us define

$$h_k(x) = \mathbb{E}^Q(Y(t) | X_t = x, Y(t) \in \mathcal{Y}_k), k = 1, \dots, K, \tag{3.1}$$

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

$$h(x) = \mathbb{E}^Q(Y(t) | X_t = x) = \sum_{k=1}^K Q(Y(t) \in \mathcal{Y}_k | X_t = x) h_k(x). \tag{3.2}$$

Based on this decomposition, we approximate the K unknown functions $h_k(\cdot)$ by polynomial regressions $\widehat{h}_k(\cdot)$ of $Y(t) \in \mathcal{Y}_k$ on risk factors. In a second step, we use a multinomial logistic regression to estimate the probabilities $Q(Y(t) \in \mathcal{Y}_k | X_t)$ for $k = 1, \dots, K$. Figure 2 provides a graphical representation of the LLSMC algorithm when $K = 3$. Once the model is fitted, the estimated asset value \widehat{a} , for a vector of risk factors x at time t , is determined in two steps. First, we calculate the probabilities that responses belong to clusters $(\mathcal{Y}_k)_{k=1,\dots,K}$, conditionally to $X_t = x$. Next, we compute $\widehat{h}_k(x)$ which approximates the conditional expectations of responses within clusters, as defined in Equation (3.1). Finally, the estimated asset value is the sum of the function $a(x)$ and of $\widehat{h}_k(x)$ weighted by probabilities $Q(Y(t) \in \mathcal{Y}_k | X_t)$.

In practice, the simulated sample, \mathcal{S} defined in Equation (2.6), is the union of sampled risk factors, noted \mathcal{X} , and of corresponding responses \mathcal{Y} . In a first stage, we partition the sample data set $\mathcal{S} = (\mathcal{X}, \mathcal{Y})$ into $K \ll n$ subsets, denoted by $(\mathcal{S}_k)_{k=1,\dots,K}$:

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

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

3.2. The clustering algorithm

The K -means algorithm (see MacQueen, 1967 or Jain, 2010) 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 are denoted $c_u \in \mathbb{R}$,

$u = 1, \dots, K$. If $d(\cdot, \cdot)$ is the Euclidian distance, we define the clusters \mathcal{Y}_k for $k = 1, \dots, K$ as follows:

$$\mathcal{Y}_k = \{y_i : d(y_i, c_k) \leq d(y_i, c_j) \forall j \in \{1, \dots, K\}\} \quad k = 1, \dots, K. \tag{3.3}$$

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

$$\mathcal{S}_k = \{(x_i, y_i) : d(y_i, c_k) \leq d(y_i, c_j) \forall j \in \{1, \dots, K\}\} \quad k = 1, \dots, K. \tag{3.4}$$

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 $I_Y = \frac{1}{n} \sum_{i=1}^n d(y_i, g)^2$ and the interclass inertia I_c is the inertia of the cloud of centers of gravity:

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

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

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

According to the König–Huyghens theorem, the global inertia is the sum of the intraclass and interclass inertia: $I_Y = I_c + I_a$. 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 cannot be solved numerically in polynomial time (NP-hard problem, see Mahajan *et al.*, 2012) but efficient heuristic procedures exist. The most common method uses an iterative refinement technique called the K -means which is detailed in Algorithm 2, provided in Appendix B. Given an initial set of K random centroids $c_1(0), \dots, c_K(0)$, we construct a partition $\{\mathcal{Y}_1(0), \dots, \mathcal{Y}_K(0)\}$ of responses. Next, we replace the K random centroids by the K centers of gravity $(c_u(1))_{u=1, \dots, K} = (c_u(0))_{u=1, \dots, K}$ of these classes and we iterate till convergence. At each iteration, we can prove that the intraclass inertia is reduced but we do not have any warranty that the partition found by this way is a global solution. Nevertheless, this risk is limited in our approach because we cluster unidimensional data ($y_{i=1, \dots, n} \in \mathbb{R}^+$). As dimensionality increases, the distance to the nearest neighbor approaches the distance to the farthest neighbor but for one dimension data, K -means is highly efficient as discussed in Beyer *et al.* (1999). In numerical illustrations, we use an improved version of the algorithm, the K -means++ of Arthur and Vassilvitskii (2007). This enhances the quality of the resulting clusters by providing initial centroids that are well spread out across the data space. The initial centroids are selected in a probabilistic manner based on their distance from already chosen centroids. In practice, this procedure is repeated several times (20 times in this article) and we keep the partition with the lowest intraclass inertia. Notice that there exists a large variety of clustering methods (Gaussian mixture model (GMM), DBscan, spectral K -means, etc.) that are substitutable to the K -means. The impact is nevertheless limited given that we cluster unidimensional data. In the first case study, we have indeed replaced the K -means algorithm by the GMM and have not observed any significant differences.

3.3. Local regressions

After having found a partition of \mathcal{S} in $\mathcal{S}_k = (\mathcal{X}_k, \mathcal{Y}_k)$, $k = 1, \dots, K$, we approximate functions $(h_k)_{k=1, \dots, K}$ by K polynomials of order d_h , denoted by $(\widehat{h}_k(\cdot))_{k=1, \dots, K}$. Let us recall that \mathbf{z} as defined in Equation (2.8) is the vector of powers of risk factors up to $d_h \in \mathbb{N}$. We assume that $\widehat{h}_k(\mathbf{x}) = \mathbf{z}^\top \boldsymbol{\beta}_k$ is a polynomial of order d_h where $\boldsymbol{\beta}_k$ is a real vector of dimension equal to the one of \mathbf{z} . In a similar manner to LSMC, the $(\widehat{h}_k(\cdot))_{k=1, \dots, K}$ are estimated by least squares minimization over the set \mathcal{P}_h of polynomials of degree d_h :

$$\boldsymbol{\beta}_k = \arg \min_{\widehat{h}_k \in \mathcal{P}_h} \left(\sum_{(x_i, y_i) \in \mathcal{S}_k} (y_i - \widehat{h}_k(x_i))^2 \right). \tag{3.5}$$

Algorithm 1 : Summary of the LLSMC estimation procedure.

Generation of $\mathcal{S} = (\mathcal{X}, \mathcal{Y})$:

n simulations of risk factors, $\mathcal{X} = (\mathbf{x}_{i=1, \dots, n})$, at time t , under \mathbb{P} .
 Simulations of corresponding responses $\mathcal{Y} = (\mathbf{y}_{i=1, \dots, n})$ under \mathbb{Q} .

Clustering of \mathcal{S} into $\mathcal{S}_k = (\mathcal{X}_k, \mathcal{Y}_k)$ for $k = 1, \dots, K$:

Partitioning of \mathcal{Y} , into $\mathcal{Y}_{k=1, \dots, K}$ with the K -means algorithm.
 $(\mathcal{X}_k)_{k=1, \dots, K}$, subsets of corresponding risk factors.

Local regression step:

Fit of polynomial $\widehat{h}_k(\mathbf{x})$ regressing \mathcal{Y}_k on \mathcal{X}_k , for $k = 1, \dots, K$.

Logistic regression step:

Fit of polynomial $\widehat{\gamma}_k(\mathbf{x})$ for $k = 2, \dots, K$ with

$$\mathbb{Q}(Y(t) \in \mathcal{Y}_k | \mathbf{X}_t = \mathbf{x}) = \frac{e^{-\widehat{\gamma}_k(\mathbf{x})}}{1 + \sum_{j=2}^K e^{-\widehat{\gamma}_j(\mathbf{x})}}.$$

Asset valuation:

$$\widehat{a}_i = a(\mathbf{x}_i) + \sum_{j=1}^K \mathbb{Q}(Y(t) \in \mathcal{Y}_k | \mathbf{X}_t = \mathbf{x}_i) \widehat{h}_k(\mathbf{x}_i).$$

The sample of powers of risk factors is again $\{z_1, \dots, z_n\}$ and we denote by Z_k and \mathbf{y}_k the matrix $Z_k = (z_j^\top)_{j \in \mathcal{S}_k}$ and the vector $\mathbf{y}_k = (y_j)_{j \in \mathcal{S}_k}$ for $k = 1, \dots, K$. Using standard arguments, the polynomial coefficients minimizing (3.5) are $\widehat{\boldsymbol{\beta}}_k = (Z_k^\top Z_k)^{-1} Z_k^\top \mathbf{y}_k$.

