


RESEARCH

Open Access



Model order reduction for the simulation of parametric interest rate models in financial risk analysis

Andreas Binder¹, Onkar Jadhav^{1,2*}  and Volker Mehrmann²

*Correspondence:

jadhav@math.tu-berlin.de

¹MathConsult GmbH,
Altenbergerstraße 69, A-4040 Linz,
Austria

²Institut für Mathematik, MA 4-5, TU
Berlin, Str. des 17. Juni 136, D-10623,
Berlin, Germany

Abstract

This paper presents a model order reduction approach for large scale high dimensional parametric models arising in the analysis of financial risk. To understand the risks associated with a financial product, one has to perform several thousand computationally demanding simulations of the model which require efficient algorithms. We establish a model reduction approach based on a variant of the proper orthogonal decomposition method to generate small model approximations for the high dimensional parametric convection-diffusion-reaction partial differential equations. This approach requires to solve the full model at some selected parameter values to generate a reduced basis. We propose an adaptive greedy sampling technique based on surrogate modeling for the selection of the sample parameter set. The new technique is analyzed, implemented, and tested on industrial data of a floater with cap and floor under the Hull–White model. The results illustrate that the reduced model approach works well for short-rate models.

MSC: 35L10; 65M06; 91G30; 91G60; 91G80

Keywords: Financial risk analysis; Short-rate models; Convection-diffusion-reaction equation; Finite difference method; Parametric model order reduction; Proper orthogonal decomposition; Adaptive greedy sampling; Packaged retail investment and insurance-based products

1 Introduction

It is essential to be aware of the financial risk associated with an invested product. The risk analysis of financial instruments often requires the valuation of such instruments under a wide range of future market scenarios. The market scenarios (e.g., values of equity prices, stock indices, foreign exchange rates, interest rates) are then input parameters in a valuation function that delivers the fair value of such financial instruments. Examples of such risk analysis tasks are the calculation of Value-at-Risk (VaR) or the expected shortfall to estimate worst-case scenarios of financial holdings [1–3]. Three different VaR approaches are widely used: (1) The parametric VaR assumes a given distribution of underlying risk factors (e.g., equity prices or processes of stock indices). The future parametric value of a financial instrument can then be estimated via its sensitivities with respect to the un-

© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

derlying risk factors. (2) The historical VaR approach applies historic day-to-day changes of the underlying prices to the current status (today) to obtain an empiric distribution of tomorrow's underlying prices. When a data range of one year (approx. 250 business days) is used, the value of the financial holding has to be calculated 250 times. (3) Monte Carlo VaR applies a Monte Carlo simulation to obtain a wide range of underlying values. Again, to calculate the values of the instrument, a valuation function has to be used. To obtain reasonably accurate results, Monte Carlo VaR typically needs thousands or tens of thousands of valuations.

Credit and debt value adjustments (CVA, DVA) are estimates of possible losses when a counterparty in a financial portfolio defaults [4]. The difference between the credit-risk-free portfolio value and the true portfolio value in the presence of credit risk depends on the future liabilities of the counterparty and its default risk. For the generation of exposure scenarios, Monte Carlo simulation is the method of choice. Frequently, the positions between two market participants (often banks) contain many different risk factors (e.g., instruments in different currencies). In such cases, exposure calculation is a high-dimensional problem.

Packaged retail and insurance-based investment products (PRIIPs) are financial instruments that are *packaged* and offered to retail investors. We will come back to the definition of *packaged* and their categorization in Sect. 2. In order to make PRIIPs from different manufacturers more comparable concerning their risks and returns, the European regulation (EU) 1286/2014 requires manufacturers of PRIIPs to supply key information documents (KIDs) to possible retail investors that are easy to read and to understand [5]. The commission delegated regulation (EU) 2017/653 formulates the details of how the risk and the possible returns of a PRIIP have to be calculated [6]. There are four different categories of PRIIPs. In this paper, we concentrate on Category 3 instruments, for which at least 10,000 market data scenarios have to be generated, and the PRIIP has to be valued under these scenarios. While the time horizon (at which the distribution of future values is to be simulated) of VaR calculations is typically quite short (one business day or ten business days, sometimes until the end of the fiscal quarter), the time horizon of PRIIPs is the *recommended holding period* (RHP) of the PRIIP. This is typically in the range of years or even (e.g., in the case of structured life insurance contracts) decades. As a consequence, the calculation of PRIIPs key figures is computationally much more challenging than a VaR calculation of the same instrument. The commission delegated regulation for PRIIPs does not prescribe which valuation functions should be used. Therefore, one can use any appropriate valuation function, namely plain discounting cash flows, closed-form solutions of Black–Scholes or Black 76 [7], numerical schemes for finance-related partial differential equations (PDEs) [1], or Monte Carlo and Quasi Monte Carlos Schemes for stochastic differential equations arising from e.g., Libor market models [8, 9]. When a financial instrument is equipped with optionalities like callabilities (for the issuer) or putabilities (for the investor) or when the instrument is path-dependent (e.g., autocallable instruments, target redemption notes), the valuation functions are more complex PDEs [10–12].

In this paper, the financial instruments are evaluated via the dynamics of short-rate models [13], based on convection-diffusion-reaction partial differential equations. The choice of the short-rate model depends on the underlying financial instrument. Some of the prominent financial models are the one-factor Hull–White model [14], the shifted Black–Karasinski model [15], and the two-factor Hull–White model [16]. These models

are calibrated based on market data like *yield curves* that generate a high dimensional parameter space [13]. In short, to perform the risk analysis, the financial model needs to be solved for such a high dimensional parameter space, and this requires efficient algorithms. For instruments with high complexity and long-time horizons, computing times of minutes for a single valuation are not unusual.

Thus, in this paper, we establish a parametric model order reduction (MOR) approach based on a variant of the proper orthogonal decomposition approach, which significantly reduces the overall computation time [17, 18]; (the POD method is also known as the Karhunen–Loève decomposition [19] or principal component analysis [20] in statistics). This nonlinear MOR approach is computationally feasible [21] as it determines low dimensional linear (or affine) subspaces [22, 23] via a truncated singular value decomposition of a snapshot matrix [24], that is computed by simulating the high dimensional parametric model for a small number of pre-selected training parameter values. The question of how to select these parameters is often the most challenging part of the process. Typically, one uses a fixed or uniform sampling [25], but this may neglect vital regions in the case of high dimensional parameter spaces.

We present a greedy sampling algorithm to determine the best suitable parameter set, see also [26–28]. The basic idea is to select the parameters at which the error between the reduced and the high dimensional model is maximal. We compute the snapshots using these parameters and thus obtain the best suitable reduced basis to generate a fairly accurate reduced model. Since the calculation of the relative error between the full and reduced model is expensive, we also develop error estimators similar to [29, 30]. The classical greedy sampling algorithm then picks optimal parameters which yield the highest values for the error estimator. However, it is not computationally feasible to compute an error estimator for the entire parameter space, so we locate the parameters using an adaptive method by constructing a surrogate model for the error estimator, similar to [26, 28]. We construct this surrogate model based on the principal component regression (PCR) technique [31].

To summarize, this paper presents an approach to select the most prominent parameters (or scenarios) for which we solve the high dimensional model to obtain a reduced order model. Thus, instead of performing thousands of costly simulations, we perform few expensive computations and simulate the remaining scenarios with the help of the reduced model. In this work, we focus on the one-factor Hull–White model [32], and we illustrate the results with a numerical example of a floater with cap and floor [33]. For more details, see the technical report [34].

The paper is organized as follows. Section 2 discusses European regulations of PRIIPs and presents a flowchart explaining the calculation steps for a Category 3 PRIIP. Section 3 explains the construction of a short-rate model. The projection-based model order reduction technique is presented in Sect. 4. The selection of optimal sampling parameters based on the classical greedy approach is presented in Sect. 4.1 and based on the adaptive greedy method in Sects. 4.2 and 4.3. Numerical results for the example of a floater are presented in Sect. 5. Appendix B addresses the simulation procedure for yield curves and the calibration of model parameters based on these simulated yield curves is described in Appendix C. In Appendix D a finite difference method for the Hull–White model is presented.

2 The European regulation of PRIIPs

The keywords of PRIIPs are packaged and retail investment. When a financial instrument manufacturer has to decide if a key information document (KID) has to be provided, the following two questions have to be answered with *yes*. Is the instrument offered to retail investors, and is it packaged? If the instrument is only offered to professional investors, e.g., banks, insurance firms, capital management firms, then no KID has to be provided. Examples are fix-to-floating swaps or caps between banks. If the same fix-to-floating swap is offered to a retail investor, a KID has to be supplied. Again, if an instrument is offered to a retail investor but not packaged, no KID is necessary. Examples of such instruments are bonds with fixed-rate interest, shares of companies traded at exchanges.

A KID must contain the following points on not more than three A4 pages:

1. General information on the product: Name of the product, name of the manufacturer, and its web address, date of the KID, name of the legal authorities for supervision.
2. What is this product?
3. What are the risks and what could I get in return?
4. What happens if [the name of the PRIIP manufacturer] is unable to pay out?
5. What are the costs?
6. How long should I hold it and can I take my money out early?
7. How can I complain?
8. Other relevant information.

To answer the question No. 3, one has to perform several thousand costly numerical simulations of the underlying instrument using an appropriate financial model. The commission delegated regulation (EU) 2017/653 formulates the details of how the risk and the possible returns of a PRIIP have to be calculated. The delegated regulation organizes PRIIPs into four categories.

Category 1: PRIIPs include instruments where the investor can lose more than the amount invested, but also options, futures, swaps instruments where prices are available on a basis less than monthly. A Libor cap is one of the examples.

Category 2: These instruments are either direct or synthetic linear investments for which at least two years of daily prices are available. (When prices are quoted less than daily, the time series has to be longer.) An example would be a synthetic exchange-traded fund based on DAX or the S&P500 index.

Category 3: These instruments are nonlinear investments, again with a time series of prices as with Category 2 instruments. Examples are many structured financial instruments, e.g., a floater with caps and floors.

Category 4: In these instruments the values depend on factors not observed in the market. Examples are life insurance instruments.

In this paper, we deal with the Category 3 instruments. The calculation for a Category 3 instrument starts by determining a recommended holding period. In most cases, manufacturers recommend holding the instrument until maturity. At the recommended holding period, a VaR equivalent volatility (VEV) has to be determined to calculate a market risk indicator (1 is very low risk, 7 is very high risk). Market risk classifications (MRM) are shown in Table 1. For the market risk indicator, 10,000 different simulations of a market risk factor (e.g., yield curve) have to be performed. For each of those scenarios, the in-

Table 1 Market risk classification

Market risk classification	VaR equivalent volatility
1	$VEV < 0.5\%$
2	$0.50\% \leq VEV < 5.00\%$
3	$5.00\% \leq VEV < 12.0\%$
4	$12.0\% \leq VEV < 20.0\%$
5	$20.0\% \leq VEV < 30.0\%$
6	$30.0\% \leq VEV < 80.0\%$
7	$VEV > 80.0\%$

strument has to be valued. From the distribution of the resulting values, the VaR can be calculated. The VaR shall be the value of the PRIIP at a confidence level of 97.5% at the end of the recommended holding period discounted to the present date using the expected risk-free discount factor (DF) from the present date to the end of the recommended holding period. Let $P_{97.5}$ be the value of the PRIIP at a confidence level of 97.5% at the end of the recommended holding period. Furthermore, we calculate $VaR_{\text{price space}}$ by multiplying $P_{97.5}$ by the discount factor

$$DF = \exp(r(t_0, t_k)(t_k - t_0)),$$

where $r(t_0, t_k)$ is the yield at the tenor point t_k , t_0 is today and t_k is the maturity, i.e., the recommended holding period T . If the PRIIP prices are log-normal, then the distribution of its logarithm can be given as $\mathcal{N}(-VEV^2/2)T, VEV^2 T)$. Using this lognormal assumption, the 97.5 percentile of VaR in price space would be

$$\begin{aligned} VaR_{\text{price space}} &= \exp\left(-\frac{T}{2}VEV^2 + z_{\alpha=0.025}VEV\sqrt{T}\right), \\ \ln(VaR_{\text{price space}}) &= -\frac{T}{2}VEV^2 - 1.96 \times VEV\sqrt{T}, \\ VEV^2 \cdot T + 3.92 \times VEV\sqrt{T} + 2 \times \ln(VaR_{\text{price space}}) &= 0, \end{aligned}$$

