The stochastic collocation Monte Carlo sampler
Highly efficient sampling from 'expensive' distributions

Grzelak, L.A.; Witteveen, J.A.S.; Oosterlee, C.W.; Suárez-Taboada, M.

DOI
10.1080/14697688.2018.1459807

Publication date
2018

Published in
Quantitative Finance

Citation (APA)
Grzelak, L. A., Witteveen, J. A. S., Oosterlee, C. W., & Suárez-Taboada, M. (2018). The stochastic collocation Monte Carlo sampler: Highly efficient sampling from 'expensive' distributions. *Quantitative Finance*, 1-18. https://doi.org/10.1080/14697688.2018.1459807

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: https://www.tandfonline.com/loi/rquf20

The stochastic collocation Monte Carlo sampler: highly efficient sampling from 'expensive' distributions

L. A. Grzelak, J. A. S. Witteveen, M. Suárez-Taboada & C. W. Oosterlee

To cite this article: L. A. Grzelak, J. A. S. Witteveen, M. Suárez-Taboada & C. W. Oosterlee (2019) The stochastic collocation Monte Carlo sampler: highly efficient sampling from 'expensive' distributions, Quantitative Finance, 19:2, 339-356, DOI: 10.1080/14697688.2018.1459807

To link to this article: https://doi.org/10.1080/14697688.2018.1459807

Published online: 08 Jun 2018.

Quantitative Finance, 2019
Vol. 19, No. 2, 339–356, https://doi.org/10.1080/14697688.2018.1459807

The stochastic collocation Monte Carlo sampler: highly efficient sampling from 'expensive' distributions

L. A. GRZELAK*†‡, J. A. S. WITTEVEEN§, M. SUÁREZ-TABOADA¶ and C. W. OOSTERLEE†§

†Delft Institute of Applied Mathematics (DIAM), TU Delft, Delft University of Technology, Delft, The Netherlands
‡Financial Engineering Department, Rabobank Nederland, Utrecht, The Netherlands
§Centrum Wiskunde & Informatica (CWI)-National Research Institute for Mathematics and Computer Science, Amsterdam, The Netherlands
¶Department of Mathematics, University of A Coruña, Campus Elviña s/n, 15071 A Coruña, Spain

(*Received 21 October 2016; accepted 27 March 2018; published online 8 June 2018*)

In this article, we propose an efficient approach for inverting computationally expensive cumulative distribution functions. A collocation method, called the Stochastic Collocation Monte Carlo sampler (SCMC sampler), within a polynomial chaos expansion framework, allows us the generation of any number of Monte Carlo samples based on only a few inversions of the original distribution plus independent samples from a standard normal variable. We will show that with this path-independent collocation approach the exact simulation of the Heston stochastic volatility model, as proposed in Broadie and Kaya [*Oper. Res*., 2006, **54**, 217–231], can be performed efficiently and accurately. We also show how to efficiently generate samples from the squared Bessel process and perform the exact simulation of the SABR model.

Keywords: Exact sampling; Heston; Squared Bessel; SABR; Stochastic collocation; Lagrange interpolation; Monte Carlo

1. Introduction

In computational finance, we often require rapid and accurate approximations of functions of stochastic variables. As an example, think of the rapid evaluation of the integrated variance process, which is a function of the variance process. The integrated variance plays a role within the exact Heston stochastic volatility Monte Carlo simulation scheme, by Broadie and Kaya (2006). Another example is represented by arithmetic Asian options where an accurate representation of integrated asset prices is sought.

For these quantities, we generally do not know analytic closed-form solutions, so numerical approximations are developed to estimate them. The Monte Carlo (MC) method is considered an accurate simulation technique to approximate these quantities, but quite a few samples are required to obtain sufficient accuracy.

In another context, Uncertainty Quantification (UQ) techniques are typically employed to uncover the probabilistic dependence of solutions to partial differential equations on uncertainty in boundary conditions or model parameters (Witteveen and Iaccarino 2013). The MC method can also be used to quantify the impact of uncertainty on PDE solutions. However, also in this case MC methods require a large number of samples.

The Stochastic Collocation (SC) method (Xiu and Hesthaven 2005, Nobile *et al*., 2008a, 2008b, Bieri and Schwab 2009, Babuška *et al*. 2010, Beck *et al*. 2012) has been developed as an efficient alternative for MC methods based on deterministic sampling at quadrature points and Lagrangian polynomial interpolation. The main idea of SC is to project the uncertainty onto a probability space with known properties and conditions. Suitable basis functions are determined for this space and suitable interpolation points are computed, based on the input distribution. With an increasing polynomial order of the expansion, exponential convergence can be obtained.