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

$$\widehat{h}(\mathbf{x}) = \sum_{k=1}^K \mathbb{Q}(Y(t) \in \mathcal{Y}_k | \mathbf{X}_t = \mathbf{x}) \widehat{h}_k(\mathbf{x}), \tag{3.6}$$

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

$$\mathbb{Q}(Y(t) \in \mathcal{Y}_k | \mathbf{X}_t = \mathbf{x}) = \begin{cases} \frac{e^{-\widehat{\gamma}_k(\mathbf{x})}}{1 + \sum_{j=2}^K e^{-\widehat{\gamma}_j(\mathbf{x})}} & k = 2, \dots, K, \\ \frac{1}{1 + \sum_{j=2}^K e^{-\widehat{\gamma}_j(\mathbf{x})}} & k = 1, \end{cases} \tag{3.7}$$

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

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

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

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

and $(\boldsymbol{\zeta}_k)_{k=2, \dots, K} = \arg \max_{\gamma_k \in \mathcal{P}_\gamma} \mathcal{L}((\gamma_k)_{k=2, \dots, K})$. The full procedure to fit the LLSMC is summarized into Algorithm 1. We first simulate $\mathcal{S} = (\mathcal{X}, \mathcal{Y})$ and create clusters on which we fit local polynomial regressors by least squares minimization. Next, we estimate conditional probabilities (3.7) and compute asset values.

In a similar manner to Cheridito *et al.* (2020), we can replace the polynomial functions $\widehat{\gamma}_k(\mathbf{x})$ by neural networks or by any other machine learning regressor.

At this stage, we have not discussed yet the convergence of the LLSMC. The framework is based on the Equality (3.2) that is true whatever the number of clusters. The question of convergence arises from the approximations of $\mathbb{Q}(Y(t) \in \mathcal{Y}_k | \mathbf{x}) = \frac{e^{-\gamma_k(\mathbf{x})}}{1 + \sum_{j=2}^K e^{-\gamma_j(\mathbf{x})}}$ and $h_k(\mathbf{x})$ by a logistic and polynomial regressions. If functions $\gamma_k(\mathbf{x})$ and $h_k(\mathbf{x})$ are C^∞ , the Taylor's theorem ensures the (local) convergence when the polynomial orders d_h and $d_\gamma \rightarrow \infty$.

If a set of L^2 -orthogonal basis functions $(e_j(\mathbf{x}))_{j \in \mathbb{N}}$ from \mathbb{R}^p to \mathbb{R} is well defined, an alternative is to approximate $h_k(\mathbf{x})$ and $\gamma_k(\mathbf{x})$ by a finite combination of m basis functions:

$$\widehat{h}_k(\mathbf{x}) = \sum_{j=1}^m \alpha_{k,j} e_j(\mathbf{x}) \quad , \quad \gamma_k(\mathbf{x}) = \sum_{j=1}^m \beta_{k,j} e_j(\mathbf{x}) \quad ,$$

for $k = 1, \dots, K$. In this case, the convergence is guaranteed when $m \rightarrow \infty$. Nevertheless, finding such an orthogonal basis function $(e_j(\mathbf{x}))_{j \in \mathbb{N}}$ is a challenging task and we refer the reader to Ha and Bauer (2022) for details. This motivate us to opt for polynomial approximations.

3.4. Measures of goodness of fit

We have not 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:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \widehat{h}(\mathbf{x}_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad , \tag{3.8}$$

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

$$R_{loc}^2 = 1 - \frac{\sum_{k=1}^K \sum_{(x_i, y_i) \in \mathcal{S}_k} (y_i - \widehat{h}_k(\mathbf{x}_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad . \tag{3.9}$$

Contrary to R^2 , we may expect a R_{loc}^2 close to 1 and should exclude any models with a low R_{loc}^2 . The R^2 and R_{loc}^2 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:

$$MSE = \frac{\sum_{i=1}^n (y_i - \widehat{h}(\mathbf{x}_i))^2}{n - p} \quad , \tag{3.10}$$

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 calculate the sum of squared errors between exact values of $A(t, \mathbf{x})$ and their LLSMC estimates, $h(\mathbf{x})$ 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 $A(t, \mathbf{x}_i)$ are 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 in combining quantiles of risk factors. Let us detail this approach. We denote by $(x_{(i)}^{(k)})_{i=1, \dots, n}$ the ordered sample $(x_i^{(k)})_{i=1, \dots, n}$ of the k^{th} risk factor:

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

We select a small number of $q \in \mathbb{N}$ quantiles $(x_{j(\alpha_i)}^{(k)}, \dots, x_{j(\alpha_q)}^{(k)})$ where $(\alpha_i)_{i=1, \dots, q}$ are probabilities and $j(\alpha_i)$ is the quantile index such as defined in Equation (2.9). We repeat this operation for each $k = 1, \dots, m$. 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

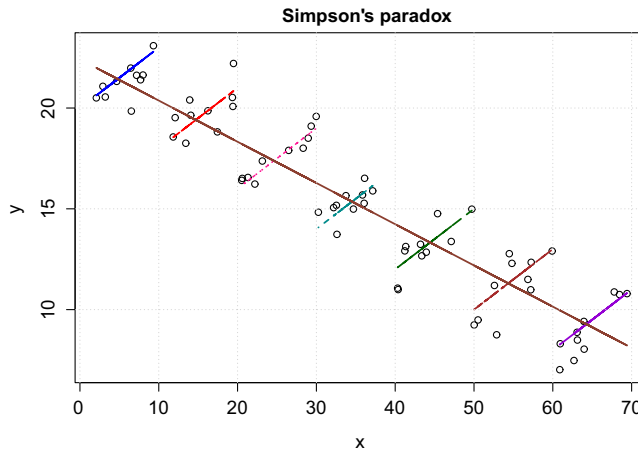


Figure 3. Illustration of the Simpson’s paradox.

$$MSE(\mathcal{V}) = \frac{1}{|\mathcal{V}|} \sum_{x \in \mathcal{V}} (A(t, x) - \widehat{h}(x))^2. \tag{3.11}$$

If the dimension of $|\mathcal{V}|$ is too large, we can select randomly a subset of \mathcal{V} . Besides the analysis of these indicators of goodness of fit, it is recommended to plot the function $\widehat{h}(x)$ to detect unexpected tail behaviors. If the asset value is computable in a large number of scenarios, we can appraise the goodness of fit by divergence measures (e.g. Kullback–Leibler) between distributions of $A(t, x)$ and $\widehat{h}(x)$, approached by simulations of x . Nevertheless, this approach is in most of cases computationally intensive.

3.5. Simpson’s paradox

It may appear counterintuitive to partition the data set using responses instead of risk factors. Two reasons motivate this choice. First, local regressions based on hard clustering of risk factors produce discontinuities in predictions, $\mathbb{E}^Q(Y(t) | X_t)$, on borders of clusters, even in a market in which all risk factors are continuous. This is clearly an undesirable feature for a model designed for risk management. Second, this prevents to observe the Simpson’s paradox (Simpson, 1951). 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 3 which compares local versus global linear regressions. Regressions on clusters of 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.

4. Application to options management in the Heston model

In this first example, we consider a financial market composed of cash and stocks with stochastic volatility. We choose this model, proposed by Heston (1993), because it is possible to benchmark LSMC and LLSMC approximated option prices to accurate ones computed by discrete Fourier transform (DFT).

4.1. The Heston model

In the Heston model, the cash 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, $(V_t)_{t \geq 0}$:

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

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

$$\begin{cases} dS_t &= \mu S_t dt + S_t \sqrt{V_t} (\rho dW_t^v + \sqrt{1 - \rho^2} dW_t^s), \\ dV_t &= \kappa (\gamma - V_t) dt + \sigma \sqrt{V_t} dW_t^v. \end{cases} \tag{4.1}$$

where $(W_t^s)_{t \geq 0}$ and $(W_t^v)_{t \geq 0}$ are independent Brownian motions 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} .

There is no analytical formula for option prices in the Heston model. Nevertheless, the characteristic function of log-returns admits a closed-form expression and we can calculate their probability density function (pdf) by DFT. European options are then valued by computing the expected discounted payoff with this pdf. Prices obtained by DFT are compared to these obtained with LSMC and LLSMC methods. Details about the pricing method are available in the Online Supplementary Materials.