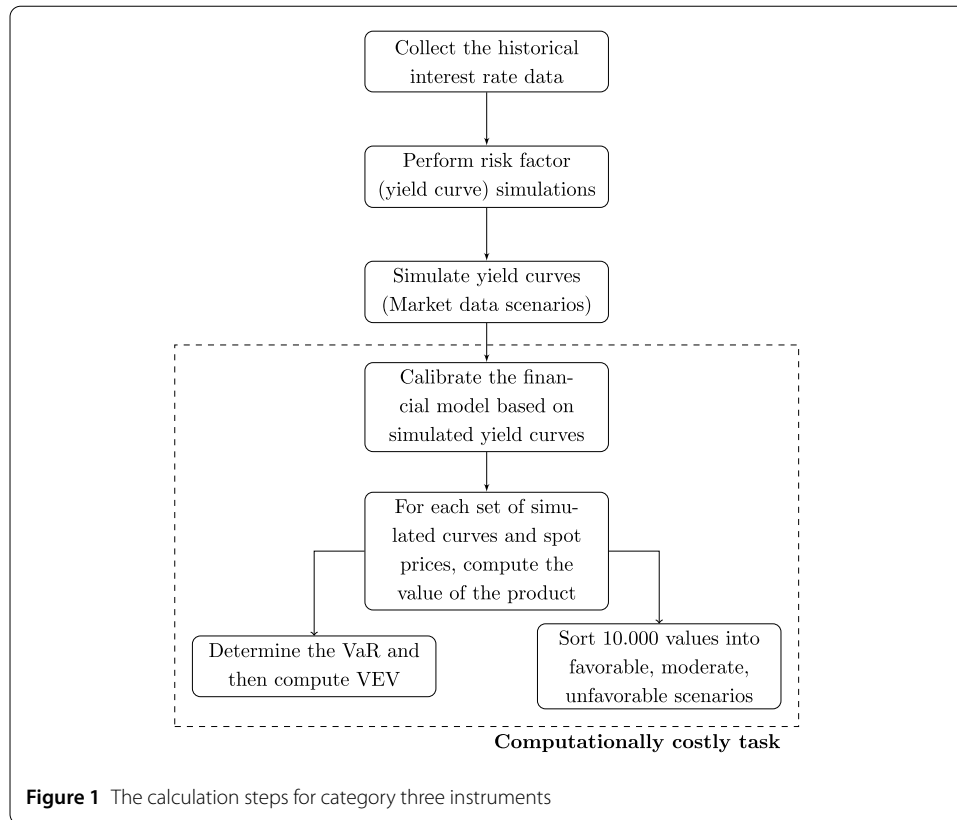
where $z_{\alpha=0.025} = -1.96$. We solve this quadratic equation for VEV

$$\begin{aligned} VEV &= \frac{-3.92\sqrt{T} \pm \sqrt{(3.92)^2 T - 8T \times \ln(VaR_{\text{price space}})}}{2T} \\ &= \frac{-1.96 \pm \sqrt{3.842 - 2 \times \ln(VaR_{\text{price space}})}}{\sqrt{T}}, \end{aligned}$$

and consider only the positive value

$$VEV = \frac{\sqrt{(3.842 - 2 \times \ln(VaR_{\text{price space}}))} - 1.96}{\sqrt{T}}.$$

Then the instrument is simulated for 10,000 different market scenarios, and the 10,000 values are sorted into favorable, moderate, and unfavorable performance scenarios, which are the values at 90th, 50th, and 10th percentile, respectively. If the recommended holding period is less than one year, then the performance scenarios are calculated only at the RHP. However, if the recommended holding period is longer than one year, then the performance scenarios at the intermediate holding periods (1 year, RHP/2, RHP) have to be



calculated. To be consistent with Category 2 instruments, the measure used after the valuation date is the risk-free measure and it is the physical measure before. Flowchart 1 shows the calculation steps for Category 3 instruments. For more details, see the joint committee of the European Supervisory Authorities flowcharts for the risk and reward calculations in the PRIIPs KID [35].

The calibration of the financial model generates a high dimensional parameter space which can be computationally costly, when no analytic solutions are available, e.g., in the case of the Black Karasinski model. Since we need to value the financial instrument for such a high dimensional parameter space, we establish a parametric model order reduction approach.

3 The mathematical model

The management of interest rate risks, i.e., the control of change in future cash flows due to the fluctuations in interest rates is of great importance. Especially, the pricing of products based on the stochastic nature of the interest rate creates the necessity for mathematical models for an underlying financial instrument. In Appendix A we describe the derivation of such a model and consider in particular the Hull–White model [32, 36] which is given by the convection-diffusion-reaction type partial differential equation (PDE) [37] of the form

$$\frac{\partial V}{\partial t} + \underbrace{(a(t) - b(t)r(t)) \frac{\partial V}{\partial r}}_{\text{convection}} + \underbrace{\frac{1}{2} \sigma(t)^2 \frac{\partial^2 V}{\partial r^2}}_{\text{diffusion}} - \underbrace{r(t)V}_{\text{reaction}} = 0, \quad (1)$$

where V is an *interest rate instrument* with *maturity* T and $r(t)$ is the *short rate* satisfying a stochastic differential equation (SDE)

$$dr(t) = (a(t) - b(t)r(t)) dt + \sigma(t) dW(t),$$

with time-dependent parameters $a(t)$, $b(t)$, and $\sigma(t)$ and white noise $W(t)$. We solve this SDE for an underlying instrument V by converting it into the PDE defined by (1). The detailed derivation of the PDE is given in Appendix A. The regulation does not prescribe how to obtain $b(t)$ and $\sigma(t)$; thus, considering them constants should lead to more robust results than time-dependent parameters, see also [38], so in our approach we consider the resulting *robust Hull–White model* with constant $b(t) = b$ and $\sigma(t) = \sigma$ and time-dependent parameter $a(t)$ for yield curve simulation and parameter calibration, which determine the average direction in which the short-rate $r(t)$ moves. Our results can, however, be extended to the more general case. According to the PRIIPs regulation, for this we have to perform at least $s = 10,000$ yield curve simulations. We construct a simulated yield curve matrix (15)

$$Y = \begin{bmatrix} y_{11} & \cdots & y_{1m} \\ \vdots & \vdots & \vdots \\ y_{s1} & \cdots & y_{sm} \end{bmatrix} \in \mathbb{R}^{s \times m},$$

which is then used to calibrate the parameter $a(t)$. We obtain s different piecewise constant parameters $a_\ell(t)$, which change their values $\alpha_{\ell,i}$ only at the m tenor points. We incorporate these in a matrix (19)

$$\mathcal{A} = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \vdots & \vdots & \vdots \\ \alpha_{s1} & \cdots & \alpha_{sm} \end{bmatrix}.$$

Appendix B presents a detailed procedure for the yield curve simulation and the calibration of $a(t)$ based on simulated yield curves is described in Appendix C.

To solve (1) numerically, we apply an upwind scheme for the convection term [39], and a standard second order finite difference scheme for the diffusion term, combined with a trapezoidal rule [40] for the time discretization, for more details see the technical report [34] and Appendix D. This discretization is a parametric high dimensional model of the form

$$A(\rho_\ell(t))V^{n-1} = B(\rho_\ell(t))V^n, \quad (2)$$

with a given terminal vector, and matrices $A(\rho_\ell) \in \mathbb{R}^{M \times M}$, and $B(\rho_\ell) \in \mathbb{R}^{M \times M}$. We call this the *full model* for the model reduction procedure in the next section. We solve (2) by propagating backward in time. Here again $\ell = 1, \dots, s = 10,000$, m is the total number of tenor points, and we need to solve this system at each time step n with an appropriate boundary condition and a known terminal value for the underlying instrument. Altogether we have a parameter space \mathcal{P} of size $10,000 \times m$ to which we now apply model reduction.

4 Parametric model order reduction

To perform the parametric model reduction for system (2), we employ Galerkin projection onto a low dimensional subspace via

$$\tilde{V}^n = QV_d^n, \quad (3)$$

where the columns of $Q \in \mathbb{R}^{M \times d}$ represent the reduced-order basis with $d \ll M$, V_d^n is a vector of reduced coordinates, and $\tilde{V}^n \in \mathbb{R}^M$ is the solution in the n th time step obtained using the reduced order model. For the Galerkin projection we require that the residual of the reduced state

$$p^n(V_d^n, \rho_\ell) = A(\rho_\ell)QV_d^{n-1} - B(\rho_\ell)QV_d^n$$

is orthogonal to the reduced basis matrix Q , i.e.,

$$Q^T p^n(V_d^n, \rho_\ell) = 0, \quad (4)$$

so that by multiplying $p^n(V_d^n, \rho_\ell)$ with Q^T , we get

$$\begin{aligned} Q^T A(\rho_\ell)QV_d^{n-1} &= Q^T B(\rho_\ell)QV_d^n, \\ A_d(\rho_\ell)V_d^{n-1} &= B_d(\rho_\ell)V_d^n, \end{aligned} \quad (5)$$

where $A_d(\rho_\ell) \in \mathbb{R}^{d \times d}$ and $B_d(\rho_\ell) \in \mathbb{R}^{d \times d}$ are the parameter dependent reduced matrices.

We obtain the Galerkin projection matrix Q in (3) based on a proper orthogonal decomposition (POD) approach, which generates an optimal order orthonormal basis Q in the least square sense that is independent of the parameter space \mathcal{P} and we do this by the method of snapshots. The snapshots are nothing but the state solutions obtained by simulating the full model for selected parameter groups. We assume that we have a training set of parameter groups $\rho_1, \dots, \rho_k \in [\rho_1, \rho_s]$. We compute the solutions of the full model for this training set and combine them in a snapshot matrix $\hat{V} = [V(\rho_1), V(\rho_2), \dots, V(\rho_k)]$. The POD method computes

$$\text{POD}(\hat{V}) := \underset{Q}{\operatorname{argmin}} \frac{1}{k} \sum_{i=1}^k \|V_i - QQ^T V_i\|^2,$$

for an orthogonal matrix $Q \in \mathbb{R}^{M \times d}$ via a *truncated SVD*, see [41],

$$\hat{V} = \Phi \Sigma \Psi^T = \sum_{i=1}^k \Sigma_i \phi_i \psi_i^T,$$

where ϕ_i and ψ_i are the left and right singular vectors of the matrix \hat{V} respectively, and Σ_i are the singular values. The truncated SVD computes only the first k columns of the matrix Φ . The optimal projection subspace Q then consists of d left singular vectors ϕ_i known as *POD modes*. The dimension d of the subspace Q is chosen such that we get a good approximation of the snapshot matrix. According to [22], large singular values correspond to the main characteristics of the system, while small singular values give only

Algorithm 1 Parametric POD**Input:** Parameter groups $\rho_s = \{a(t), b, \sigma\}$, energy level EL , number of samples k .**Output:** Projection matrix Q .Choose sample parameter groups ρ_1, \dots, ρ_k .FOR $i = 1$ to k .Solve the full order model for the parameter group ρ_i and determine $V(\rho_i)$.

END

Construct snapshot matrix $\hat{V} = [V(\rho_1), \dots, V(\rho_k)]$.Compute leading singular values and vectors of \hat{V} using truncated SVD $\hat{V} = \Phi \Sigma \Psi^T$.Set $\Xi = \text{diag}(\Sigma) / \text{sum}(\text{diag}(\Sigma))$.FOR $j = 1$ to $\text{length}(\Sigma)$. $\tilde{\Xi} = \text{sum}(\Xi(1:j)) \times 100$.IF $\tilde{\Xi} > EL$ then set $d = j$.

END

 $Q = [\phi_1 \cdots \phi_d]$.

small perturbations of the overall dynamics. The relative importance of the i th POD mode of the matrix \hat{V} is determined by the *relative energy* Ξ_i of that mode

$$\Xi_i = \frac{\Sigma_i}{\sum_{i=1}^k \Sigma_i}. \quad (6)$$

If the sum of the energies of the generated modes is 1, then these modes can be used to reconstruct a snapshot matrix completely [42]. In general, the number of modes required to generate the complete data set is significantly less than the total number of POD modes [43]. Thus, a matrix \hat{V} can be accurately approximated by using POD modes whose corresponding energies sum to almost all of the total energy. Thus, we choose only d out of k POD modes to construct $Q = [\phi_1 \cdots \phi_d]$ which is a parameter independent projection space based on (6).

We summarize the procedure in Algorithm 1, where we denote by $V(\rho_i)$ the solution of the full model for a parameter set ρ_i , by \hat{V} the snapshot matrix, by Φ (Ψ) the matrix of left (right) singular vectors, associated with the diagonal matrix of singular values Σ , and by Ξ_j the relative energy of the j th POD mode.

It is evident that the quality of the reduced model strongly depends on the selection of parameter groups ρ_1, \dots, ρ_k that are used to compute the snapshots. Hence it is essential to introduce an efficient sampling technique for the parameter space. We could consider standard sampling techniques, like uniform sampling or random sampling [44]. However, these techniques may neglect vital regions within the parameter space. As an alternative, a greedy sampling method has been suggested in the framework of model order reduction, see [26–28].

4.1 Greedy sampling method

The greedy sampling technique selects the parameter groups at which the error between the reduced model and the full model is maximal. We compute the snapshots using these parameter groups in such a way that we obtain an optimal reduced basis Q . Let $\|e\| = \frac{\|V - \tilde{V}\|}{\|V\|}$

Algorithm 2 Classical greedy sampling

Input: Maximum number of iterations I_{\max} , maximum number of parameter groups c , parameter space \mathcal{P} , tolerance ε_{tol} .

Output: Galerkin projection matrix Q .

Choose first parameter group $\rho_1 = [(a_{11}, \dots, a_{1m}), b, \sigma]$ from \mathcal{P} .

Solve the full order model for the parameter group ρ_1 and store the results in V_1 .

Compute a truncated SVD of the matrix V_1 and construct Q_1 .

Randomly select a set of c parameter groups $\hat{\mathcal{P}} = \{\rho_1, \rho_2, \dots, \rho_c\} \subset \mathcal{P}$.

FOR $i = 2$ to I_{\max}

FOR $j = 1$ to c

 Compute a reduced model for parameter group ρ_j with reduced basis Q_{i-1} .

 Compute the error estimator $\varepsilon(\rho_j)$.

END

 Compute $\rho_I = \operatorname{argmax}_{\rho \in \hat{\mathcal{P}}} \varepsilon(\rho)$.

 IF $\varepsilon(\rho_I) \leq \varepsilon_{\text{tol}}$, then $Q = Q_{i-1}$, STOP.

 Simulate the full model for the parameter group ρ_I and store the result in V_i .

 Construct a snapshot matrix \hat{V} by concatenating the solutions V_ℓ for $\ell = 1, \dots, i$.

 Compute a truncated SVD of the matrix \hat{V} and construct Q_i .