In this paper, we will apply the SC method for the efficient generation of samples of a stochastic variable with an expensive distribution. To that end, we introduce another stochastic variable with a distribution that is much cheaper to evaluate and we use SC to approximate the relation between these two

*Corresponding author. Email: L.A.Grzelak@TUDelft.nl

stochastic variables, thus avoiding many expensive computations or simulations. *We will use the standard normal variable for the cheap collocation variable here*. One specific aim is to efficiently sample the integrated variance process from a Heston stochastic volatility model and provide an accurate simulation scheme for the dynamics of the stock price process with large time steps. Although an (involved) characteristic function is in principle available for the integrated variance, the computation of the distribution of the integrated variance remains an expensive computational task where Fourier inversion techniques to invert the characteristic function need to be employed.

A variant of the sampling technique for processes that have significant mass at zero is also discussed. Stochastic processes for which the underlying asset may *hit and stay* at zero are often encountered in finance. This feature is desired as it describes realistically stock movements and the possibility of default events, i.e. once an asset price reaches the zero value it may not recover in the future. We discuss two particular examples of such models, the squared-Bessel process which, for certain sets of parameters, models absorption at zero and the SABR model in which also the volatility is modeled by a stochastic process. We will see that the SC expansion can be efficiently applied to these problems.

The paper is organised as follows: section 2 introduces a general, basic description of the SC method and the specific methodology used. An error analysis is presented in sections 3 and 4 presents analytic tests to show the performance of the method and numerical results for the squared Bessel, the Heston and the SABR models. Conclusions are drawn in section 5. Basic background information on stochastic collocation and tabulated values are placed in the appendices.

2. Lagrange polynomials and stochastic collocation

We consider two scalar random variables X and Y. The original idea behind the SC method is to approximate a problem variable of interest, Y, as function of another random variable X. Random variable X is governed by a probability density function (PDF) $f_X(x)$ and $F_X(x)$ is its cumulative distribution function (CDF) with $F_X(x) \in [0, 1]$. As $f_X(x)$ is not zero in the interior of its domain, $F_X(x)$ is strictly monotonic and as a consequence the transformation between the original probability space and the new space is bijective. For X we choose the standard normal variable here. In the general setting however, when X is not normally distributed, we would require the first $2N$ moments, where N is the number of collocation points, to exist and that for $X : \Omega_X \to \mathbb{R}$ and $Y : \Omega_Y \to \mathbb{R}$ we have $\Omega_Y \subseteq \Omega_X$.

2.1. Sampling by the stochastic collocation method

We consider the problem of generating a sample from a random variable Y with a strictly monotonic CDF $F_Y(y)$ and its PDF given by $f_Y(y)$. $F_Y(Y)$ is uniformly distributed† on [0, 1], and we can use a standard approach to generate uniform samples u_n and any desired sample by inverting the CDF,

$$F_Y(Y) \overset{\mathrm{d}}{=} U, \quad \text{thus} \quad y_n = F_Y^{-1}(u_n), \tag{2.1}$$

for $U \sim \mathcal{U}([0, 1])$, and u_n is a sample from $\mathcal{U}([0, 1])$. Inverting the CDF in (2.1) is typically computationally expensive, especially when the inversion of the CDF of the random variable Y is not known analytically. In such a case, it is important to reduce the number of expensive inversions as much as possible.

We introduce another random variable, X, for which the inversion $F_X^{-1}(u_n)$ is computationally less expensive than the one in (2.1). As both $F_X(X)$ and $F_Y(Y)$ are uniformly distributed we have $F_Y(Y) \overset{\mathrm{d}}{=} F_X(X)$. This equality in distribution does not imply that X and Y are equal in distribution, but only the CDFs can be equated to the same uniform distribution. Samples of Y, y_n, and X, ξ_n, are related via the following inversion,

$$y_n = F_Y^{-1}(F_X(\xi_n)). \tag{2.2}$$

Obviously, the sampling via (2.2) is considered expensive as for each *cheap* realisation of X one needs to calculate the expensive inverse of the CDF of Y. The objective is to find an alternative relation which does not require the inversions $F_Y^{-1}(\cdot)$ for all samples of X.

The task is to find a function $g(\cdot) = F_Y^{-1}(F_X(\cdot))$, i.e. $F_X(x) = F_Y(g(x))$ which yields $Y \overset{\mathrm{d}}{=} g(X)$, where the evaluations of function $g(\cdot)$ do not require expensive inversions $F_Y^{-1}(\cdot)$, as in (2.2). Once we determine the mapping function $g(\cdot)$, the CDFs $F_X(x)$ and $F_Y(g(x))$ are equal not only in distribution sense but also element-wise.