4.2. Butterfly options

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

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

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

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

and its price $A(t, X_t)$, at time $t \leq T$, is equal to the \mathbb{Q} -expected discounted payoff, $A(t, X_t) = \mathbb{E}^{\mathbb{Q}}(e^{-r(T-t)} H(S_T) | \mathcal{F}_t)$. 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 with the LSMC. We will next consider a bull trap option which has an 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). We perform 10,000 primary simulations of S_t and V_t under \mathbb{P} up to $t = 1$ year. For each primary simulation, we simulate a single secondary sample path under \mathbb{Q} up to $T = 2$ years. We consider 350 steps of time per year and responses are equal to $Y(t) = e^{-r(T-t)} H(S_T)$.

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

d_h	R^2	$\sqrt{MSE(\mathcal{V})}$	\sqrt{MSE}	d.f.	Time (s)
2	0.0397	0.36	2.10	6	1.95
3	0.0451	0.57	2.10	10	1.65
4	0.0499	1.07	2.09	15	1.39
5	0.0522	2.39	2.09	21	2.00
6	0.0536	1.93	2.09	28	1.71

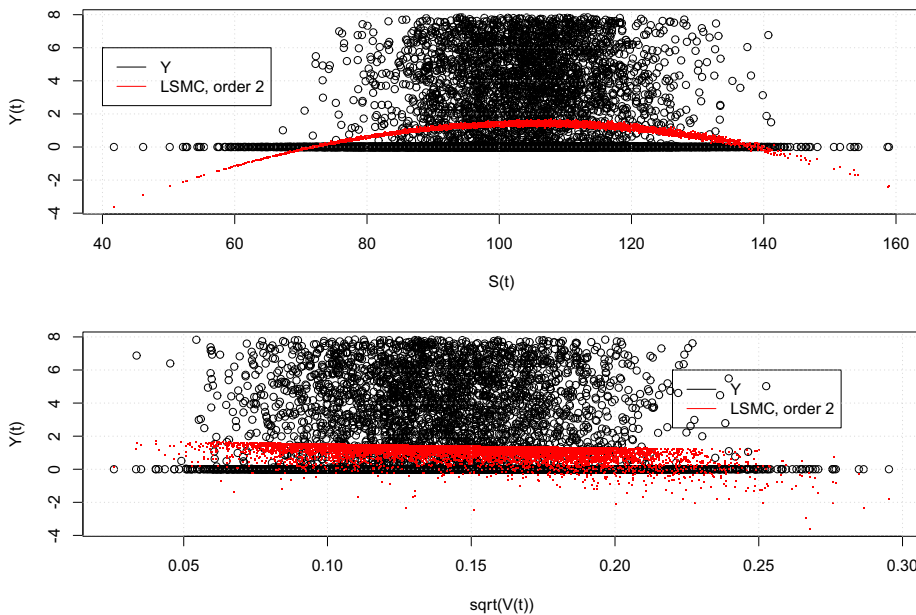


Figure 4. Simulated responses $Y(t) = e^{-r(T-t)}H(S_T)$ versus stock prices, S_t , and volatilities, $\sqrt{V_t}$. The red points are the predictions $h(\mathbf{X}_t)$ from the LSMC with a second-order polynomial regression.

4.3. Numerical analysis of LSMC and LLSMC

The upper and lower plots of Figure 4 depict the simulated responses plotted against stock prices and volatilities. The red points represent the LSMC estimates of butterfly option prices 1 year ahead, using a second-order polynomial of risk factors. These graphs expose a limitation of the conventional LSMC approach: it is unable to predict a positive option price for extremely high and low values of stock prices. This issue is particularly significant when using LSMC for computing risk measures.

Table 2 reports the R^2 , the MSE, and $MSE(\mathcal{V})$ of the LSMC, such as defined by Equations (3.8), (3.10), and (3.11). 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 observation that extreme values of risk factors tend to produce extremely high and low option prices. The exact prices of butterfly options in these 100 scenarios are computed by DFT.

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 $\mathbb{E}^Q(Y(t) | \mathbf{x})$. The R^2 s also increase with the complexity of the model. The MSE on the training set is nearly constant whatever the polynomial order, whereas the lowest $MSE(\mathcal{V})$ on the validation set is achieved with an order 2 polynomial. The

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

K	d_y	d_h	R^2	$\sqrt{MSE(\mathcal{V})}$	\sqrt{MSE}	d.f.	R^2_{loc}	Time (s)
3	2	3	0.0526	0.19	2.09	42	0.95	4.66
2	2	3	0.0525	0.20	2.09	26	0.87	4.48
5	2	3	0.0527	0.20	2.10	74	0.98	5.14
6	2	4	0.0524	0.20	2.10	120	0.99	5.59
2	2	2	0.0521	0.21	2.09	18	0.87	5.75
3	2	2	0.0524	0.21	2.09	30	0.95	4.28
4	2	2	0.0525	0.21	2.09	42	0.98	4.40
5	2	2	0.0525	0.21	2.10	54	0.98	5.05
5	2	4	0.0527	0.21	2.10	99	0.99	4.87
6	2	2	0.0525	0.22	2.10	66	0.99	5.50

LSMC is estimated in less than 2 s.¹ We next analyze the tail behavior of these approximations. This is done by plotting the function $\hat{h}(x)$. We will come back on this point later and focus first 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 4 and d_y 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 optimal goodness of fit is achieved by using three clusters, applying cubic regression to each cluster, and utilizing a second-order polynomial for the multinomial logistic regression. Upon comparing the results with the LSMC figures in Table 2, it is evident that the LLSMC method significantly reduces the mean squared error $MSE(\mathcal{V})$ on the validation data set, while the MSEs on the training set remain comparable. This indicates that the LLSMC model better replicates extremely high and low option prices. It is worth noting that the LLSMC- R^2_{loc} and LSMC- R^2 values are comparable. The runtime for the LLSMC method ranges from 4.28 to 5.59 s.

We provide in the Online Supplementary Materials the goodness of fit statistics and runtimes for the LSMC and LLSMC, computed with 100,000 simulations instead of 10,000. A comparison with Tables 2 and 3 does not reveal any significant increases of MSE and R^2 . This validates our choice to limit the number of scenarios to 10,000.

We compare the LSMC and the LLSMC with hyperparameters $K = 3$, $d_y = 2$, $d_h = 3$ (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 5 shows the segmentation of responses in 3 clusters with the K-means algorithm. The mid and lower plots show the responses and local predictions $\hat{h}_k(X_t)$ (red points) on clusters, with respect to stock prices and volatilities. In contrast to the LSMC approach, the LLSMC method does not produce significant negative responses.

Figure 6 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 $\sqrt{V_t} \in \{7\%, 14\%, 23\%\}$, and 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 middle plot illustrates the option prices under standard market conditions, while the right and left plots represent extreme volatility conditions. We observe that LSMC models of order 2 or 4 produce negative prices in the tails. To evaluate the overall accuracy of methods in these three scenarios of volatility, we provide in Table 4 the average pricing errors. This table confirms that the LLSMC globally outperforms LSMC. Tables 5 and 6 present the VaRs and TVaRs of the butterfly option for various quantiles. The LSMC models yield negative values for the lowest percentiles, whereas the LLSMC provides slightly lower VaRs and TVaRs than the LSMC of orders 3–6 for the highest percentiles. It would be interesting to compare these results to VaRs and TVaRs based on exact prices

¹ Computations are performed on a laptop with a AMD Ryzen 7 processor and 16 Gb of RAM.