END

be the relative error between the reduced and full model and set

$$\rho_I = \operatorname{argmax}_{\rho \in \mathcal{P}} \|e\|.$$

At each greedy iteration $i = 1, \dots, I_{\max}$, the greedy sampling algorithm selects the optimal parameter group ρ_I that maximizes the relative error $\|e\|$. However, the computation of the relative error $\|e\|$ is computationally costly as it entails the solution of the full model. Thus, usually, the relative error is replaced by error bounds or the residual $\|p\|$. However, in some cases, it is not possible to obtain error bounds, see [28–30]. Let ε be the error estimator, i.e., in our case the norm of the residual. At each iteration $i = 1, \dots, I_{\max}$, the classical greedy sampling algorithm chooses the parameter group as the maximizer

$$\rho_I = \operatorname{argmax}_{\rho \in \mathcal{P}} \varepsilon(\rho). \quad (7)$$

The algorithm is initiated by selecting a parameter group ρ_1 from the parameter set \mathcal{P} and computing a reduced basis Q_1 as in Sect. 4. Choosing a pre-defined parameter set $\hat{\mathcal{P}}$ of cardinality c randomly from the set \mathcal{P} , at each point of $\hat{\mathcal{P}}$ the algorithm determines a reduced order model with reduced basis Q_1 and then computes error estimator values, $\varepsilon(\rho_j)_{j=1}^c$. The parameter group in $\hat{\mathcal{P}}$ at which the error estimator is maximal is then selected as the optimal parameter group ρ_I . Then the full order model is simulated for this parameter group and the snapshot matrix \hat{V} is updated. Finally, a new reduced basis is obtained by computing a truncated singular value decomposition of the updated snapshot matrix, as in Algorithm 1. These steps are then repeated for I_{\max} iterations or until the maximal value of the error estimator is lower than the specified tolerance ε_{tol} . The procedure for the classical greedy approach is summarized in Algorithm 2.

The classical greedy sampling method computes an inexpensive *a posteriori* error estimator for the reduced model. However, it is not feasible to calculate the error estimator values for the entire parameter space \mathcal{P} . To address this, the classical greedy sampling technique chooses the pre-defined parameter set $\hat{\mathcal{P}}$ randomly as a subset of \mathcal{P} . Random sampling is designed to represent the whole parameter space \mathcal{P} , but there is no guarantee that $\hat{\mathcal{P}}$ will reflect the complete space \mathcal{P} , since the random selection may neglect parameter groups corresponding to the most significant error. These observations motivate us to design a new criterion for the selection of the subset $\hat{\mathcal{P}}$.

Another drawback of the classical greedy sampling technique is that we have to specify the maximum error estimator tolerance ε_{tol} . The error estimator usually depends on some error bound, which may not be tight or may not exist. To overcome this drawback, following ideas in [28], we establish a strategy to construct a surrogate model for the error as a function of the error estimator and we use this error model to control the convergence of the greedy sampling algorithm.

4.2 Adaptive greedy sampling method

To overcome drawbacks of the classical greedy sampling approach, we implemented the adaptive sampling approach which selects the parameter groups adaptively at each iteration of the greedy procedure, using an optimized search based on surrogate modeling. We construct a surrogate model $\bar{\varepsilon}$ of the error estimator to approximate the error estimator ε over the entire parameter space. Further, we use this surrogate model to select the parameter groups $\hat{\mathcal{P}}_k = \{\rho_1, \dots, \rho_{c_k}\}$ with $c_k < c$, where the values of the error estimator are highest. For each parameter group within the parameter set $\hat{\mathcal{P}}_k$, we determine a reduced model and compute the values of the error estimator. Then the process repeats itself until the total number of parameter groups reaches c , resulting in the desired parameter set $\hat{\mathcal{P}}$.

The first stage of the adaptive greedy sampling algorithm computes the error estimator over a randomly selected parameter set $\hat{\mathcal{P}}_0$ of cardinality c_0 . The algorithm uses these error estimator values $\{\varepsilon_i\}_{i=1}^{c_0}$ to build a surrogate model $\bar{\varepsilon}_0$ and locates the c_k parameter groups corresponding to the c_k maximal values of the surrogate model. This process repeats itself for $k = 1, \dots, K$ iterations until the total number of parameter groups reaches c . Finally, the optimal parameter group ρ_l is the one that maximizes the error estimator within the parameter set

$$\hat{\mathcal{P}} = \hat{\mathcal{P}}_0 \cup \hat{\mathcal{P}}_1 \cup \hat{\mathcal{P}}_2 \cup \dots \cup \hat{\mathcal{P}}_K.$$

Thus, at the k th iteration, a surrogate model $\bar{\varepsilon}_k$ is constructed that approximates the error estimator over the entire parameter space \mathcal{P} . There are different choices to build a surrogate model [44]. In this paper, we use the principal component regression (PCR) technique. Suppose $\hat{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_{c_k}) \in \mathbb{R}^{c_k \times 1}$ is the vector of error estimator values at the k th iteration. Since the parameters b and σ are assumed constant, we build a surrogate model with the parameter $a(t)$ only. Let $\hat{\mathcal{P}}_k = [\rho_1, \dots, \rho_{c_k}] \in \mathbb{R}^{c_k \times m}$ be the matrix composed of c_k parameter groups at the k th iteration. The rows of the matrix $\hat{\mathcal{P}}_k$ represent c_k parameter vectors, while the m columns represent m tensor points for the parameter vector $a(t)$. We can fit a simple multiple regression model as

$$\hat{\varepsilon} = \hat{\mathcal{P}}_k \cdot \eta + \text{err}, \quad (8)$$

where $\eta = [\eta_1, \dots, \eta_m]$ is an array containing regression coefficients and err is an array of residuals. The least square estimate of η is obtained as

$$\hat{\eta} = \underset{\eta}{\operatorname{argmin}} \|\hat{\varepsilon} - \hat{\mathcal{P}}_k \cdot \eta\|_2^2 = \underset{\eta}{\operatorname{argmin}} \left\| \hat{\varepsilon} - \sum_{i=1}^m \rho_i \eta_i \right\|_2^2.$$

If c_k is not much larger than m , then the model may give weak predictions due to the risk of over-fitting for the parameter groups which are not used in model training. Also, if c_k is smaller than m , then the least square approach cannot produce a unique solution, restricting the use of the simple linear regression model. We may face this problem during the first few iterations of the adaptive greedy sampling algorithm, as we will have few error estimator values to build a reasonably accurate model.

The principal component regression (PCR) technique is a dimension reduction technique in which m explanatory variables are replaced by p linearly uncorrelated variables called principal components. The dimension reduction is achieved by considering only a few relevant principal components. The PCR approach helps to reduce the problem of estimating m coefficients to the more simpler problem of determining p coefficients. In the following, we describe the method to construct a surrogate model at the k th iteration in detail.

Before performing a principal component analysis (PCA), we center both the response vector $\hat{\varepsilon}$ and the data matrix $\hat{\mathcal{P}}_k$. The PCR starts by performing a PCA of the matrix $\hat{\mathcal{P}}_k$. For this, we compute an SVD

$$\hat{\mathcal{P}}_k = \hat{\Phi} \hat{\Sigma} \hat{\Psi},$$

and then the principal components are the columns of the matrix $\hat{\mathcal{P}}_k \hat{\Psi}$. For dimension reduction, we select only p columns of the matrix $\hat{\Psi}$ to construct a fairly accurate reduced model. In [45] it is suggested that the first three or four principal components are enough to analyze the yield curve changes. Let $Z = \hat{\mathcal{P}}_k \hat{\Psi}_p = [\hat{\mathcal{P}}_k \hat{\psi}_1, \dots, \hat{\mathcal{P}}_k \hat{\psi}_p]$ be the matrix containing the first p principal components. We regress $\hat{\varepsilon}$ on these principal components via

$$\hat{\varepsilon} = Z\Omega + \text{err},$$

where $\Omega = [\omega_1, \dots, \omega_p]$ is the vector containing the regression coefficients obtained using the principal components. The least square estimate for Ω is given as

$$\hat{\Omega} = \underset{\Omega}{\operatorname{argmin}} \|\hat{\varepsilon} - Z\Omega\|_2^2 = \underset{\omega}{\operatorname{argmin}} \left\| \hat{\varepsilon} - \sum_{i=1}^p z_i \omega_i \right\|_2^2.$$

We obtain the PCR estimate $\eta_{\text{PCR}} \in \mathbb{R}^m$ of the regression coefficients η as

$$\eta_{\text{PCR}} = \hat{\Psi}_p \hat{\Omega}.$$

Finally, the value of the surrogate model for any parameter vector $a_\ell = [a_{\ell,1}, \dots, a_{\ell,m}]$ is

$$\bar{\varepsilon}(\rho_\ell) = \eta_1 a_{\ell,1} + \dots + \eta_m a_{\ell,m}.$$

Algorithm 3 Surrogate error model using PCR

Input: Vector $\hat{\varepsilon} = [\varepsilon_1, \dots, \varepsilon_{c_k}]$, $\hat{\mathcal{P}}_k = [\rho_1, \dots, \rho_{c_k}] \in \mathbb{R}^{c_k \times \bar{m}}$, number of principal components p .

Output: Vector of regression coefficients η .

Standardize $\hat{\mathcal{P}}_k$ and $\hat{\varepsilon}$ with zero mean and variance one.

Compute the SVD $\hat{\mathcal{P}}_k = \hat{\Phi} \hat{\Sigma} \hat{\Psi}$.

Construct the matrix $Z = \hat{\mathcal{P}}_k \hat{\Psi}_p = [\hat{\mathcal{P}}_k \hat{\psi}_1, \dots, \hat{\mathcal{P}}_k \hat{\psi}_p]$ composed of principal components.

Compute the least square regression using the principal components as independent variables: $\hat{\Omega} = \operatorname{argmin}_{\Omega} \|\hat{\varepsilon} - Z\Omega\|_2^2$.

Compute the PCR estimate $\eta_{\text{PCR}} = \hat{\Psi}_p \hat{\Omega}$ of the regression coefficients η .

Algorithm 3 presents this surrogate error model, which we will use in the construction of an adaptive greedy sampling method.

4.3 Adaptive greedy sampling algorithm

The adaptive greedy sampling algorithm utilizes the designed surrogate model to locate optimal parameter groups adaptively at each greedy iteration $i = 1, \dots, I_{\max}$. The first few steps of the algorithm resemble the classical greedy sampling approach. First the parameter group ρ_1 is selected from the parameter space \mathcal{P} and the reduced basis Q_1 is constructed. Furthermore, the algorithm randomly selects c_0 parameter groups and constructs a temporary parameter set $\hat{\mathcal{P}}_0 = \{\rho_1, \dots, \rho_{c_0}\}$. For each parameter group in the parameter set $\hat{\mathcal{P}}_0$, the algorithm determines a reduced order model and computes an array of corresponding residual errors $\varepsilon^0 = \{\varepsilon(\rho_1), \dots, \varepsilon(\rho_{c_0})\}$. Then a surrogate model for the error estimator $\bar{\varepsilon}$ is constructed based on the estimator values $\{\varepsilon(\rho_j)\}_{j=1}^{c_0}$, as discussed in Sect. 4.2. The obtained surrogate model is then simulated for the entire parameter space \mathcal{P} . Furthermore, we locate c_k parameter groups corresponding to the first c_k maximal values of the surrogate model. We then construct a new parameter set $\hat{\mathcal{P}}_k = \{\rho_1, \dots, \rho_k\}$ composed of these c_k parameter groups.

The algorithm determines a reduced model for each parameter group within the parameter set $\hat{\mathcal{P}}_k$ and obtains an array of error estimator values $\varepsilon^k = \{\varepsilon(\rho_1), \dots, \varepsilon(\rho_{c_k})\}$ for the parameter set $\hat{\mathcal{P}}_k$. Furthermore, we concatenate the set $\hat{\mathcal{P}}_k$ and the set $\hat{\mathcal{P}}_0$ to form a new parameter set $\hat{\mathcal{P}} = \hat{\mathcal{P}}_k \cup \hat{\mathcal{P}}_0$. Let $e_{\text{sg}} = \varepsilon^0 \cup \dots \cup \varepsilon^k$ be the set composed of all the error estimator values available at the k th iteration. The algorithm then uses this error estimator set e_{sg} to build a new surrogate model. The quality of the surrogate model increases with each iteration as we get more error estimator values. This process is repeated until the cardinality of the set $\hat{\mathcal{P}}$ reaches c , giving

$$\hat{\mathcal{P}} = \hat{\mathcal{P}}_0 \cup \hat{\mathcal{P}}_1 \cup \hat{\mathcal{P}}_2 \cup \dots \cup \hat{\mathcal{P}}_K.$$

Finally, the optimal parameter group ρ_l which maximizes the error estimator (7) is extracted from the parameter set $\hat{\mathcal{P}}$. Note that typically it is not necessary to obtain a very accurate sampling using the designed surrogate model. Sampling the full model in the neighborhood of the parameter group with maximal error is sufficient to obtain good re-

Table 2 List of symbols used in Algorithm 4

ρ	group of model parameters $\{a(t), b, \sigma\}$;
c_k	number of parameters selected adaptively based on surrogate modeling;
V	solution obtained by solving a full order model;
c_0	Number of randomly selected parameter groups to initiate the algorithm;
ε	error estimator;
ε^k	set comprised of error estimator values at k th iteration;
e_{sg}	set composed of all error estimator values at the k th iteration;
\bar{e}	surrogate error model;
$\hat{\rho}$	parameter set used to obtain the optimal parameter group;
ρ_l	optimal parameter group which maximizes the error estimator;
e	relative error between a reduced and full model;
\bar{V}	solution obtained using a reduced order model;
\hat{V}	snapshot matrix;
E_p	error set;
\bar{e}	error model as an approximate error for the exact error e ;
e_{tol}^{max}	tolerance for the relative error, greedy iterations terminates if $\bar{e} < e_{tol}^{max}$;
γ_e	slope of the error model;
$\hat{\tau}$	intersection with the logarithmic axis $\log(y)$.

sults. To monitor the convergence of the adaptive greedy algorithm, we have designed an error model \bar{e} for the relative error e as a function of the residual error ε .

To build an approximate error model, we simulate one full model at each greedy iteration for the optimal parameter group ρ_l , and update the snapshot matrix \hat{V} . A new reduced basis Q is then obtained by computing the truncated singular value decomposition of the updated snapshot matrix as explained in Sect. 4. Furthermore, we solve the reduced model for the optimal parameter group before and, after updating the reduced basis, obtain the respective error estimator values $\varepsilon^{bef}(\rho_l)$, and $\varepsilon^{aft}(\rho_l)$. Then we compute the relative errors e^{bef}, e^{aft} between the full and reduced model constructed before and after updating the reduced basis.