With $g(\cdot)$ determined, the sampling from the *expensive* random variable Y can be decomposed into sampling from a *cheap* random variable X and a transformation to Y via an evaluation of function $g(\cdot)$, i.e. $y_n = g(\xi_n)$. It is therefore crucial that $g(\cdot)$ is as simple as possible.

The SC method is used here to efficiently approximate $g(\cdot)$. The method approximates Y as a function g of X in terms of an expansion in Lagrange polynomials $\ell_i(\xi_n)$, i.e.

$$y_n \approx g_N(\xi_n) = \sum_{i=1}^{N} y_i \ell_i(\xi_n), \qquad \ell_i(\xi_n) = \prod_{j=1,\ i\neq j}^{N} \frac{\xi_n - x_j}{x_i - x_j}, \tag{2.3}$$

where ξ_n is a sample from X and x_i and x_j are so-called *collocation points*, y_i is the exact evaluation at a collocation point x_i given in (2.2), i.e. $y_i = F_Y^{-1}(F_X(x_i))$. In the expression above $\boldsymbol{\ell}(x) = (\ell_1(x), \ell_2(x), \ldots, \ell_N(x))^{\mathrm{T}}$ is called the Lagrange basis. Each element of the basis satisfies $\ell_i(x_j) = \delta_{ij}$, with the Kronecker delta $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ otherwise. A special choice of the collocation points x_i will be discussed. We take a closer look at the equation (2.3). Once we have determined N collocation points x_i and N expensive inversions $F_Y^{-1}(F_X(x_i))$, we can generate any number of samples by evaluating the polynomial $g_N(\xi_n)$. The resulting collocation method is called the *Stochastic Collocation Monte Carlo sampler (SCMC sampler)* here.

For any random variable X, the optimal collocation points are based on the moments of X. The optimal collocation points are chosen to be Gauss quadrature points that are defined as the zeros of the corresponding orthogonal polynomial. This leads to a stable interpolation under the probability distribution of X.

†Consider $u \in [0, 1]$ for which we have $\mathbb{P}[F_Y(Y) \leq u] = \mathbb{P}[Y \leq F_Y^{-1}(u)] = F_Y(F_Y^{-1}(u)) = u$. This equals the CDF of the uniform distribution on [0, 1], i.e. $F_U(u) = u$ for $U \sim \mathcal{U}([0, 1])$.

A brief introduction to Lagrange polynomials, to the optimal points based on the moments of X by means of the so-called Gram matrix M, and a basic example involving the collocation method are presented, for convenience, in appendix 1.

Remark [Monotonicity of $g_N(x)$] With the collocation points x_i and the corresponding inversions $y_i = F_Y^{-1}(F_X(x_i))$, the task is to construct an approximating function $g_N(x)$ which is (ideally) monotonic, differentiable and which satisfies $y_i = g_N(x_i)$. In this paper we choose for $g_N(x)$ the Lagrange polynomial which is a well-accepted choice in the field of Uncertainty Quantification. With this choice of polynomial, monotonicity is not guaranteed; however, the convergence of the SCMC sampler does not really depend on the monotonicity of $g_N(x)$.

Within $[x_i, x_N]$ monotonicity of the function $g_N(x)$ can be guaranteed by the use of e.g. the monotone cubic Hermite interpolation (Fritsch and Carlson 1980) which also guarantees continuity of the first derivative. The resulting formulas for the occurring expectations (for which we have analytic solutions with the Lagrange polynomials) are somewhat more involved in the case of monotone cubic Hermite interpolations which is the main reason for us to stay with the Lagrange interpolation.

2.1.1. Sampling from the non-central chi squared distribution. As a first numerical example, we consider the problem of generating samples from the non-central chi squared distribution $Y \sim \chi^2(d, \lambda)$ with the number of degrees of freedom parameter d and non-centrality parameter λ. We perform an experiment with the following set of parameters,† $d = 1.2$ and $\lambda = 0.1$.