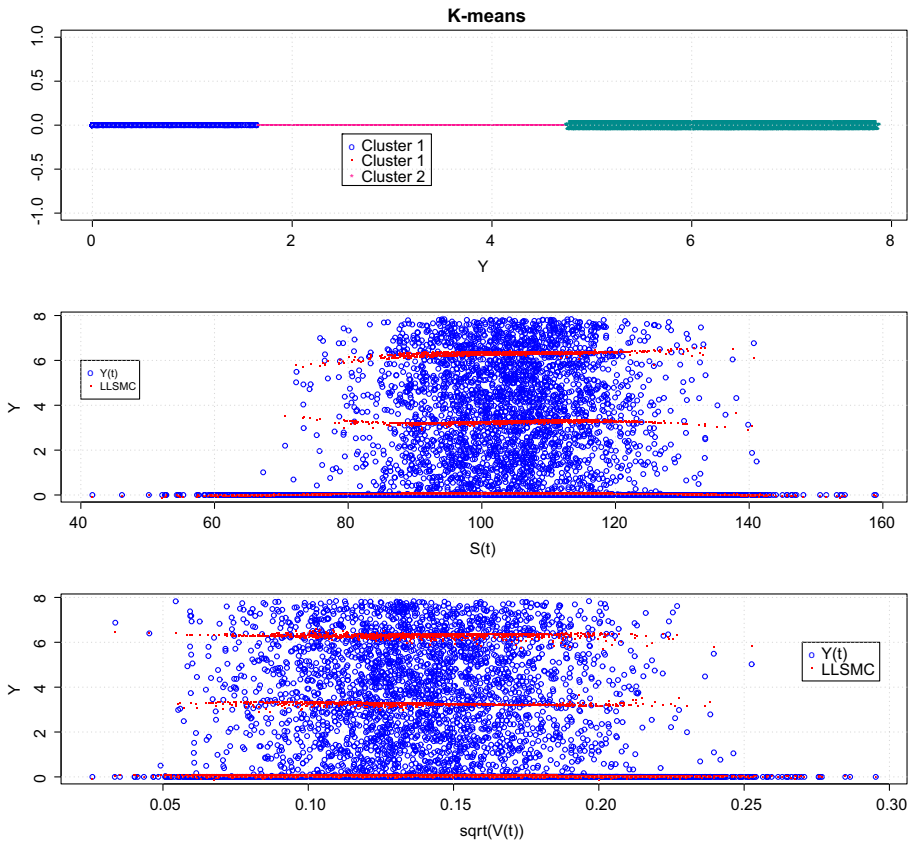


Figure 5. Upper plot: a-priori segmentation of responses in three clusters. Mid and lower plots: responses (blue points) and local regressions (red points) with respect to stock prices and volatilities.

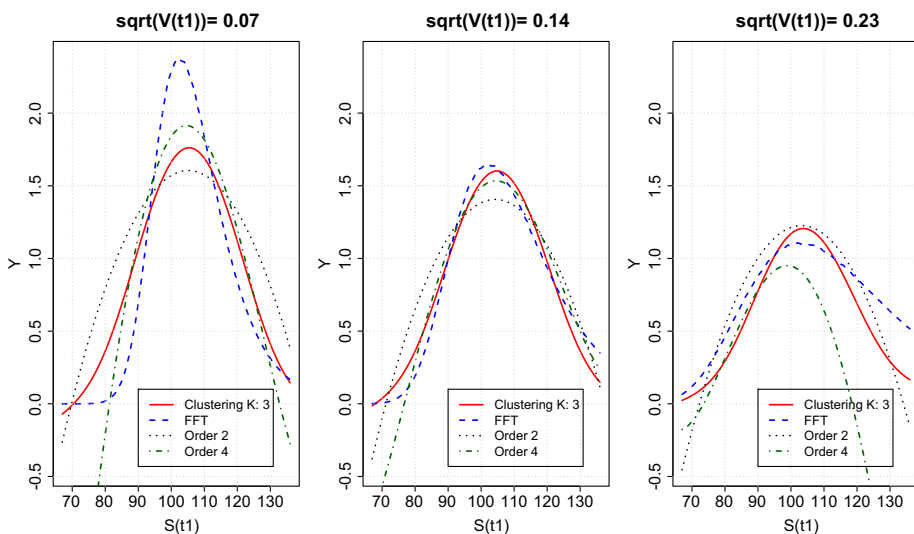


Figure 6. Butterfly options for S , ranging from 68 to 139 and volatilities $\sqrt{V_t} \in \{7\%, 14\%, 23\%$.

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

	$\sqrt{V_t} = 7\%$	$\sqrt{V_t} = 14\%$	$\sqrt{V_t} = 23\%$
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 5. VaR 1 year, LSMC model, and LLSMC.

	0.05%	0.1%	1%	5%	95%	99%	99.9%	99.95%
LLSMC, 3-2-3	-0.08	-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 6. Expected shortfall, 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	1.70	1.85	2.17	2.37
LSMC, $d_h = 5$	-0.71	-0.64	-0.35	-0.16	1.76	1.92	2.10	2.17
LSMC, $d_h = 6$	-0.77	-0.45	-0.08	0.05	1.81	1.97	2.22	2.33

computed by DFT. Unfortunately, the valuation by DFT of 10,000 butterfly options is computationally intensive. Nonetheless, such a comparison is feasible in the second case (Section 5).

We provide a detailed explanation of how the LLSMC 3-2-3 method operates, using Figure 7. The left plot shows the local regression functions $\hat{h}_k(x)$, for various stock prices S_t and a volatility of 14%. The middle plot depicts the probabilities that a pair of risk factors leads to a response belonging to the k th cluster. The first cluster explains the left and right tails of butterfly option prices. When S_t is below 80 or above 130, the response is assigned to cluster 1 with a probability exceeding 90% and the correspond function $\hat{h}_3(x)$ is nearly flat and close to zero. The probabilities of belonging to clusters 2 and 3 are relatively similar and exceed 5% for $S_t \in [80, 130]$. The right plot displays the products of the regression and probabilities functions. According to Equation (3.6), the estimated option price is obtained by summing of these three terms.

4.4. Bull trap options

We have considered a butterfly option because its payoff is not strictly increasing or decreasing function of S_T . In this paragraph, we check 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

Table 7. Average pricing errors, bull trap portfolio.

	$\sqrt{V_t} = 7\%$	$\sqrt{V_t} = 14\%$	$\sqrt{V_t} = 23\%$
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

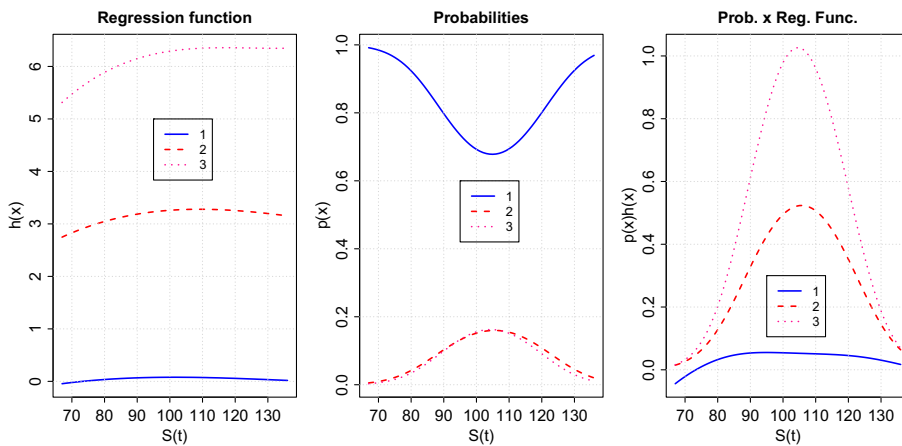


Figure 7. Plots of regression functions $h_k(x)$ for $k = 1, \dots, K$, probabilities $\mathbb{Q}(Y(t) \in \mathcal{Y}_k | X_t = x)$, and their products, for $\sqrt{V_t} = 14\%$.

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

We use the same 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 $K = 3, d_y = 2, d_h = 1$. Figure 7 compares LSMC and LLSMC bull trap options for stock prices S_t , ranging from 68 to 139, and $\sqrt{V_t} \in \{7\%, 14\%, 23\%\}$. Table 7 presents the average pricing errors in the three scenarios. These results provide confirmation that the LLSMC method achieves a higher level of overall accuracy compared to the LSMC method.

4.5. Partitioning of risk factors and the Simpson’s paradox

To conclude this section, we illustrate the Simpson’s paradox in butterfly option pricing. For this purpose, we divide the sample \mathcal{S} into $K \ll n$ subsets $(\mathcal{S}_k)_{k=1, \dots, K}$:

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

where the partition is based on the pooling of risk factors. Each cluster is defined by a centroid $c_k \in \mathbb{R}^m$ of dimension m such that

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

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

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

Table 8. R^2 , MSE , $MSE(\mathcal{V})$, and R_{loc}^2 for the \mathcal{X} -LLSMC model. $d.f.$ is the number of parameters.