In this way, at each greedy iteration, we get a set of error values E_p that we use to construct a linear approximate error model for the exact error e based on the error estimator ε .

$$E_p = \{(e_1^{bef}, \varepsilon_1^{bef}) \cup (e_1^{aft}, \varepsilon_1^{aft}), \dots, (e_i^{bef}, \varepsilon_i^{bef}) \cup (e_i^{aft}, \varepsilon_i^{aft})\}, \quad (9)$$

as

$$\log(\bar{e}_i) = \gamma_i \log(\varepsilon) + \log \tau. \quad (10)$$

Setting $\mathcal{Y} = \log(\bar{e})$, $\mathcal{X} = \log(\varepsilon)$ and $\hat{\tau} = \log(\tau)$ we get

$$\mathcal{Y} = \gamma_e \mathcal{X} + \hat{\tau},$$

where γ_e is the slope of the linear model and $\hat{\tau}$ is the intersection with the logarithmic axis $\log(y)$.

After each greedy iteration, we get more data points in the error set E_p , which increases the accuracy of the error model. In Sect. 5, we illustrate that this linear model is sufficient to achieve an accurate error model. The adaptive greedy sampling approach is summarized in Algorithm 4. The list of symbols used in Algorithm 4 is presented in Table 2.

Algorithm 4 Adaptive greedy sampling algorithm

Input: Maximal number of iterations I_{\max} , maximal number of parameter groups c , number of adaptive candidates c_k , parameter space \mathcal{P} , tolerance e_{tol}^{\max} .

Output: Q .

Choose first parameter group $\rho_1 = [[a_{11}, \dots, a_{1m}], b, \sigma]$ from \mathcal{P} .

Simulate the full model for the parameter group ρ_1 and store the results in V_1 .

Compute a truncated SVD of the matrix V_1 and construct Q_1 .

FOR $i = 2, \dots, I_{\max}$

Randomly select a set of parameter groups $\hat{\mathcal{P}}_0 = \{\rho_1, \rho_2, \dots, \rho_{c_0}\} \subset \mathcal{P}$.

FOR $j = 1, \dots, c_0$

determine a reduced model for parameter group ρ_j with reduced basis Q_{i-1} .

Compute the error estimator $\varepsilon(\rho_j)$.

END

Let $\varepsilon^0 = \{\varepsilon(\rho_1), \dots, \varepsilon(\rho_{c_0})\}$ be error estimators for $\hat{\mathcal{P}}_0$.

Let $k = 1$ and $e_{\text{sg}} = \varepsilon^0$.

WHILE $n(\hat{\mathcal{P}}) < c$

Construct a surrogate model $\bar{\varepsilon}(\rho)$ using the values e_{sg} .

Compute the values $\bar{\varepsilon}(\rho)$ of the surrogate model over \mathcal{P} .

Determine first c_k maximal values of $\bar{\varepsilon}(\rho)$ and parameter set $\hat{\mathcal{P}}_k =$

$\{\rho_1, \dots, \rho_{c_k}\}$.

For $x = 1, \dots, n(\hat{\mathcal{P}}_k)$

determine a reduced model for parameter group ρ_x .

Compute the error estimator $\varepsilon(\rho_x)$.

END

Let $\varepsilon^k = \{\varepsilon(\rho_1), \dots, \varepsilon(\rho_{c_k})\}$ be the error estimators for $\hat{\mathcal{P}}_k$.

Update $e_{\text{sg}} = \{\varepsilon^0 \cup \dots \cup \varepsilon^k\}$.

Construct new parameter set $\hat{\mathcal{P}} = \hat{\mathcal{P}}_0 \cup \hat{\mathcal{P}}_k$.

$k = k + 1$.

END

Find $\rho_I = \operatorname{argmax}_{\rho \in \hat{\mathcal{P}}} \varepsilon(\rho)$.

IF $i > 2$ and $\bar{e}_i \leq e_{\text{tol}}^{\max}$, set $Q = Q_{i-1}$, STOP.

Solve the full model for the parameter group ρ_I and store result in V_i .

Solve reduced model for the parameter group ρ_I using Q_{i-1} and store result in \bar{V}_i .

Compute relative error e_i^{bef} and error estimator $\varepsilon_i^{\text{bef}}$ using the reduced model

obtained with Q_{i-1} , before updating reduced basis. Set

$e_i^{\text{bef}} = \|V_i(\rho_I) - \bar{V}_i(\rho_I)\| / \|V_i(\rho_I)\|$.

Construct a snapshot matrix \hat{V} by concatenating the solutions V_ℓ for $\ell = 1, \dots, i$.

Compute an SVD of the matrix \hat{V} and construct Q_i .

Simulate reduced model for parameter group ρ_I using Q_i and store result in \bar{V}_{i+1} .

Compute relative error e_i^{aft} and error estimator $\varepsilon_i^{\text{aft}}$ using the reduced model

obtained with Q_i , after updating the reduced basis. Set

$e_i^{\text{aft}} = \|V_i(\rho_I) - \bar{V}_{i+1}(\rho_I)\| / \|V_i(\rho_I)\|$.

Construct error set $E_p = \{(e_1^{\text{bef}}, \varepsilon_1^{\text{bef}}) \cup (e_1^{\text{aft}}, \varepsilon_1^{\text{aft}}), \dots, (e_i^{\text{bef}}, \varepsilon_i^{\text{bef}}) \cup (e_i^{\text{aft}}, \varepsilon_i^{\text{aft}})\}$.

Construct model for error \bar{e} using error set E_p : $\log(\bar{e}_i) = \gamma_i \log(\varepsilon) + \log \tau$.

END

Table 3 Numerical Example of a floater with cap and floor

Coupon frequency	quarterly
Cap rate, C_R	2.25 % p.a.
Floor rate, F_R	0.5 % p.a.
Currency	EURO
Maturity	10 years
Nominal amount	1.0
b	0.015
σ	0.006

5 Numerical example

A numerical example of a floater with cap and floor [33] is used to test the developed algorithms and methods. We model the floater instrument using the Hull–White model and compare the results of the PDE model by discretizing as in Sect. D with the results of the reduced model. The reduced model is generated by implementing the POD method along with the classical and the adaptive greedy sampling techniques. The characteristics of the floater instrument are as shown in Table 3. The interest rates are capped at $C_R = 2.25\%$ p.a. and floored at $C_F = 0.5\%$ p.a. with the reference rate as Euribor3M. The coupon rates can be written as

$$\text{coup} = \min(2.25\%, \max(0.5\%, \text{Euribor3M})). \quad (11)$$

Note that the coupon rate $\text{coup}^{(n)}$ at time t_n is set in advance as the coupon rate at t_{n-1} . All computations are carried out on a PC with 4 cores and 8 logical processors at 2.90 GHz (Intel i7 7th generation). We used MATLAB R2018a for the yield curve simulations and the model reduction. The numerical method for the yield curve simulations is tested with real market based historical data. Market data are available from market data providers like Thomson Reuters, Bloomberg, and several others. We obtained this data from MathConsult within their UnRisk Omega datasets [46]. The daily interest rate data are collected at 21 tenor points in time over the past 5 years, where each year has 260 working days, so there are 1300 observation periods. We have used the inbuilt UnRisk tool for the parameter calibration, which is well integrated with Mathematica (version used: Mathematica 11.3). Further, we used calibrated parameters for the construction of a Hull–White model.

We have computed the model parameters as explained in Appendix C. The yield curve simulation is the first step to compute the model parameters. Based on the procedure described in Appendix B, we have performed the bootstrapping process for the recommended holding period of 10 years, i.e., for the maturity of the floater. The collected historical data has 21 tenor points and 1306 observation periods as follows (D: Day, M: Month, Y: Year):

$$m =: \{1D, 3M, 6M, 1Y, 2Y, 3Y, \dots, 10Y, 12Y, 15Y, 20Y, 25Y, 30Y, 40Y, 50Y\},$$

$$n =: \{1306 \text{ daily interest rates at each tenor point}\}.$$

The 10,000 simulated yield curves in 10 years in the future are presented in Fig. 2. For the floater example, we need parameter values only until the 10Y tenor point (maturity of the floater). Henceforth, we consider the simulated yield curves with only the first 13 tenor points. The calibration generates the real parameter space of dimension $\mathbb{R}^{10,000 \times 13}$ for the parameter $a(t)$. We considered the constant volatility $\sigma = 0.006$ and the constant

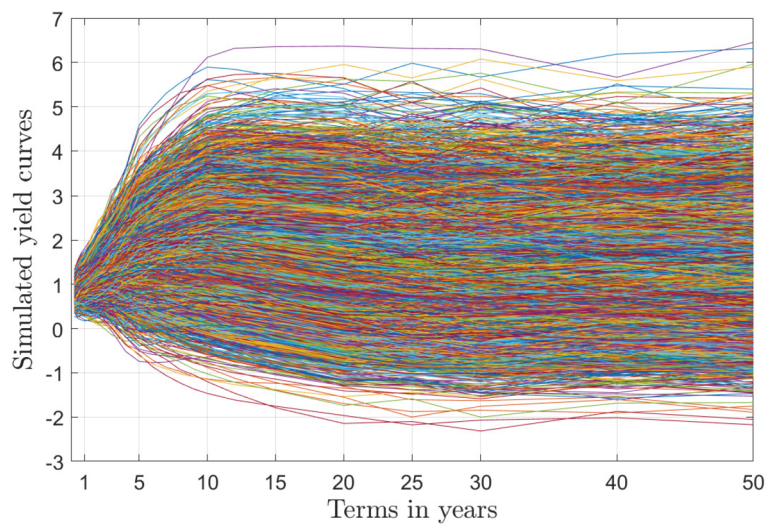


Figure 2 10,000 simulated yield curves obtained by bootstrapping for next 10 years

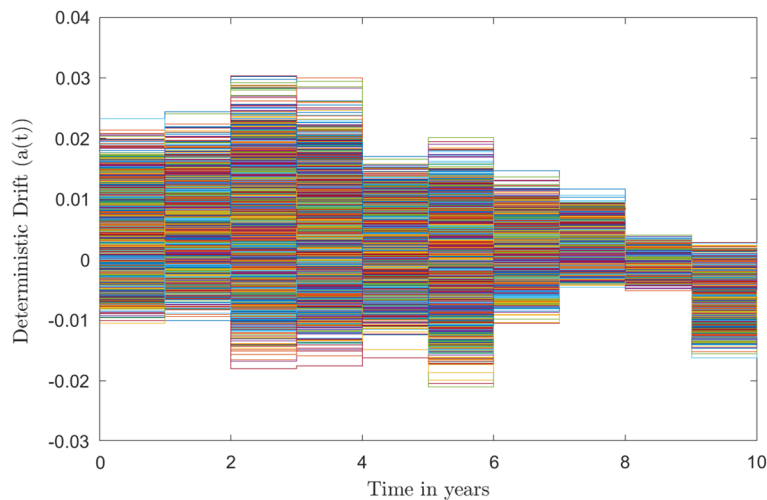
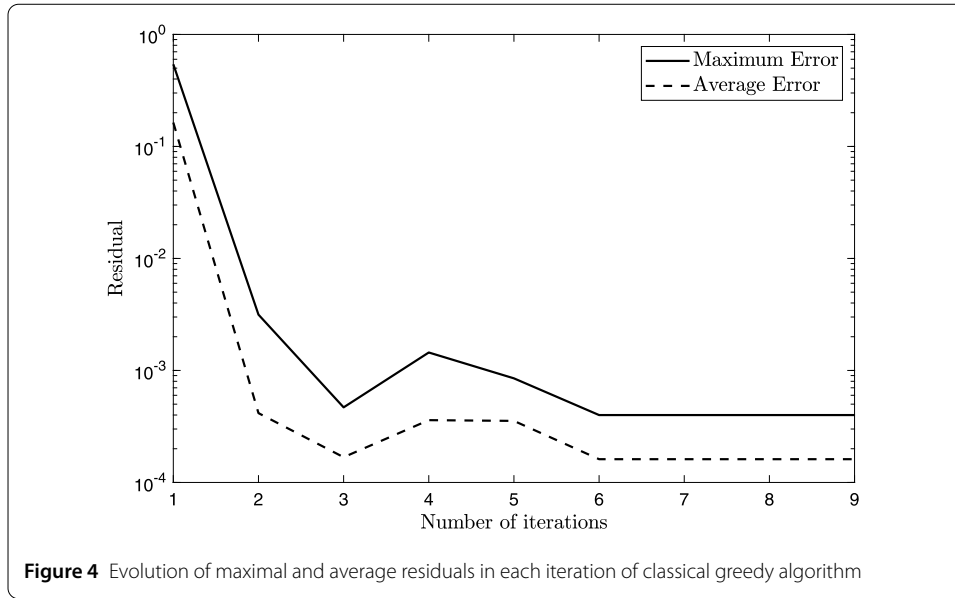


Figure 3 10,000 parameter vectors $a(t)$ as a piecewise function of time

mean reversion $b = 0.015$. All variable parameters are assumed to be piecewise constants between the tenor points $(0 - 3M, 3M - 6M, 6M - 1Y, 1Y - 2Y, 2Y - 3Y, \dots, 9Y - 10Y)$. Figure 3 shows 10,000 different piecewise constant parameters $a(t)$. The computational domain for a spatial dimension r is restricted as described in Sect. D. Here, $r_{\text{low}} = -0.1$ and $r_{\text{up}} = 0.1$. We applied homogeneous Neumann boundary conditions of the form

$$\left. \frac{\partial V}{\partial r} \right|_{r_{\text{low}}} = 0, \quad \left. \frac{\partial V}{\partial r} \right|_{r_{\text{up}}} = 0. \quad (12)$$

We divided the spatial domain into $M = 600$ equidistant grid points $\{r(1), r(2), \dots, r(M)\}$ and used the N time points (measured in days) starting from $t = 0$ until maturity T , i.e., in our case, the number of days until maturity are assumed to be $3600 \approx 10Y$ with an interval



$\tau = 1$. Rewriting (2), for given final value of V at the last tenor point, we obtain

$$A(\rho_\ell(t))V^{n-1} = B(\rho_\ell(t))V^n.$$

We can apply the first boundary condition in (12) by updating the first and the last rows (A_1 and A_M) of the matrix $A(\rho_\ell)$ which yields

$$A_1 = (-1, 1, 0, \dots, 0) \quad \text{and} \quad A_M = (0, \dots, 0, 1, -1).$$

The second Neumann boundary condition can be applied by changing the last entry of the vector BV^n to zero. Starting at $t = T$ with the known terminal condition $V(T)$ as the principal amount, at each time step, we solve the system of linear equations (2).