K	d_h	R^2	$\sqrt{MSE(\mathcal{V})}$	\sqrt{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

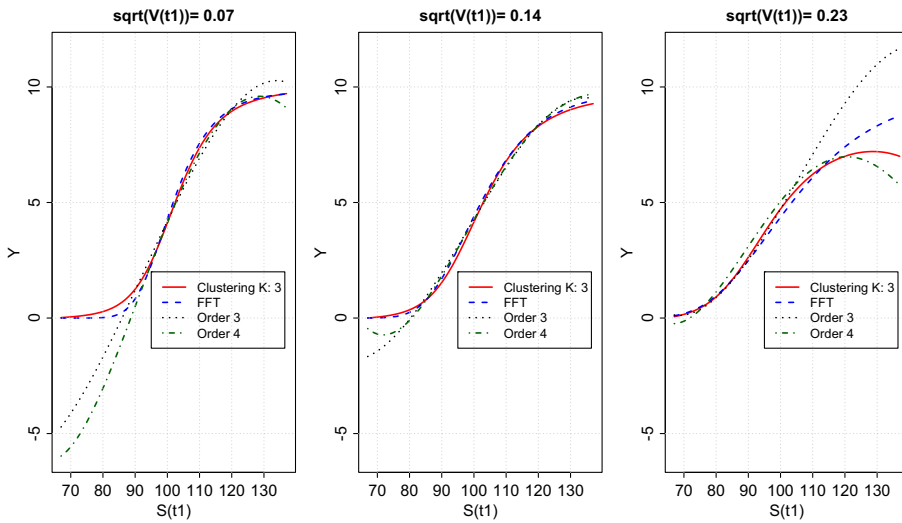


Figure 8. Bull trap options for S_i ranging from 68 to 139 and volatilities $\sqrt{V_i} \in \{7\%, 14\%, 23\%\}$.

where $\hat{h}_k \in \mathcal{P}_h$ is the set of polynomials of order d_h . As for a LLSMC regression, the \hat{h}_k are estimated by least squares minimization (Equation (2.4)). This variant of local model is denoted by \mathcal{X} -LLSMC.

Table 8 provides statistics of goodness of fit for models with K ranging from 2 to 6 and d_h from 1 to 4. Models are sorted in ascending order of $MSE(\mathcal{V})$ s and the table displays statistics of the 10 best models according to this criterion. Compared to Table 3, the \mathcal{X} -LLSMC 4-1 achieves a similar accuracy with less parameters. If we restrict our analysis to a comparison of goodness of fit statistics, both the \mathcal{X} -LLSMC and LLSMC methods appear to be equivalent for computing VaR or TVaR. Plotting the \mathcal{X} -LLSMC regression function leads to another conclusion. Figures 8 and 9 compares \mathcal{X} -LLSMC 4-1 and DFT Butterfly prices for S_i ranging from 68 to 139, and $\sqrt{V_i} \in \{7\%, 14\%, 23\%\}$. We observe that local regressions based on clusters of risk factors produce discontinuities in predicted responses on borders of clusters. Second, we identify local trends not relevant to the global slope of price curves. These two elements disqualify the \mathcal{X} -LLSMC for risk management purposes.

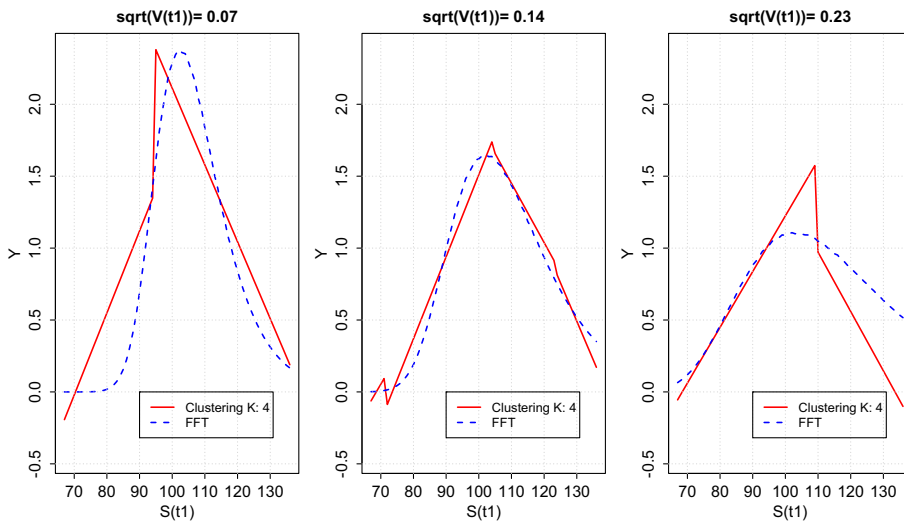


Figure 9. Butterfly options for S_t , ranging from 68 to 139 and volatilities $\sqrt{V_t} \in \{7\%, 14\%, 23\%\}$.

5. Application to life insurance management

In this second example, we evaluate the performances of LSMC and LLSMC in assessing the risk of a participating pure endowment. The following subsection provides a brief overview of the characteristics of this product in a market with three risk factors. Since the contract has a closed-form valuation formula, we will compare the approximate VaRs and TVaRs obtained from LSMC and LLSMC with the exact values.

5.1. A participating pure endowment

We consider a joint life insurance and financial market. The stock price indices, 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:

$$\begin{pmatrix} dS_t \\ dr_t \\ d\mu_{x+t} \end{pmatrix} = \begin{pmatrix} \mu S_t \\ \kappa_r (\gamma_r(t) - r_t) \\ \kappa_\mu (\gamma_x(t) - \mu_{x+t}) \end{pmatrix} dt + \begin{pmatrix} S_t \sigma_S & 0 & 0 \\ 0 & \sigma_r & 0 \\ 0 & 0 & \sigma_x(t) \end{pmatrix} \Sigma \begin{pmatrix} dW_t^{(1)} \\ dW_t^{(2)} \\ dW_t^{(3)} \end{pmatrix}. \tag{5.1}$$

$(W_t^{(1)}, W_t^{(2)}, W_t^{(3)})_{t \geq 0}$ are independent Brownian motions. $\mu, \kappa_r, \kappa_\mu, \sigma_S$ 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 fitted to term structures of interest and mortality rates. The initial age of the insured is denoted by $x \in [0, x_{\max})$, $x_{\max} > 0$. 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 the Online Supplementary Materials. The matrix Σ is the (upper) Cholesky decomposition of the correlation matrix and is such that

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

where $\rho_{Sr}, \rho_{S\mu}$ and $\rho_{r\mu} \in (-1, 1)$. This model incorporates the correlation between financial and mortality shocks, which can be relevant in the context of events such as a pandemic like Covid-19. It is worth noting that Ha and Bauer (2022) have also explored a similar framework, although our model differs in terms of mortality dynamics, where we incorporate mean reversion with age-dependent volatility. Additionally, our focus is on benchmarking the LLSMC algorithm using a participating pure endowment contract, for which we derive a closed-form expression for its price. The contract which is subscribed by an individual

of age x guarantees a payout at the expiry date (T) equal to the maximum between a capital C_T and the stock indices S_T , in the event of survival. However, the benefit is 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:

$$\begin{aligned}
 V_t &= \mathbb{E}^{\mathbb{Q}} \left(e^{-\int_t^T r_s ds} \mathbf{1}_{\{\tau \geq T\}} (C_T + (S_T - C_T)_+ - (S_T - C_M)_+) \mid \mathcal{F}_t \right) \\
 &= \mathbf{1}_{\{\tau \geq t\}} C_T \mathbb{E}^{\mathbb{Q}} \left(e^{-\int_t^T (r_s + \mu_{x+s}) ds} \mid \mathcal{F}_t \right) \\
 &\quad + \mathbf{1}_{\{\tau \geq t\}} \mathbb{E}^{\mathbb{Q}} \left(e^{-\int_t^T (r_s + \mu_{x+s}) ds} ((S_T - C_T)_+) \mid \mathcal{F}_t \right) \\
 &\quad - \mathbf{1}_{\{\tau \geq t\}} \mathbb{E}^{\mathbb{Q}} \left(e^{-\int_t^T (r_s + \mu_{x+s}) ds} ((S_T - C_M)_+) \mid \mathcal{F}_t \right).
 \end{aligned}
 \tag{5.2}$$

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 impact on our conclusions). The instantaneous return of the stock indices is r_t under \mathbb{Q} . The zero-coupon bond, the survival probabilities, and the pure endowment² are, respectively, defined by the following expectations:

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

The model being affine, we can easily derive the closed-form expressions of these products. In the remainder of this article, we adopt the following 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 $B_y(\cdot, \cdot)$:

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

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

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

In order to match the initial yield curve of zero-coupon bonds, the function $\gamma_r(T)$ satisfies the following relationship:

$$\gamma_r(T) = -\frac{1}{\kappa_r} \partial_T^2 \ln P(0, T) - \partial_T \ln P(0, T) + \frac{\sigma_r^2}{2\kappa_r^2} (1 - e^{-2\kappa_r T}),$$

where $-\partial_T \ln P(0, T)$ is the instantaneous forward rate. For a given initial mortality curve ${}_tP_x$, the function $\gamma_x(u)$ is such that

$$\gamma_x(T) = -\frac{1}{\kappa_\mu} \partial_T^2 \ln {}_tP_x - \partial_T \ln {}_tP_x + \frac{\alpha^2 e^{2\beta x}}{2\kappa_\mu(\kappa_\mu + \beta)} (e^{2\beta T} - e^{-2\kappa_\mu T}).$$

²The pure endowment pays one monetary unit at time T if the individual is alive.