Note that we need to update the value of the grid point $r(i)$ every three months as the coupon frequency is quarterly by adding coupon f^n based on the coupon rate given by (11). To calculate the value at the intermediate holding period, we calculate the fair value of the floater and add it to the accreted coupons.

We have implemented the parametric model reduction approach for the floater example, as discussed in Sect. 4 using both the classical and the adaptive greedy sampling algorithms.

At each iteration of the classical greedy sampling approach, the algorithm constructs a reduced basis via Algorithm 1. We have specified a maximum number of 40 pre-defined candidates to construct a set $\hat{\mathcal{P}}$ and a maximum number of iterations $I_{\max} = 10$. The progression of the maximal and average residuals with each iteration of the greedy algorithm is presented in Fig. 4. It is observed that the maximal residual error typically decreases in the process and the proposed greedy algorithm efficiently locates the optimal parameter groups and constructs the desired reduced basis Q . Furthermore, we tested the effect of change in the cardinalities of the set $\hat{\mathcal{P}}$. The proposed algorithm is applied with three different cardinalities of $\hat{\mathcal{P}}$: $|\hat{\mathcal{P}}_1| = 20$, $|\hat{\mathcal{P}}_2| = 30$, $|\hat{\mathcal{P}}_3| = 40$. Note that we have constructed $\hat{\mathcal{P}}$ by randomly selecting the parameter groups from the parameter space \mathcal{P} . Figure 5 shows

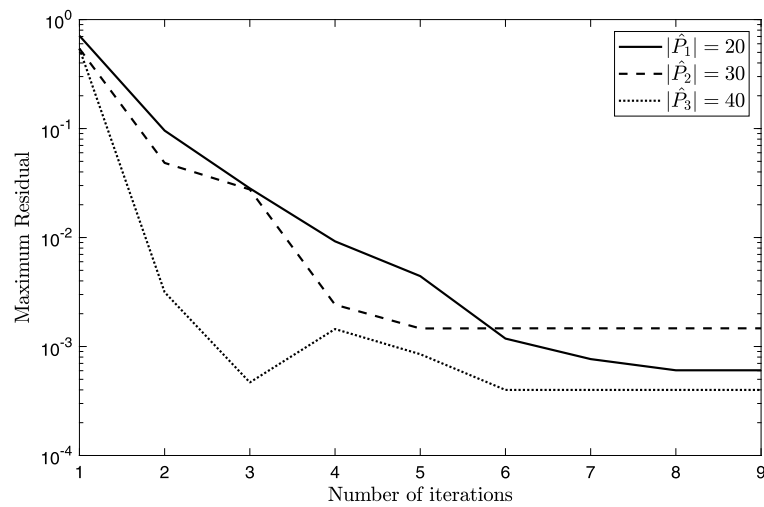


Figure 5 Evolution of the maximum residual error for three different cardinalities of set $\hat{\mathcal{P}}$

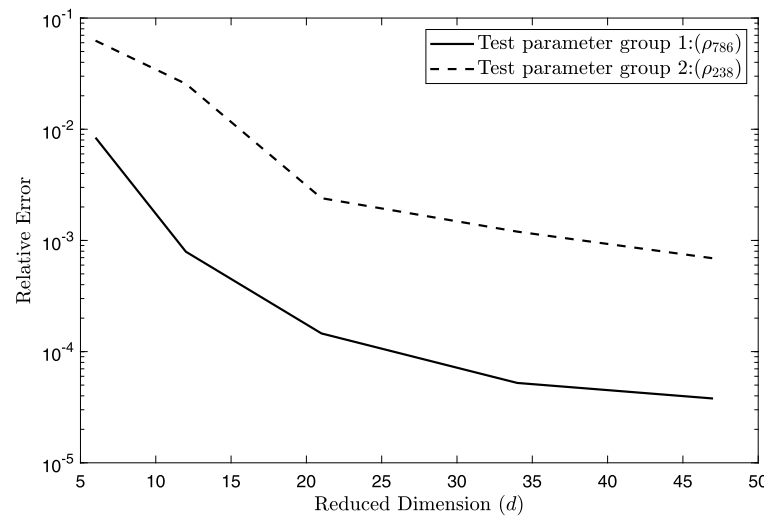
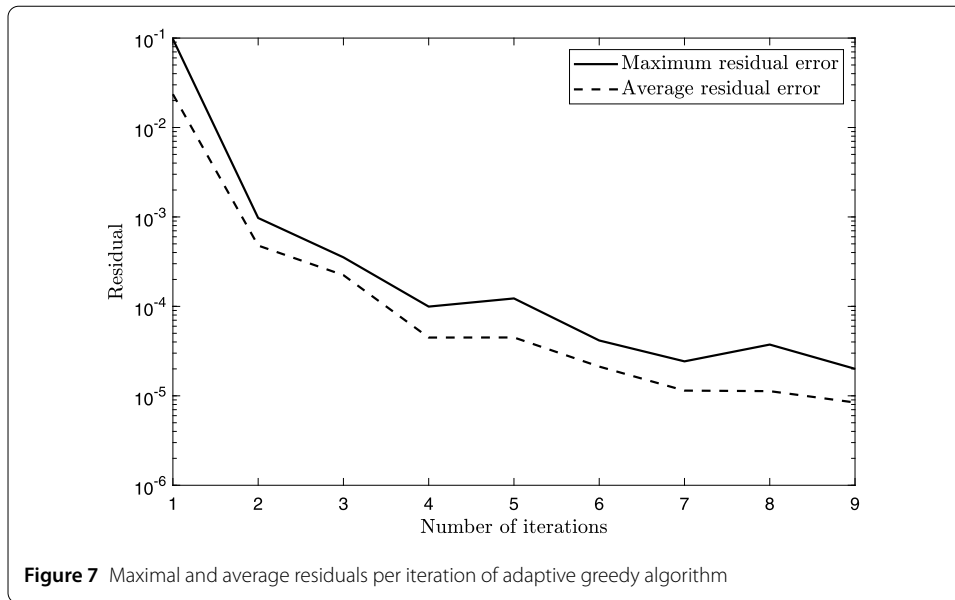


Figure 6 Relative error between full and reduced model for two different parameter groups

the plot of the maximal residual against the number of iterations for three different cardinalities.

It is evident that with an increasing number of candidates, the maximal residual error decreases and the decrease is sufficient for a cardinality 20 of $\hat{\mathcal{P}}$. Figure 6 illustrates the relative error between the full model and the reduced-order model for two different parameter groups. During our tests we noticed that there are some parameter groups (e.g., ρ_{238}) for which the reduced model gives unsatisfactory results.

One can observe that the reduced model for the parameter group ρ_{238} (dashed line) shows inferior results as compared to the reduced model for the parameter group ρ_{786} (solid line). Even an increase in the reduced dimension d does not improve the quality of the result substantially.



This reveals that the selection of trial candidates by random sampling may neglect parameter groups corresponding to the significant error.

To overcome this drawback, we have also implemented the adaptive greedy sampling approach for the floater example. At each greedy iteration, the algorithm locates $c_k = 10$ parameter groups adaptively using the surrogate modeling technique, as described in Sect. 4.2. We have fixed the maximum number of elements within the parameter set $\hat{\mathcal{P}}$ to 40. Furthermore, the adaptively obtained parameter set $\hat{\mathcal{P}}$ has been used to locate the optimal parameter group ρ_I . These steps are repeated for a maximum of $I_{\max} = 10$ iterations or until convergence. The algorithm has been initiated by selecting $c_0 = 20$ random parameter groups.

The optimal parameter group updates the snapshot matrix, and consecutively the algorithm generates a new reduced basis at each greedy iteration. Figure 7 shows the evolution of maximum and average residual errors with each iteration of the adaptive greedy algorithm.

The residual error decreases with each incrementing iteration and hence the algorithm succeeded in locating the optimal parameter group efficiently. To monitor the convergence of the adaptive greedy algorithm, we have designed an error model \bar{e} for the relative error e as a function of the residual error ε . Figure 8 shows the designed error model compared to the available error set E_p for four different greedy iterations. The error plot exhibits a strong correlation between the relative error and the residual error. The results indicate that a linear error model is satisfactory to capture the overall behavior of the exact error as a function of the residual error. We have used the reduced basis obtained from the adaptive greedy sampling procedure to design the reduced model. Figure 9 presents the relative error plot for the parameter groups ρ_{238} , and ρ_{786} . We see that the adaptive greedy approach gives better results than the classical greedy method. With a reduced dimension of $d = 6$, we obtained an excellent result as the relative error is less than 10^{-3} .

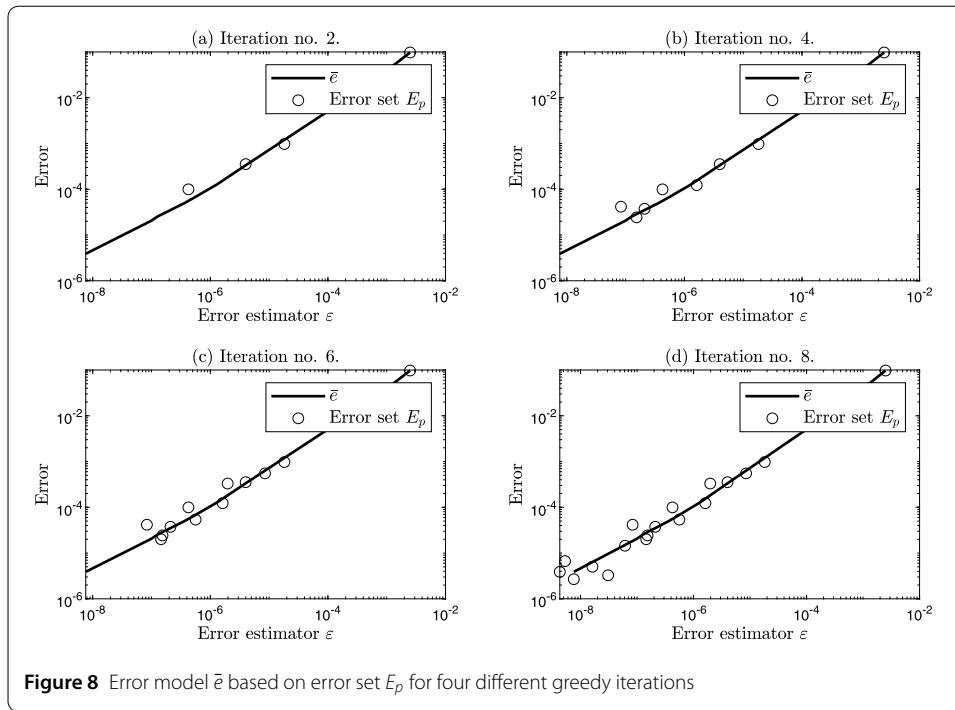


Figure 8 Error model $\bar{\epsilon}$ based on error set E_p for four different greedy iterations

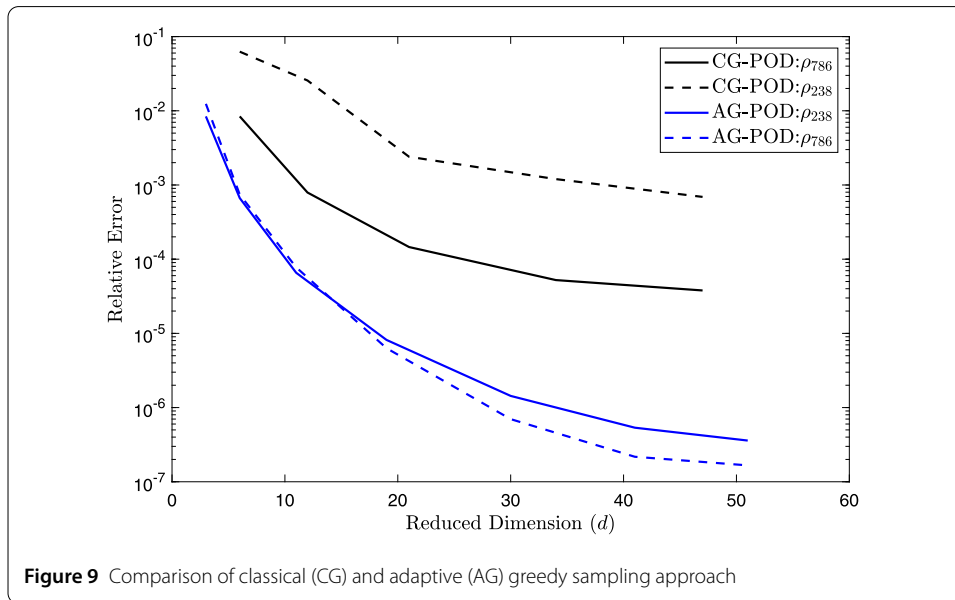


Figure 9 Comparison of classical (CG) and adaptive (AG) greedy sampling approach

5.1 Computational cost

In the case of the classical greedy sampling approach, the algorithm solves c reduced models and one full model at each greedy iteration. It also computes a truncated SVD of the updated snapshot matrix with each proceeding iteration. Let t_{RM} be the time required to solve one reduced model, t_{FM} the computational time required for one full model, and t_{SVD} the time required to obtain a truncated SVD of the snapshot matrix. The total computational time T_Q^{CG} required to obtain the reduced basis after i steps of the greedy algorithm

Table 4 Computing time/reduction time (T_Q) to generate projection subspace

Algorithm	Cardinality $ \hat{\mathcal{P}} $	Max. no. iterations I_{\max}	Computing time
Classical greedy sampling	20	10	56.82 s
Classical greedy sampling	30	10	82.54 s
Classical greedy sampling	40	10	95.04 s
Adaptive greedy sampling	40	10	183.21 s

Table 5 Evaluation time

Algorithm	Model	Eva. time single ρ_s	Total Eva. time (T_{eva})	Total time $T_Q + T_{\text{eva}}$
	FM, $M = 600$	0.2193 s	2193.72 s	2193.72 s
Classical greedy sampling	RM, $d = 5$	0.0102 s	102.56 s	197.60 s
Classical greedy sampling	RM, $d = 10$	0.0125 s	125.48 s	220.52 s
Adaptive greedy sampling	RM, $d = 6$	0.0104 s	104.38 s	287.59 s
Adaptive greedy sampling	RM, $d = 10$	0.0124 s	124.32 s	307.53 s

in the case of the classical greedy sampling approach can be given as

$$T_Q^{CG} \approx [c \times t_{\text{RM}} + (t_{\text{FM}} + t_{\text{SVD}})] \times i.$$

Similarly, in the case of the adaptive greedy approach, the total computational time T_Q^{AG} can be given as

$$T_Q^{AG} \approx [c_0 \times t_{\text{RM}} + \underbrace{k(C_k \times t_{\text{RM}} + t_{\text{SM}} + t_{\text{SM}}^{\text{ev}})}_{t_{\rho_I}} + t_{\text{FM}} + t_{\text{SVD}} + 2t_{\text{RM}}^{\text{af,bf}} + t_{\text{EM}}] \times i,$$

where t_{SM} and $t_{\text{SM}}^{\text{ev}}$ denote the computational times required to build and evaluate a surrogate model for the entire parameter space respectively. t_{EM} is the time required to build an error model. The term $2t_{\text{RM}}^{\text{af,bf}}$ shows the computational time needed to solve the reduced model after and before updating the reduced basis.

Table 4 compares the computing times required to generate the reduced basis Q for different sets of $\hat{\mathcal{P}}$ in case of the classical greedy sampling approach with that needed to obtain the reduced basis in case of the adaptive greedy sampling approach. The computing time t_{ρ_I} required to locate the optimal parameter group by constructing a surrogate model for one greedy iteration is approximately 8 seconds. t_{ρ_I} is nothing but the time required to complete a while loop outlined in Algorithm 4 for a single greedy iteration. Thus, the total time contributed to generate the reduced basis via surrogate modeling is $I_{\max} \times t_{\rho_I} = 78.56\text{s}$, considering the adaptive greedy algorithm runs for I_{\max} iterations. Figure 8 shows that we can truncate the algorithm after 4 or 5 iterations as the residual error falls below 10^{-4} . The computing times required to simulate reduced models and full models are presented in Table 5. The time required to solve the complete system with a parameter space of $10,000 \times m$ for both a full and reduced model is given in the total time column.

We can see that the evaluation time required for the reduced model is at least 18–20 times less than that of the full model. However, there is a slight increase in total time due to the addition of the reduction time T_Q . One can also observe that with an increase in the dimension d of the reduced model, the evaluation time increases significantly. The reduced model with the classical greedy sampling approach is at least 10–11 times faster

than that for the full model. The time required to simulate the reduced model with the adaptive greedy sampling approach is a bit higher due to the time invested in building surrogate and error models. Despite of that, the reduced model is at least 8–9 times faster than the full model. The computational time presented in both tables considers that the greedy algorithms run for the maximal number of iterations I_{\max} . However, we can truncate the algorithms after 4 or 5 iterations, i.e., we can practically achieve even more speedup than discussed here.

5.2 Floater performance scenario values

To design a key information document, we need the values of the floater at different spot rates. The spot rate r_{sp} is the yield rate at the first tenor point from the simulated yield curve. The value of a floater at the spot rate r_{sp} is nothing but the value at that short rate $r = r_{sp}$. For 10,000 simulated yield curves, we get 10,000 different spot rates and the corresponding values for the floater. At the recommended holding period, we determine a VaR equivalent volatility to calculate a market risk indicator. For the floater example, the calculated VEV = 0.3989%, which gives the market risk indicator as 1.

The 10,000 values are further used to calculate three different scenarios: (i) favorable scenario, (ii) moderate scenario, (iii) unfavorable scenario which are the values at 90th percentile, 50th percentile and 10th percentile of 10,000 values, respectively. Table 6 shows the floater performance scenario results obtained using the reduced-order model and the commercially available UnRisk software for the comparison. Figure 10 shows the distribution of fair values of the floater, plus the coupons in the respective path obtained so far, after five years and ten years. At five years, the main contribution to price variability arises from discounting the future cash flows (between five and ten years) with a discount factor arising from log-normally distributed rates. Therefore we have (in our example) a left-skewed distribution at five years, and a right-skewed distribution at ten years.

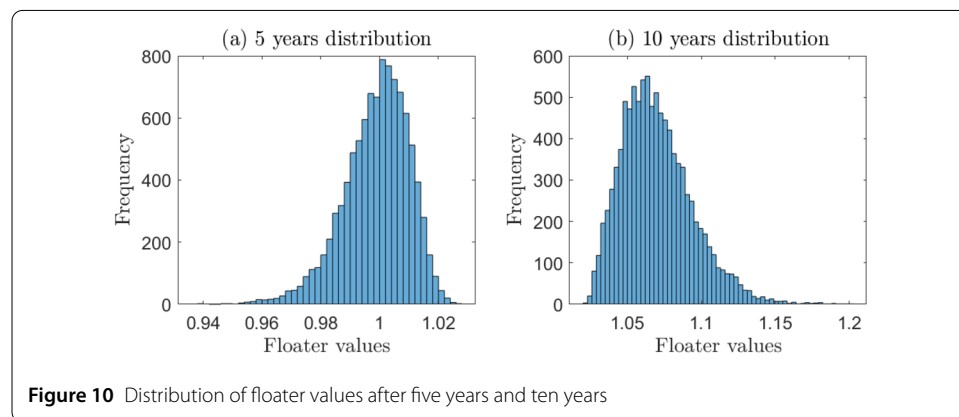


Table 6 Results for a floater with cap and floor

Performance Scenario	5 years		10 years	
	ROM	UnRisk	ROM	UnRisk
Favorable (90th percentile)	1.013	1.014	1.105	1.101
Moderate (50th percentile)	1.009	1.002	1.067	1.066
Unfavorable (10th percentile)	0.988	0.984	1.038	1.041

6 Conclusion

This paper presents a parametric model reduction approach for a discretized convection-diffusion-reaction PDE that arises in the analysis of financial risk. The parameter space with time-dependent parameters is generated via the calibration of financial models based on market structures. A finite difference method has been implemented to solve this PDE. The selection of parameters to obtain the reduced basis is of utmost importance. We have established a greedy approach for parameter sampling, and we noticed that there are some parameter groups for which the classical greedy sampling approach gave unsatisfactory results. To overcome this drawback, we have applied an adaptive greedy sampling method using a surrogate model for the error estimator that is constructed for the entire parameter space and further used to locate the parameters which most likely maximize the error estimator. The surrogate model is built using the principal component regression technique. We tested the designed algorithms for a numerical example of a floater with cap and floor solved using the Hull–White model. The results indicate the computational advantage of the parametric model reduction technique for the short-rate models. A reduced model of dimension $d = 6$ was enough to reach an accuracy of 0.01%. The reduced model was at least 10–12 times faster than the full model. The developed model order reduction approach shows potential applications in the historical or Monte Carlo value at risk calculations as well, where a large number of simulations need to be performed for the underlying instrument.

7 Outlook

In this paper, we solved the numerical example of the floater with cap and floor based on the Hull–White model, and the obtained results show that the model order reduction approach works well. However, there are more complicated financial instruments under the Category 3 PRIIPs, e.g., instruments with Bermudan or American callabilities, callable/puttable steepener, target redemption notes, snowball. The coupons for the steepener instrument depend on the difference between two rates, while in the case of snowball, a new coupon depends on the old coupon. Thus, to solve such complicated instruments, we need to consider multi-factor models like the two-factor Hull–White model. These challenges demand to explore the developed model order reduction algorithms for more complicated financial models. In the future, we will implement the developed approaches for financial instruments like steepener, snowball.

Appendix A: Derivation of the model

In this appendix, we derive the model for a risk-less investment with a continuous profit at a low risk-free rate. We follow the standard derivation in [1]. It is based on the same principle of constructing a risk-free portfolio as the seminal theory of Black–Scholes.

Let $B(t)$ be the value of a bank account at time $t \geq 0$. We assume that the bank account evolves according to the differential equation

$$dB(t) = B(t)r(t) dt,$$

with initial value $B(0) = 1$, where $r(t)$ is the *short-rate*, i.e. the growth rate of the bank account B within a small time interval $(t, t + dt)$. This leads to the formula

$$B(t) = \exp\left(\int_0^t r(\tau) d\tau\right).$$

When working with interest-rate products, the study of the variability of interest rates is essential. Therefore, the short-rate is modelled as a stochastic process. Let S be the price of stock at the end of the n th trading day. The daily return from days n to $n + 1$ is given by $(S_{n+1} - S_n)/S_n$. In general, it is common to work with log returns, since the log return of k days can be easily computed by adding up the daily log returns,

$$\log(S_k/S_0) = \log(S_1/S_0) + \cdots + \log(S_k/S_{k-1}).$$

Based on the assumption that the log returns over disjoint time intervals are stochastically independent and equally distributed, the central limit theorem [47] implies that the log returns are normally distributed [48]. These properties [49] can be realized by a standard *Brownian motion* $W(t)$, i.e., a family of random variables $W(t)$, indexed by $t \geq 0$, with the properties

- $W(0) = 0$.
- With probability 1, the function $W(t)$ is continuous in t .
- For $t \geq 0$, the increment $W(t + \tau) - W(\tau)$ is normally distributed with mean 0 and variance t , i.e.,

$$W(t + \tau) - W(\tau) \sim N(0, t).$$

- For all N and times $t_0 < t_1 < \cdots < t_{N-1} < t_N$, the increments $W(t_j) - W(t_{j-1})$ are stochastically independent.

These properties lead to the *Langevin equation*, see [50], which is a stochastic differential equation

$$\frac{dX(t)}{dt} = q(t)X(t),$$

with initial condition $X(0) = X_0$ with probability 1, where the stochastic parameter $q(t)$ is given, see [51], by

$$q(t) = f(r(t), t) + h(r(t), t)w(t).$$

Here $w(t)$ is a white noise process, and f, h are given functions of the interest rate $r(t)$. This leads to

$$\frac{dX(t)}{dt} = f(r(t), t)X(t) + g(r(t), t)X(t)w(t). \quad (13)$$

The force $w(t) = dW(t)/dt$ is a fluctuating quantity with Gaussian distribution. Substituting $dW(t) = w(t)dt$ in (13), we get

$$dX(t) = f(r(t), t)X(t)dt + g(r(t), t)X(t)dW(t)$$

and we obtain an SDE for the short-rate $r(t)$ via

$$dr(t) = f(t, r(t)) dt + g(t, r(t)) dW(t).$$

Based on the Ito lemma [3], we can derive a general PDE for any underlying instrument depending on the short-rate. Consider a *risk neutral portfolio* $\Pi(t)$ that depends on the short-rate $r(t)$ and consists of two *interest rate instruments* V_1 and V_2 with different *maturities* T_1 and T_2 , respectively. Suppose that there are $\Delta = (\frac{\partial V_1}{\partial r(t)} / \frac{\partial V_2}{\partial r(t)})$ units of the instrument V_2 . For an infinitesimal time interval, the value change of the portfolio is $d\Pi(t) = \Delta dV_2 - dV_1$. To avoid the arbitrage, we consider a risk-free rate [1], which gives

$$d\Pi(t) = \Delta dV_2 + (V_1 - \Delta V_2)r(t) dt - dV_1,$$

and obtain the PDE

$$\begin{aligned} d\Pi(t) = & (V_1 - \Delta V_2)r(t) dt \\ & - \left[\left(\frac{\partial V_1}{\partial r(t)} f(r(t), t) + \frac{\partial V_1}{\partial t} + \frac{1}{2} \frac{\partial^2 V_1}{\partial r_t^2} g^2(r(t), t) \right) dt + \frac{\partial V_1}{\partial r(t)} g(r(t), t) dW(t) \right] \\ & + \Delta \left[\left(\frac{\partial V_2}{\partial r(t)} f(r(t), t) + \frac{\partial V_2}{\partial t} + \frac{1}{2} \frac{\partial^2 V_2}{\partial r_t^2} g^2(r(t), t) \right) dt + \frac{\partial V_2}{\partial r(t)} g(r(t), t) dW(t) \right]. \end{aligned}$$

Assuming a *zero net investment requirement*, i.e., $d\Pi(t) = 0$, we obtain

$$\begin{aligned} 0 = & \left[V_1 - \left(\frac{\partial V_1}{\partial r(t)} / \frac{\partial V_2}{\partial r(t)} \right) V_2 \right] r(t) dt \\ & - \left[\left(\frac{\partial V_1}{\partial r(t)} f(r(t), t) + \frac{\partial V_1}{\partial t} + \frac{1}{2} \frac{\partial^2 V_1}{\partial r_t^2} g^2(r(t), t) \right) dt + \frac{\partial V_1}{\partial r(t)} g(r(t), t) dW(t) \right] \\ & + \left(\frac{\partial V_1}{\partial r(t)} / \frac{\partial V_2}{\partial r(t)} \right) \left[\left(\frac{\partial V_2}{\partial r(t)} f(r(t), t) + \frac{\partial V_2}{\partial t} + \frac{1}{2} \frac{\partial^2 V_2}{\partial r_t^2} g^2(r(t), t) \right) dt \right. \\ & \left. + \frac{\partial V_2}{\partial r(t)} g(r(t), t) dW(t) \right]. \end{aligned}$$