Bond prices, survival probabilities, and endowments admit closed-form expressions presented in the next proposition.

Proposition 5.1. *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*

$$P(t, T) = \exp \left(-r_t B_{\kappa_r}(t, T) - (\partial_t \ln P(0, t)) B_{\kappa_r}(t, T) + \ln \frac{P(0, T)}{P(0, t)} \right) \times \exp \left(-\frac{\sigma_r^2}{4\kappa_r} \left((1 - e^{-2\kappa_r t}) B_{\kappa_r}(t, T)^2 \right) \right). \tag{5.3}$$

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

$$\begin{aligned} {}_t p_{x+t} &= \exp \left(-\mu_{x+t} B_{\kappa_\mu}(t, T) - (\partial_t \ln {}_t p_x) B_{\kappa_\mu}(t, T) + \ln \frac{{}_t p_x}{{}_t p_x} \right) \times \\ &\exp \left(\frac{\alpha^2 e^{2\beta(x+T)}}{2\kappa_\mu^2} (B_{2\beta}(t, T) - 2B_{2\beta+\kappa_\mu}(t, T) + B_{2\beta+2\kappa_\mu}(t, T)) \right) \times \\ &\exp \left(\frac{\alpha^2 e^{2\beta x}}{2\kappa_\mu(\kappa_\mu + \beta)} (e^{2\beta T} B_{2\beta+\kappa_\mu}(t, T) - e^{2\beta T} B_{2\beta}(t, T)) \right) \times \\ &\exp \left(\frac{\alpha^2 e^{2\beta x}}{2\kappa_\mu(\kappa_\mu + \beta)} (e^{-2\kappa_\mu t} B_{2\kappa_\mu}(t, T) - e^{-\kappa_\mu(T+t)} B_{\kappa_\mu}(t, T)) \right), \end{aligned} \tag{5.4}$$

whereas the pure endowment contracts are valued by:

$${}_t E_t = \mathbf{I}_{\{\tau \geq t\}} {}_t p_{x+t} P(t, T) \times \exp \left(\frac{\sigma_r \epsilon_{r\mu} \alpha e^{\beta(x+T)}}{\kappa_\mu \kappa_r} (B_\beta(t, T) - B_{\kappa_\mu+\beta}(t, T) - B_{\kappa_r+\beta}(t, T) + B_{\kappa_\mu+\kappa_r+\beta}(t, T)) \right). \tag{5.5}$$

The sketch of the proof is provided in the Online Supplementary Materials. In order to obtain a closed-form expression of the saving contract (5.2), we perform a change of measure using as Radon–Nikodym derivative:

$$\frac{d\mathbb{F}}{d\mathbb{Q}} \Big|_T = \mathbb{E}^{\mathbb{Q}} \left(\frac{d\mathbb{F}}{d\mathbb{Q}} \mid \mathcal{F}_T \right) = \frac{e^{-\int_0^T (r_s + \mu_{x+s}) ds}}{\mathbb{E}^{\mathbb{Q}} \left(e^{-\int_0^T (r_s + \mu_{x+s}) ds} \mid \mathcal{F}_0 \right)}. \tag{5.6}$$

Taking advantage the log-normality of S_T under the \mathbb{F} -measure, we can derive a closed-form expression for the call options embedded in the benefits, such as defined in Equation (5.2).

Proposition 5.2. *The log-return under the \mathbb{F} -measure is log-normal*

$$\ln(S_T/S_t) \sim N(\mu_{\mathbb{F}}(t, T), v_{\mathbb{F}}(t, T)^2)$$

with a mean and variance, respectively, given by

$$\left\{ \begin{aligned} \mu_{\mathbb{F}}(t, T) &= -\frac{\sigma_r^2}{2} \int_t^T B_{\kappa_r}(u, T)^2 du - \sigma_r \epsilon_{r\mu} \int_t^T \sigma_x(u) B_{\kappa_r}(u, T) B_{\kappa_\mu}(u, T) du \\ &\quad - \frac{\sigma_S^2(T-t)}{2} - \sigma_S \sigma_r (\epsilon_{S_r} \epsilon_{rr} + \epsilon_{S\mu} \epsilon_{r\mu}) \int_t^T B_{\kappa_r}(u, T) du \\ &\quad - \sigma_S \epsilon_{S\mu} \int_t^T \sigma_x(u) B_{\kappa_\mu}(u, T) du, \\ v_{\mathbb{F}}(t, T)^2 &= \sigma_S^2(T-t) + \sigma_r^2 \int_t^T B_{\kappa_r}(u, T)^2 du \\ &\quad + 2\sigma_S \sigma_r (\epsilon_{S_r} \epsilon_{rr} + \epsilon_{S\mu} \epsilon_{r\mu}) \int_t^T B_{\kappa_r}(u, T) du. \end{aligned} \right. \tag{5.7}$$

If we adopt the following notations:

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

Table 9. Model parameters and features of the contract.

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

the embedded call options in the participating pure endowment (5.2) are valued by:

$$\begin{aligned}
 & \mathbf{1}_{\{\tau \geq t\}} \mathbb{E}^{\mathbb{Q}} \left(e^{-\int_t^T (r_s + \mu_{x+s}) ds} ((S_T - C)_+) \mid \mathcal{F}_t \right) \\
 &= {}_T E_t \left[\frac{S_t e^{\mu_{\mathbb{P}}(t,T) + \frac{v_{\mathbb{P}}(t,T)^2}{2}}}{P(t, T)} \Phi(-d_1(t, T)) - C_T \Phi(-d_2(t, T)) \right].
 \end{aligned}
 \tag{5.8}$$

We refer to the Online Supplementary Materials for a proof of this result. The exact value of the pure endowment is obtained by combining Equations (5.5) and (5.8). This allows us to compare LSMC and LLSMC approximated values to exact prices in the next subsection.

5.2. Numerical analysis

We fit a Nelson–Siegel model to the Belgian state yield curve³ on the 23/11/22. Initial survival probabilities are described by a Makeham’s model adjusted to male Belgian mortality rates.⁴ Details are provided in the Online Supplementary Materials. 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. The correlations $\rho_{S\mu}$ and $\rho_{r\mu}$ are set to -5% and 0%. Parameter estimates are reported in Table 1.

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 :

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

Expectations and variances are approached by empirical averages and variances of the simulated sample. The contract features are reported in Table 1. We simulate 10,000 primary scenarios and a single secondary response per scenario,

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

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

The plots of Figure 10 show responses versus stock prices, interest, and mortality rates. The red dots correspond to LSMC estimates of the endowment 1 year ahead, with a second-order polynomial of risk factors.

³Source: national bank of Belgium (www.nbb.be).

⁴Source: Human mortality database (www.mortality.org).