Eliminating the stochastic term, we obtain

$$\begin{aligned} & \left[V_1 - \left(\frac{\partial V_1}{\partial r(t)} / \frac{\partial V_2}{\partial r(t)} \right) V_2 \right] r(t) dt \\ & = \left[\frac{\partial V_1}{\partial t} + \frac{1}{2} \frac{\partial^2 V_1}{\partial r_t^2} g^2(r(t), t) - \left(\frac{\partial V_1}{\partial r(t)} / \frac{\partial V_2}{\partial r(t)} \right) \left(\frac{\partial V_2}{\partial t} + \frac{1}{2} \frac{\partial^2 V_2}{\partial r_t^2} g^2(r(t), t) \right) \right] dt. \end{aligned}$$

Rearranging the terms, we get

$$\frac{\frac{\partial V_1}{\partial t} + \frac{1}{2} \frac{\partial^2 V_1}{\partial r_t^2} g^2(r(t), t) - rV_1}{\frac{\partial V_1}{\partial r(t)}} = \frac{\frac{\partial V_2}{\partial t} + \frac{1}{2} \frac{\partial^2 V_2}{\partial r_t^2} g^2(r(t), t) - r(t)V_2}{\frac{\partial V_2}{\partial r(t)}} =: u(r(t), t),$$

and we obtain a PDE for the financial instrument V_1 depending on $r(t)$ given by

$$\frac{\partial V_1}{\partial t} + \frac{1}{2} g^2(r(t), t) \frac{\partial^2 V_1}{\partial r(t)^2} - u(r(t), t) \frac{\partial V_1}{\partial r(t)} - r(t) V_1 = 0.$$

There exist several well-known one-factor short-rate models, such as the Vasicek model [52], the Cox–Ingersoll–Ross model [53], or the Hull–White model [32, 36] which is an extension of the Vasicek model. The stochastic differential equation in the Hull–White model is given as

$$dr(t) = (a(t) - b(t)r(t)) dt + \sigma(t) dW(t),$$

with time-dependent parameters $a(t)$, $b(t)$, and $\sigma(t)$. The term $(a(t) - b(t)r(t))$ is a drift term and $a(t)$ is known as *deterministic drift*. Setting $g(r(t), t) = \sigma(t)$, and $-u(r(t), t) = (a(t) - b(t)r(t))$.

The Hull–White model is calibrated based on today's (t_0) market data for bond prices $B(t, T)$. To project the parameters $b(t)$ and $\sigma(t)$ into the future (t_1), one could use either $B(t_1, T + (t_1 - t_0))$ or $B(t_1, T)$ [14]. In the first case, the shape of parameters remains unchanged and does not cover seasonalities or expected changes like money market politics, while in the second case, we lose the information concentrated on the short end. When b and σ are constants, both approaches deliver the same parameters.

Appendix B: Yield curve simulation

The PRIIP regulation demands to perform yield curve simulations for at least 10,000 times [6], and that the data set must contain at least 2 years of daily interest rates for an underlying instrument, 4 years of weekly interest rates, or 5 years of monthly interest rates. We construct a data matrix $D \in \mathbb{R}^{n \times m}$ of the collected historical interest rates data, where each row of the matrix forms a yield curve, and the column represents the m tenor points, which are the different contract lengths of an underlying instrument. For example, if we have collected the daily interest rate data at $m \approx 20$ tenor points in time over the past five years, then since a year has approximately 260 working days, one obtains $n \approx 1306$ observation periods.

The regulations demand to take the natural logarithm of the ratio between the interest rate at each observation period and the interest rate at the preceding period. To ensure that we can form the natural logarithm, we need that all elements of the data matrix D are positive which is achieved by adding a correction term. With \mathcal{W} the matrix of all ones, we set $\tilde{D} = D + \gamma \mathcal{W}$, where γ is chosen so that all elements of matrix \tilde{D} are positive. We are compensating γ shift at the bootstrapping stage by subtracting it from the simulated rates. Then we calculate the log returns over each period and store them into a new matrix $\hat{D} = \hat{d}_{ij} \in \mathbb{R}^{n \times m}$ as

$$\hat{d}_{ij} = \frac{\ln(\tilde{d}_{ij})}{\ln(\tilde{d}_{i-1,j})}.$$

We calculate the arithmetic mean μ_j of each column of the matrix \hat{D} ,

$$\mu_j = \frac{1}{n} \sum_{i=1}^n \hat{d}_{ij},$$

subtract μ_j from each element of the corresponding j th column of \hat{D} and store the obtained results in a matrix $\bar{\bar{D}}$ with entries $\bar{\bar{d}}_{ij} = \hat{d}_{ij} - \mu_j$. We then compute the singular value

decomposition (SVD) [41],

$$\bar{\bar{D}} = \Phi \Sigma \Psi^T,$$

where Σ is a diagonal matrix having singular values Σ_i arranged in descending order. The columns of Φ are the normalized left singular vectors and the columns of $\Phi \Sigma$ are known as *principal components*. The columns of Ψ are the right singular vectors or *principal directions* of the covariance matrix $C = \bar{\bar{D}}^T \bar{\bar{D}}$.

The relative importance of the i th singular value is determined by the *relative energy*

$$\Xi_i = \frac{\Sigma_i}{\sum_{i=1}^m \Sigma_i},$$

where the *total energy* is given by $\sum_{i=1}^m \Xi_i = 1$. We then select the p right singular vectors corresponding to the maximal p energies and construct a matrix

$$\bar{\Psi} = \begin{bmatrix} \psi_{11} & \cdots & \psi_{1p} \\ \vdots & \vdots & \vdots \\ \psi_{m1} & \cdots & \psi_{mp} \end{bmatrix},$$

project the matrix $\bar{\bar{D}}$ onto the matrix Φ via

$$M_p = \bar{\bar{D}} \cdot \bar{\Psi} \in \mathbb{R}^{n \times p},$$

and then calculate the matrix of returns $M_R = M_p \bar{\Psi}^T \in \mathbb{R}^{n \times m}$. The regulations suggest selecting the first three ($p = 3$) right singular vectors. This process simplifies the statistical data $\bar{\bar{D}}$ and transforms m correlated tenor points into p uncorrelated principal components, reproducing the same data by simply reducing the total size of the model.

We then perform *bootstrapping*, where large numbers of small samples of the same size are drawn repeatedly from the original data set. According to the PRIIP regulations, for the yield curve simulation we have to perform a bootstrapping procedure for at least 10,000 times. The standardized KID also has to include the recommended *holding period*, i.e., the period between the acquisition of an asset and its sale. The time step in the simulation of yield curves is typically one observation period. If H is the recommended holding period in days, e.g., $H \approx 2600$ days, then there are H observation periods in the recommended holding period.

For each such observation period, we select a random row from the matrix M_R , i.e., altogether H random rows, and construct a matrix $[\chi_{ij}] \in \mathbb{R}^{H \times m}$ from these selected rows. Then we sum over the selected rows of the columns corresponding to the tenor point j , i.e.,

$$\bar{\chi}_j = \sum_{i=1}^h \chi_{ij}, \quad j = 1, \dots, m.$$

In this way, we obtain a row vector $\bar{\chi} = [\bar{\chi}_1 \bar{\chi}_2 \cdots \bar{\chi}_m] \in \mathbb{R}^{1 \times m}$. The final simulated yield rate y_j at tenor point j is then the rate \bar{d}_{nj} of the last observation period at the corresponding

tenor point j , multiplied by the exponential of $\bar{\chi}_j$, adjusted for any shift γ used to ensure positive values for all tenor points, and adjusted for the forward rate so that the expected mean matches current expectations.

The forward rate between time points t_k and t_ℓ starting from a time point t_0 is given as

$$r_{k,\ell} = \frac{R(t_0, t_\ell)(t_\ell - t_0) - R(t_0, t_k)(t_k - t_0)}{t_\ell - t_k},$$

where t_k and t_ℓ are measured in years and $R(t_0, t_k)$ and $R(t_0, t_\ell)$ are the interest rates available from the data matrix for the time periods (t_0, t_k) and (t_0, t_ℓ) , respectively. Thus, the final simulated yield curve between time points t_k and t_ℓ is given by

$$y(t_\ell) = \bar{d}_{k,\ell} \exp(\bar{\chi}_\ell) - \gamma + r_{k,\ell}, \quad \ell = 1, \dots, m, \quad (14)$$

and the simulated yield curve from the calculated simulated returns is given by

$$y = [y_1 y_2 \cdots y_m].$$

We then perform the bootstrapping procedure for at least $s = 10,000$ times and construct a simulated yield curve matrix

$$Y = \begin{bmatrix} y_{11} & \cdots & y_{1m} \\ \vdots & \vdots & \vdots \\ y_{s1} & \cdots & y_{sm} \end{bmatrix} \in \mathbb{R}^{s \times m}, \quad (15)$$

which is then used to calibrate the parameter $a(t)$.

Appendix C: Parameter calibration

For a *zero-coupon bond* $B(t, T)$ maturing at time T , based on the Hull–White model, one obtains a closed-form solution, see [54], as

$$B(t, T) = \exp\{-r(t)\Gamma(t, T) - \Lambda(t, T)\}, \quad (16)$$

where $\kappa(t) = \int_0^t b(s) ds = bt$, since b is assumed constant,

$$\begin{aligned} \Gamma(t, T) &= \int_t^T e^{-\kappa(t)} dt, \\ \Lambda(t, T) &= \int_t^T \left[e^{\kappa(v)} a(v) \left(\int_v^T e^{-\kappa(z)} dz \right) - \frac{1}{2} e^{2\kappa(v)} \sigma^2 \left(\int_v^T e^{-\kappa(z)} dz \right)^2 \right] dv. \end{aligned}$$

Here we have again used that σ is constant.

To perform the calibration, we use as input data i) the initial value of $a(0)$ at $t = 0$, ii) the zero-coupon bond prices, iii) the constant value of the volatility σ of the short-rate $r(t)$, and iv) the constant value b each for all maturities T_m , $0 \leq T_m \leq T$, where T_m is the maturity at the m th tenor point. Then we compute $\kappa(t)$ from $\frac{\partial}{\partial T} \kappa(T) = \frac{\partial}{\partial T} \int_0^T b(s) ds = b$

and use

$$\frac{\partial}{\partial T} \Gamma(0, T) = e^{-\kappa(T)}$$

to compute $\Gamma(t)$.

Then, for $0 \leq T_m \leq T$, we get

$$\begin{aligned} \frac{\partial}{\partial T} \Lambda(0, T) &= \int_0^T \left[e^{\kappa(v)} a(v) e^{-\kappa(T)} - e^{2\kappa(v)} \sigma^2 e^{-\kappa(T)} \left(\int_v^T e^{-\kappa(z)} dz \right) \right] dv, \\ e^{\kappa(T)} \frac{\partial}{\partial T} \Lambda(0, T) &= \int_0^T \left[e^{\kappa(v)} a(v) - e^{2\kappa(v)} \sigma^2 \left(\int_v^T e^{-\kappa(z)} dz \right) \right] dv, \\ \frac{\partial}{\partial T} \left[e^{\kappa(T)} \frac{\partial}{\partial T} \Lambda(0, T) \right] &= e^{\kappa(T)} a(T) - \int_0^T e^{2\kappa(v)} \sigma^2 e^{-\kappa(T)} dv, \\ e^{\kappa(T)} \left[e^{\kappa(T)} \frac{\partial}{\partial T} \Lambda(0, T) \right] &= e^{2\kappa(T)} a(T) - \int_0^T e^{2\kappa(v)} \sigma^2 dv, \\ \frac{\partial}{\partial T} \left[e^{\kappa(T)} \left[e^{\kappa(T)} \frac{\partial}{\partial T} \Lambda(0, T) \right] \right] &= \frac{\partial a(T)}{\partial T} e^{2\kappa(T)} + 2a(T) e^{2\kappa(T)} \frac{\partial}{\partial T} \kappa(T) - e^{2\kappa(T)} \sigma^2, \\ \frac{\partial}{\partial T} \left[e^{\kappa(T)} \left[e^{\kappa(T)} \frac{\partial}{\partial T} \Lambda(0, T) \right] \right] &= \frac{\partial a(T)}{\partial T} e^{2\kappa(T)} + 2a(T) e^{2\kappa(T)} b(T) - e^{2\kappa(T)} \sigma^2. \end{aligned}$$

The simulated yield $y(T)$ at the tenor point T is then given by

$$y(T) = -\ln B(0, T), \quad (17)$$

and from (17) we obtain $\Lambda(0, T) = [y(T) - r(0)\Gamma]$. In this way, for $a(t)$ we obtain the ordinary differential equation (ODE)

$$\frac{\partial}{\partial t} a(t) e^{2\kappa(t)} + 2a(t) \cdot b \cdot e^{2\kappa(t)} - e^{2\kappa(t)} \sigma^2 = \frac{\partial}{\partial t} \left[e^{\kappa(t)} \left[e^{\kappa(t)} \frac{\partial}{\partial t} (y(t) - r(0)\Gamma(0, t)) \right] \right],$$

which we solve numerically with the given initial conditions. If we approximate $a(t)$ by a piecewise constant function with values $a(i)$ which change at the tenor point i , then we obtain a linear system

$$L\alpha = F,$$

for the vector $\alpha = [a(i)]$, where L is lower triangular with non-zero diagonal elements. In [55] it is noted that the integral equation Λ is of the first kind with L2 kernel and a small perturbation (noise) in the market data that are used to obtain the yield curves leads to large changes in the model parameter $a(t)$. This means that the problem to compute $a(t)$ from the data is an ill-posed problem and for this reason we determine the vector α via Tikhonov regularization as

$$\alpha_\mu^\delta = \operatorname{argmin} \|L\alpha - F^\delta\|^2 + \mu \|\alpha\|^2, \quad (18)$$

where α_μ^δ is an approximation to α , μ is the regularization parameter, $\delta = \|F - F^\delta\|$ is the noise level, and $\mu \|\alpha\|^2$ is a regularization term. We then solve the optimization problem