Table 10. R^2 , MSE , $MSE(\mathcal{V})$, and runtimes of regressions of Y_t on X_t in the LSMC model. \sqrt{EMSE} is the MSE valued with analytical prices. *d.f.* is the number of parameters.

d_h	R^2	$\sqrt{MSE(\mathcal{V})}$	\sqrt{MSE}	\sqrt{EMSE}	d.f.	Time (s)
2	0.3815	1.98	11.48	1.14	10	12.83
3	0.3874	2.09	11.35	0.97	20	11.58
4	0.3864	2.14	11.29	0.96	35	11.42
5	0.3875	2.90	11.40	0.80	56	11.56
6	0.3955	3.93	11.25	0.91	84	11.58

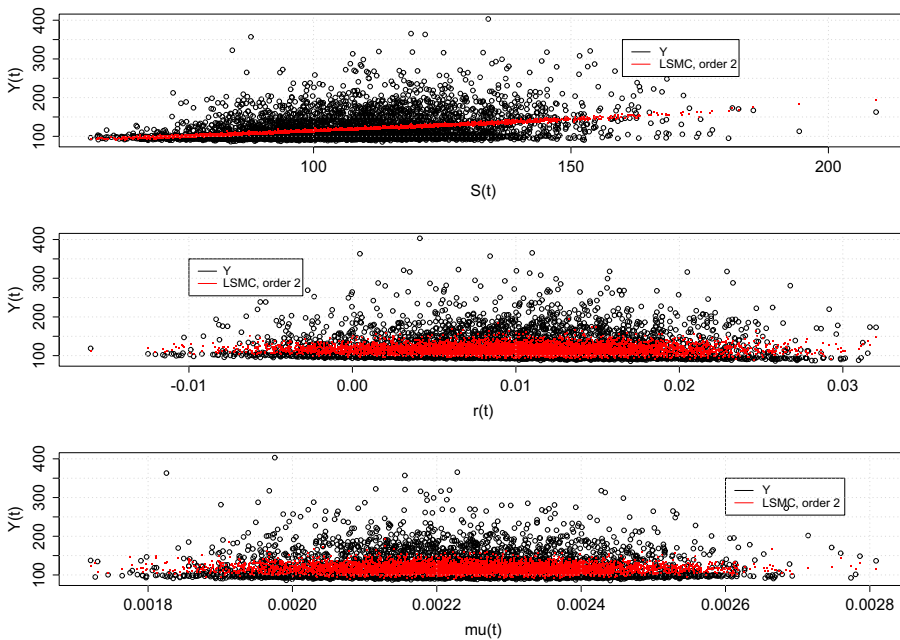


Figure 10. Simulated responses $Y(t) =$ versus stock prices S_t , r_t and $\mu_x + t$. The red dots are the predictions $h(X_t)$ from the LSMC with a second-order polynomials regression.

Table 10 reports the R^2 , the MSE, and $MSE(\mathcal{V})$ of the LSMC, such as defined by Equations (3.8), (3.10), and (3.11). 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.

Table 10 provides statistics about LSMC polynomial regressions of order d_h from 2 to 6. The R^2 s slightly increase with the complexity of the model. The $MSE(\mathcal{V})$ on the validation set is minimized by a polynomial of second degree. The LSMC is estimated in 12 s. 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_v 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 optimal goodness of fit is achieved by using 5–3 clusters, a square or cubic regression on each cluster, and a cubic multinomial logistic regression. Compared to the LSMC, the LLSMC reduces the $MSE(\mathcal{V})$ and the EMSE by more than half, whereas MSE on the training set remains comparable. This indicates that the LLSMC model provides a

Table 11. R^2 , MSE , $MSE(\mathcal{V})$, R^2_{loc} , and runtimes for the LLSMC model. \sqrt{MSE} , exact is the MSE valued with analytical prices. d.f. is the number of parameters.

K	d_y	d_h	R^2	$\sqrt{MSE(\mathcal{V})}$	\sqrt{MSE}	\sqrt{EMSE}	d.f.	R^2_{loc}	Time (s)
5	3	2	0.3952	0.65	11.28	0.47	130	0.97	76.44
4	3	3	0.3952	0.69	11.28	0.41	140	0.96	46.72
3	3	2	0.3952	0.76	11.25	0.37	70	0.93	71.14
4	3	2	0.3952	0.77	11.26	0.41	100	0.95	59.80
5	2	2	0.3918	0.79	11.29	0.70	90	0.97	68.32
5	3	3	0.3953	0.81	11.31	0.47	180	0.97	68.00
2	3	2	0.3946	0.84	11.23	0.37	40	0.88	72.97
5	3	1	0.3949	0.86	11.26	0.49	100	0.97	78.56
4	2	3	0.3923	0.87	11.30	0.71	110	0.96	55.33
4	2	2	0.3923	0.88	11.27	0.71	70	0.95	68.57

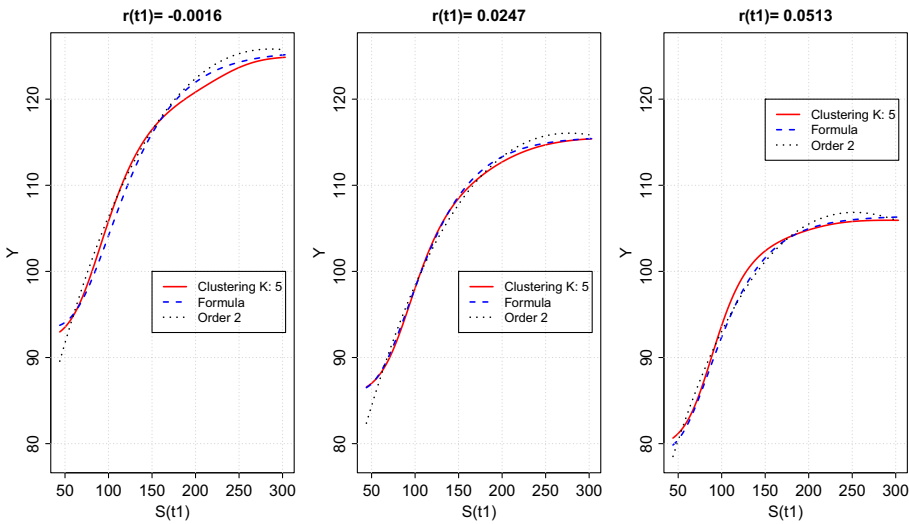


Figure 11. Endowment prices for S_t , ranging from 40 to 310, $r_t \in \{-0.16\%, 2.47\%, 5.13\% \}$ and $\mu_{x+t} = 0.0017$.

better fit. The computational times range from 47 to 79 s depending on the model setting. We provide in the Online Supplementary Materials the goodness of fit statistics and runtimes for LSMC and LLSMC, computed with 100,000 simulations instead of 10,000. We do not observe significant differences in MSE and R^2 . This validates our choice to set the number of scenarios to 10,000.

We next compare the LSMC of order 3 and the LLSMC with hyperparameters $K = 5, d_y = 3, d_h = 2$ as this setting leads to a low $MSE(\mathcal{V})$ and a high R^2_{loc} . Figure 6 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 $r_t \in \{-0.16\%, 2.47\%, 5.13\% \}$, and 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 12, are slightly lower on average than those of the LSMC when $r_t \in \{-0.16\%, 2.47\% \}$. In particular, the LLSMC better fits extremely low values. This will be confirmed by the comparison of VaRs and TVaRs.

Figure 12 displays VaRs and TVaRs computed with the LLSMC 5-3-2, the LSMC of order 3, and analytical values. Tables 13 and 14 report the relative spread between VaR/TVaRs computed with

Table 12. Average pricing errors for the three cases presented in Figure 11.

	$r_t = -0.16\%$	$r_t = 2.47\%$	$r_t = 5.13\%$
LLSMC, 5-3-2	0.9441	0.3353	0.7314
LSMC, order 3	1.2602	0.8872	0.7606

Table 13. VaR 5 years, LSMC model, and LLSMC.

VaR	Relative errors in %							
	0.1%	1%	2%	3%	97%	98%	99%	99.9%
LLSMC 5-3-2	0.73	0.28	0.25	0.09	-0.35	-0.43	-0.51	-0.6
LLSMC 4-3-3	0.35	0.44	0.32	0.12	-0.19	-0.2	-0.27	-0.64
LLSMC 3-3-2	1.32	1.11	0.85	0.76	-0.09	-0.11	-0.22	-0.57
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 14. Expected shortfall, 5 years, LSMC model, and LLSMC.