(18) to obtain an approximation to the parameter $a(t)$ via the commercial software *UnRisk PRICING ENGINE* for the parameter calibrations [46]. By providing the simulated yield curve, the UnRisk pricing function returns the calibrated parameter $a(t)$ for that yield curve. Based on $s = 10,000$ different simulated yield curves, we obtain s different piecewise constant parameters $a_\ell(t)$, which change their values $\alpha_{\ell,i}$ only at the m tenor points. We incorporate these in a matrix

$$\mathcal{A} = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \vdots & \vdots & \vdots \\ \alpha_{s1} & \cdots & \alpha_{sm} \end{bmatrix}. \quad (19)$$

Appendix D: Numerical methods

The Hull–White model (3) is discretized by applying a finite difference method. As computational domain for the interest rate $r(t)$ we use an interval $[r_{\text{low}}, r_{\text{up}}]$, according to [1] given by

$$r_{\text{low}} = r(T) - 7\sigma\sqrt{T}, \quad r_{\text{up}} = r(T) + 7\sigma\sqrt{T},$$

where $r(T)$ is the yield at the maturity T also known as a *spot rate*. We divide the spatial domain into M equidistant grid points $\{r(1), r(2), \dots, r(M)\}$, $r(i) = r(i-1) + h$ with spacial step size h , and the time interval $[0, T]$ in N points $t_0 = 0, t_1, \dots, t_N = T$, $t_n = n\tau$, with time step τ . Using the spatial discretization operator $\mathcal{L}(n)$ at time point n , we get a system of ODEs for the vector $V = [V_i] = [V(r(i))]$ of values at the spatial grid points

$$\frac{V(t) - V(t - \tau)}{\tau} = (1 - \Theta)(\mathcal{L}(t)V(t)) + \Theta(\mathcal{L}(t - \tau)V(t - \tau)),$$

which is given componentwise by

for $(a(n) - br(i)) > 0$

$$\mathcal{L}(n)V_i^n := \frac{1}{2}\sigma^2 \frac{V_{i+1}^n - 2V_i^n + V_{i-1}^n}{h^2} + (a(n) - br(i)) \frac{V_i^n - V_{i-1}^n}{h} - r(i)V_i^n,$$

for $(a(n) - br(i)) < 0$

$$\mathcal{L}(n)V_i^n := \frac{1}{2}\sigma^2 \frac{V_{i+1}^n - 2V_i^n + V_{i-1}^n}{h^2} + (a(n) - br(i)) \frac{V_{i+1}^n - V_i^n}{h} - r(i)V_i^n,$$

so that the Crank–Nicolson scheme in time gives a linear system

$$\underbrace{\left(1 - \frac{1}{2}\tau\mathcal{L}(t - \tau)\right)}_{A(\rho_\ell(t)) \in \mathbb{R}^{M \times M}} V(t - \tau) = \underbrace{\left(1 + \frac{1}{2}\tau\mathcal{L}(t)\right)}_{B(\rho_\ell(t)) \in \mathbb{R}^{M \times M}} V(t),$$

where the matrices $A(\rho_\ell(t))$, and $B(\rho_\ell(t))$ depend on $\rho_\ell(t) = \{a(t), b, \sigma\}$ for the ℓ th group of these parameters and are given by

$$A(\rho_\ell(t)) = I - \frac{\sigma^2\tau}{2h^2}J - \frac{\tau}{2h}(H^+G + H^-G^T) + R_o,$$

and

$$B(\rho_\ell(t)) = I + \frac{\sigma^2 \tau}{2h^2} J + \frac{\tau}{2h} (H^+ G + H^- G^T) - R_o,$$

where $R_o = \text{diag}(r(1), \dots, r(M))$, $H^+ = \text{diag}(\max(a(n) - br(1)), \dots, \max(a(n) - br(M)))$, $H^- = \text{diag}(\min(a(n) - br(1)), \dots, \min(a(n) - br(M)))$, and

$$J = \begin{bmatrix} -2 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \ddots & \vdots \\ 0 & 1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & 1 & -2 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \ddots & \vdots \\ 0 & -1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}.$$

This discretization is a parametric high dimensional model of the form

$$A(\rho_\ell(t)) V^{n-1} = B(\rho_\ell(t)) V^n,$$

with given terminal vector V^T , and matrices $A(\rho_\ell) \in \mathbb{R}^{M \times M}$, and $B(\rho_\ell) \in \mathbb{R}^{M \times M}$. We solve this model by propagating backward in time. Here again $\ell = 1, \dots, s = 10,000$, m is the total number of tenor points, and we need to solve this system at each time step n with an appropriate boundary condition and a known terminal value for the underlying instrument. Altogether we have a parameter space \mathcal{P} of size $10,000 \times m$ to which we now apply model reduction.

Acknowledgements

We thank our colleagues Michael Schwaiger and Diana Hufnagl from MathConsult who provided insight and expertise that greatly assisted the research.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 765374. Open Access funding enabled and organized by Projekt DEAL.

Abbreviations

VaR, Value-at-Risk; CVA, credit value adjustments; DVA, debt value adjustments; PRIIPs, packaged retail and insurance-based investment products; EU, European Union; KID, Key information document; RHP, recommended holding period; PDE, partial differential equation; POD, proper orthogonal decomposition; MOR, model order reduction; PCR, principal component regression; DAX, Deutscher Aktienindex; VEV, VaR equivalent volatility; DF, discount factor; MRM, market risk indicator; SVD, Singular value decomposition; PCA, principal component analysis; CG, classical greedy; AG, Adaptive greedy.

Availability of data and materials

The interest rate data that support the findings of this study are available from MathConsult within their UnRisk Omega datasets but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Authors' contributions

OJ and VM conceived the presented idea of the model order reduction. OJ developed the theoretical formalism and performed the numerical simulations. AB wrote the introduction and provided valuable insights for the financial part. VM encouraged OJ to investigate surrogate modeling and supervised the findings of the work. AB and VM supervised the project. All authors discussed the results and contributed to the final manuscript. All authors read and approved the final manuscript.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 21 July 2020 Accepted: 24 May 2021 Published online: 03 June 2021

References

1. Aichinger M, Binder A. A workout in computational finance. 1st ed. West Sussex: Wiley; 2013.
2. Jorion P. Financial risk manager handbook. 2nd ed. New York: Wiley; 2007.
3. Wilmott P, Howison S, Dewynne J. The mathematics of financial derivatives: a student introduction. 1st ed. London: Cambridge University Press; 2002.
4. Jorion P. Counterparty credit risk and credit value adjustment: a continuing challenge for global financial markets. 2nd ed. New York: Wiley; 2012.
5. European Commission. Commission delegated regulation (EU) 1286/2014. Off J EU. 2014;1:1–23.
6. European Commission. Commission delegated regulation (EU) 2017/653. Off J EU. 2017;1:1–52.
7. Heston S. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev Financ Stud*. 1993;6(2):327–43.
8. Longstaff F, Valuing SE. American options by simulation: a simple least-squares approach. *Rev Financ Stud*. 2001;14(1):113–47.
9. Rebonato R, McKay K, White R. The SABR/LIBOR market model: pricing, calibration and hedging for complex interest-rate derivatives. 1st ed. New York: Wiley; 2009.
10. Briani M, Caramellino L, Zanette A. A hybrid tree/finite-difference approach for Heston–Hull–White-type models. *J Comput Finance*. 2017;21(3).
11. Falcó A, Navarro L, Cendón CV. Finite difference methods for Hull–White pricing of interest rate derivatives with dynamical consistent curves. *Soc Sci Res Net Elec J*. 2014.
12. Haentjens T, In't Hout K. Direction implicit finite difference schemes for the Heston–Hull–White partial differential equation. *J Comput Finance*. 2012;16:83–110.
13. Brigo D, Mercurio F. Interest rate models—theory and practice. 2nd ed. Berlin: Springer; 2006.
14. Hull J, White A. Numerical procedures for implementing term structure models I: single-factor models. *J Deriv*. 1994;2:7–16.
15. Black F, Karasinski P. Bond and option pricing when short rates are lognormal. *Financ Anal J*. 1991;47(4):52–9.
16. Hull J, White A. The general Hull–White model and supercalibration. *Financ Anal J*. 2001;57(6):34–43.
17. Berkooz G, Holmes P, Lumley JL. The proper orthogonal decomposition in the analysis of turbulent flows. *Annu Rev Fluid Mech*. 1993;25(1):539–75.
18. Chatterjee A. An introduction to the proper orthogonal decomposition. *Curr Sci*. 2000;78:808–17.
19. Graham MD, Kevrekidis IG. Alternative approaches to the Karhunen–Loève decomposition for model reduction and data analysis. *Comput Chem Eng*. 1996;20(5):495–506.
20. Jolliffe I. Principal component analysis. 1st ed. Berlin: Springer; 2014.
21. Graham MD, Proper KIG. Orthogonal decomposition and its applications part I: theory. *J Sound Vib*. 2002;252(3):527–44.
22. Rathinam M, Petzold LR. A new look at proper orthogonal decomposition. *SIAM J Numer Anal*. 2003;41(5):1893–925.
23. Vidal A, Sakrison D. On the optimality of the Karhunen–Loève expansion (corresp.). *IEEE Trans Inf Theory*. 1969;15(2):319–21.
24. Sirovich L. Turbulence and the dynamics of coherent structures. Part I: coherent structures. *Q Appl Math*. 1987;45(3):561–71.
25. Lucia JD, Beran PS, Silva WA. Reduced order modeling: new approaches for computational physics. *Prog Aerosp Sci*. 2004;40(1–2):51–117.
26. Amsallem D, Zahr M, Choi Y. Design optimization using hyper-reduced-order models. *Struct Multidiscip Optim*. 2015;51(4):919–40.
27. Grepl M, Patera A. A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *M2AN, Math Model Numer Anal*. 2005;39:157–81.
28. Paul-Dubois-Taine A, Amsallem D. An adaptive and efficient greedy procedure for the optimal training of parametric reduced-order models. *Int J Numer Methods Eng*. 2014;102:1262–92.
29. Bui-Thanh T, Willcox K, Model GO. Reduction for large-scale systems with high-dimensional parametric input space. *SIAM J Sci Comput*. 2008;30(6):3270–88.
30. Veroy K, Patera A. Certified real-time solution of the parametrized steady incompressible Navier–Stokes equations: rigorous reduced-basis a posteriori error bounds. *Int J Numer Methods Fluids*. 2005;47:773–88.
31. Lee H, Park Y, Lee S. Principal component regression by principal component selection. *Commun Stat Appl Methods*. 2015;22(2):173–80.
32. Hull J, Pricing WA. Interest-rate-derivative securities. *Rev Financ Stud*. 1990;3(4):573–92.
33. Fabozzi F. Valuation of fixed income securities and derivatives. 3rd ed. New York: Wiley; 1998.
34. Binder A, Jadhav O, Mehrmann V. Model order reduction for parametric high dimensional models in the analysis of financial risk [Technical Report]. 2020. arXiv.
35. European Supervisory Authorities [pdf]; 2017. Available from: <https://esas-joint-committee.europa.eu/Publications/Technical>

36. Hull J, White A. One-factor interest-rate models and the valuation of interest-rate derivative securities. *J Financ Quant Anal.* 1993;28(2):235–54.
37. Anderson D, Tannehill JC, Pletcher RH. Computational fluid mechanics and heat transfer. 3rd ed. London: CRC Press; 2013.
38. Darbellay GA. A note on the volatility term structure in short rate models. *J Appl Math Mech.* 1998;78:885–6.
39. Heinrich JC, Huyakorn PS, Zienkiewicz OC, Mitchell AR. An 'upwind' finite element scheme for two-dimensional convective transport equation. *Int J Numer Methods Eng.* 1977;11:131–43.
40. Sun G, Trueman CW. Efficient implementations of the Crank–Nicolson scheme for the finite-difference time-domain method. *IEEE Trans Microw Theory Tech.* 1977;11:131–43.
41. Golub GH, Singular RC. Value decomposition and least squares solutions. *Numer Math.* 1970;24:403–20.
42. Williams MO, Schmid PJ, Hybrid KJN. Reduced-order integration with proper orthogonal decomposition and dynamic mode decomposition. *SIAM J Multiscale Model Simul.* 2013;11(2):522–44.
43. Model PR. Reduction via proper orthogonal decomposition. In: Schilders W, van der Vorst H, Rommes J, editors. *Model order reduction: theory, research aspects and applications.* Berlin: Springer; 2008. p. 95–109.
44. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. 1st ed. New York: Springer; 2013.
45. Binder A. A clever handful is enough. *Wilmott.* 2007. 10–14.
46. MathConsult. Calibration of interest rate models. Linz: MathConsult GmbH; 2009.
47. Dudley RM. Unifrom central limit theorems. 1st ed. New York: Cambridge University Press; 2010.
48. Corhay A, Rad AT. Statistical properties of daily returns: evidence from European stock markets. *J Bus Finance Account.* 1994;21(2):271–82.
49. Albrecher H, Binder A, Lautscham V, Mayer P. Introduction to quantitative methods for financial markets. 1st ed. Berlin: Springer; 2010.
50. Mahnke R, Kaupuzs J, Lubashevsky I. Physics of stochastic processes: how randomness acts in time. 1st ed. Weinheim: Wiley-VCH Verlag GmbH and Co.; 2008.
51. Grigoriu M. Stochastic calculus: applications in science and engineering. 1st ed. Basel: Birkhäuser; 2002.
52. Vasicek O. An equilibrium characterization of the term structure. *J Financ Econ.* 1977;5(2):177–88.
53. Cox JC, Ingersoll Jr JE, Ross SA. A theory of the term structure of interest rates. *Econometrica.* 1985;53(2):385–407.
54. Shreve SE. Stochastic calculus and finance. 1st ed. New York: Springer; 2004.
55. Engl H. Calibration problems—An inverse problems view. *Wilmott.* 2007. 16–20.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)