Tail VaR	Relative errors in %							
	0.1%	1%	2%	3%	97%	98%	99%	99.9%
LLSMC 5-3-2	0.37	0.54	0.5	0.38	-0.41	-0.44	-0.49	-0.6
LLSMC 4-3-3	-0.19	0.45	0.49	0.39	-0.22	-0.28	-0.39	-1.37
LLSMC 3-3-2	2.13	1.45	1.26	1.14	-0.19	-0.24	-0.32	-0.51
LSMC $d_h = 2$	-8.41	-1.16	-0.6	-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

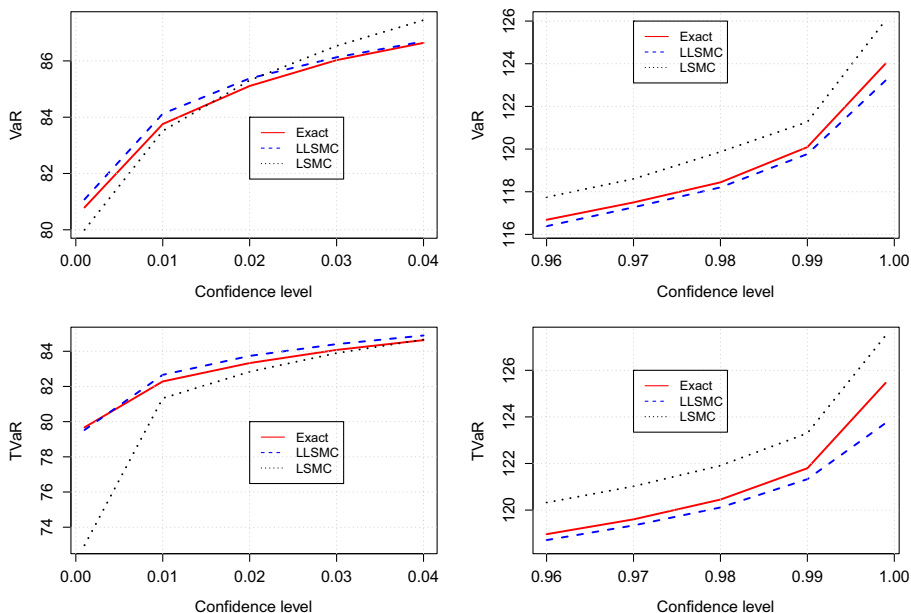


Figure 12. Lower/upper VaRs and TVaRs, LLSMC 5-3-2, LSMC of order 3.

approximated and exact analytical prices. These results emphasize that LLSMC provides a more accurate estimate of VaR/TVaRs. In particular, the LSMC's inability to closely replicate extreme values results in a significant divergence of TVaR for very low or high confidence levels.

6. Conclusions

This article proposes a straightforward and powerful extension of the LSMC method for risk management by incorporating local and logistic regressions. The novelty of our approach lies in segmenting the data set based on responses rather than risk factors. We then employ polynomial regressions $(\hat{h}_k(\cdot))_{k=1,\dots,K}$ for each cluster. For a given vector of risk factors \mathbf{x} , $\hat{h}_k(\mathbf{x})$ approximates the market value of future cash flows when the corresponding responses belong to the k th cluster. The unconditional value of future cash flows is obtained by summing $\hat{h}_k(\mathbf{x})$, weighted by probabilities of cluster membership estimated through a multinomial logistic regression.

We validate the LLSMC method through two case studies. In both cases, numerical experiments demonstrate that LLSMC achieves superior accuracy compared to LSMC across a broader range of scenarios. We also observe that LLSMC yields fewer erratic prices for lower and upper quantiles of risk factors. This confirms that LLSMC is better suited for computing risk measures such as VaR and tail value at risk (Tail VaR) compared to LSMC. Furthermore, LLSMC provides a higher level of interpretability. We also compare LLSMC to a local method that relies on partitioning risk factors (\mathcal{X} -LLSMC). Our findings reveal that this approach suffers from Simpson's paradox, where \mathcal{X} -LLSMC prices exhibit local trends that are not relevant in the global context.

This work opens avenues for further research. First, the LLSMC algorithm is likely to be more efficient than LSMC for estimating the solvency capital requirement within the Solvency II framework. Second, we can consider replacing local polynomial approximations with local machine learning regressions. This hybrid procedure would be particularly suitable for managing a large number of risk factors. Finally, LLSMC can be adapted to price American options by discretizing the time horizon and estimating backward local regressions at each time step.

Acknowledgments. 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.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/asb.2023.25>.

References

- Arthur, D. and Vassilvitskii, S. (2007) k-means++: The advantages of careful seeding. *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*.
- Bacinello, A.R., Biffis, E. and Millossovich, P. (2009) Pricing life insurance contract with early exercise features. *Journal of Computational and Applied Mathematics*, **233**(1), 27–35.
- Bacinello, A.R., Biffis, E. and Millossovich, P. (2010) Regression-based algorithms for life insurance contracts with surrender guarantees. *Quantitative Finance*, **10**(9), 1077–1090.
- Bauer, D., Reuss, A. and Singer, D. (2012) On the calculation of the solvency capital requirement based on nested simulations. *ASTIN Bulletin*, **42**, 453–99.
- Becker, S., Cheridito, P. and Jentzen, A. (2020) Pricing and hedging American-Style options with deep learning. *Journal of Risk and Financial Management*, **13**(7), 158. <https://www.mdpi.com/1911-8074/13/7/158>.
- Beyer, K., Goldstein, J., Ramakrishnan, R. and Shaft, U. (1999) When is “nearest neighbor” meaningful? *Database Theory ICDT'99*, pp. 217–235.
- Cheridito, P., Ery, J. and Wüthrich, M.V. (2020) Assessing asset-liability risk with neural networks. *Risks* **2020**, **8**(1), 16.
- Clement, E., Lamberton, D. and Protter, P. (2002) An analysis of a least squares regression method for American option pricing. *Finance and Stochastics*, **6**, 449–471.
- Floryszczak, A., Le Courtois, O. and 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.

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

Glasserman, P. and Yu, B. (2004) Number of paths versus number of basis functions in American option pricing. *Annals of Applied Probability*, **14**(4), 2090–2119.

Ha, H. and Bauer, D. (2022) A least-squares Monte Carlo approach to the estimation of enterprise risk. *Finance and Stochastic*, **26**, 417–459.

Hainaut, D. (2022) *Continuous Time Processes for Finance. Switching, Self-exciting Fractional and Other Recent Dynamics*. Springer, Bocconi University Press.

Heston, S. (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, **6**, 327–343

Hörig, M. and 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/>.

Hörig, M., Leitschkis, M., Murray, K. and 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>.

Jain, A.K. (2010) Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, **31**(8), 651–666.

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

Longstaff, F.A. and Schwartz, E.S. (2001) Valuing American options by simulation: A simple least square approach. *The Review of Financial Studies*, **14**(1), 113–147.

MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations. *Fifth Berkeley Symposium on Mathematics, Statistics and Probability*, University of California Press, pp. 281–297

Mahajan, M., Nimbhorkar, P. and Varadarajan, K. (2012) The planar K-means problem is NP-hard. *Theoretical Computer Science*, **442**, 13–21.

McNeil, A.J., Frey, R. and Embrechts, P. (2015) *Quantitative Risk Management: Concepts, Techniques and Tools*, Revised Edition. Princeton University Press.

Molnar, C. (2023) *Interpretable Machine Learning. A guide for Making Black Box Models Explainable*. <https://christophm.github.io/interpretable-ml-book/index.html>.

Moreno, M. and 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.

Pelsser, A. and Schweizer, J. (2016) The difference between LSMC and replicating portfolio in insurance liability modeling. *European Actuarial Journal*, **6**, 441–494.

Simpson, E.H. (1951) The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, **13**(2), 238–241.

Stentoft, L. (2004) Assessing the least squares Monte-Carlo approach to American option valuation. *Review of Derivatives Research*, **7**(2), 107–128.

Appendix A

Proof of Proposition 2.1. Let us denote by $v_{X,Y}(x, y)$ the joint probability density function (pdf) of $(X_t, Y(t))$ and by $v_X(x)$, $v_Y(y)$ the marginal pdfs. According to the Bayes’ rule, the conditional density of $Y(t)|X_t$ is such that

$$v_{X,Y}(x, y) = v_{Y|X}(y|x)v_X(x)$$

and the expectation in (2.5) may be rewritten as

$$\mathbb{E}^Q \left((h(X_t) - Y(t))^2 \right) = \int_{dom(X)} \int_{dom(Y)} (h(x) - y)^2 v_{Y|X}(y|x) dy v_X(x) dx.$$

The function $h(X_t)$ minimizes (2.5) if and only if

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

which is achieved for $h(x) = \mathbb{E}^Q (Y(t)|X_t = x)$.

end

Appendix B

Algorithm 2 : Algorithm for K -means clustering.

Initialization:

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

Main procedure:

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

Assignment step:

For $i = 1$ to n

(1) Assign (x_i, y_i) to a cluster $\mathcal{S}_k(e)$ and y_i to $\mathcal{Y}_k(e)$ where $k \in \{1, \dots, K\}$

$\mathcal{S}_k(e) = \{(x_i, y_i) : d(y_i, c_k(e)) \leq d(y_i, c_j(e)) \forall j \in \{1, \dots, K\}\}$,

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

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