We take $X \sim \mathcal{N}(0, 1)$, with the moments given by:

$$M = \{\mathbb{E}[X^{i+j}]\}_{i,j=0}^{N} = \begin{cases} 0, & i+j \text{ is odd}, \\ (i+j-1)!!, & i+j \text{ is even}, \end{cases} \tag{2.4}$$

with $\mathbb{E}[X^0] = 1$ and where !! is the so-called double factorial.‡ For $N = 5$ we determine the collocation points for X based on its moments. The resulting moment matrices M and $\widehat{J}$ (see appendix 1) are given by:

$$M = \begin{pmatrix} 1 & 0 & 1 & 0 & 3 & 0 \\ 0 & 1 & 0 & 3 & 0 & 15 \\ 1 & 0 & 3 & 0 & 15 & 0 \\ 0 & 3 & 0 & 15 & 0 & 105 \\ 3 & 0 & 15 & 0 & 105 & 0 \\ 0 & 15 & 0 & 105 & 0 & 945 \end{pmatrix},$$

$$\widehat{J} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1.4142 & 0 & 0 \\ 0 & 1.4142 & 0 & 1.7321 & 0 \\ 0 & 0 & 1.7321 & 0 & 2 \\ 0 & 0 & 0 & 2 & 0 \end{pmatrix},$$

where matrix $\widehat{J}$ is a symmetric tridiagonal matrix obtained from the eigenvalue method (see theorem 1.2) in which the matrix coefficients are obtained from the Cholesky decomposition of Gram matrix M, as described in detail in appendix A.1.

†Note that the non-central distribution is well defined for any positive number of degrees of freedom, d (Johnson *et al.* 1994).

‡$n!! = n(n-2)(n-4)\ldots$

Table 1. The collocation points x_i, corresponding CDF and inversions $F^{-1}_{\chi^2(d,\lambda)}(x_i)$, for $i = 1, \ldots, 5$.

	x_i	$F_{\mathcal{N}(0,1)}(x_i)$	$y_i = F^{-1}_{\chi^2(d,\lambda)}(F_{\mathcal{N}(0,1)}(x_i))$
x_1	−2.8570	0.0021	0.000063961434589
x_2	−1.3556	0.0876	0.031420172480241
x_3	0	0.5	0.685785887466036
x_4	1.3556	0.9124	3.623925068433782
x_5	2.8570	0.9979	10.846256627398553

For the standard normal random variable the collocation points have been tabulated in table A1 in appendix A.3. The corresponding values of the *cheap* CDF $F_{\mathcal{N}(0,1)}(\boldsymbol{x})$ are given in table 1. We perform $N = 5$ *expensive* inversions,§ $y_i = F^{-1}_{\chi^2(d,\lambda)}(F_{\mathcal{N}(0,1)}(x_i)))$, $i = 1, \ldots, 5$, also given in table 1. When the expensive inversions are determined, we obtain a vector of samples $\boldsymbol{y}_n$.

With the approximating polynomial, $g_N(\xi_n)$, it is not guaranteed that the obtained realisations are non-negative. In such a case one can either ignore any negative values, using $g(\xi_n) := |g(\xi_n)|$, or cap the realisations at 0, i.e. $g_N(\xi_n) := \max(g_N(\xi_n), 0)$.

The results are presented in figure 1, where $N = 5$ collocation points are considered. We see that when the distribution of interest, Y, exhibits significant mass in the neighborhood of zero, the number of collocation points should be increased for accuracy reasons. Interpolation based on $N = 5$ already provides excellent results in this rather extreme case. In table 2 the timing results, performed on an i5-2400 with 4 GB of RAM, are presented. These results show that for simulating 1.000.000 samples the SCMC sampler is about 3 times faster than MATLAB.

3. Error analysis

In this section, we discuss the errors generated by the SCMC sampler. We begin with a case for which the collocation method gives exact results. In example 3.1 below the results for a normal distribution are presented.

Example 3.1 [Exact solution for normal distribution with $N = 2$] For any two normally distributed random variables $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ and $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ the expansion in (2.3) is exact in distribution, i.e. $g_N(X) \overset{\mathrm{d}}{=} Y$, for $N = 2$.

Choosing two collocation points, x_1 and x_2, equation (2.3) gives for function $g_2(X)$,

$$g_2(X) = y_1 \frac{X - x_2}{x_1 - x_2} + y_2 \frac{X - x_1}{x_2 - x_1}. \tag{3.1}$$

Since both random variables are normally distributed we can transform them into standard normal variables, i.e. $F_{\mathcal{N}(0,1)}\left(\frac{y_i - \mu_Y}{\sigma_Y}\right) = F_{\mathcal{N}(0,1)}\left(\frac{x_i - \mu_X}{\sigma_X}\right)$. So, for each collocation point x_i, we have $y_i = \frac{x_i - \mu_X}{\sigma_X}\sigma_Y + \mu_Y$, in equation (3.1). Since $g_2(X)$ follows the normal distribution, $\mathbb{E}[g_2(X)] = \mu_Y$ and

§These inversions were calculated with the MATLAB function $ncx2inv(\cdot)$.

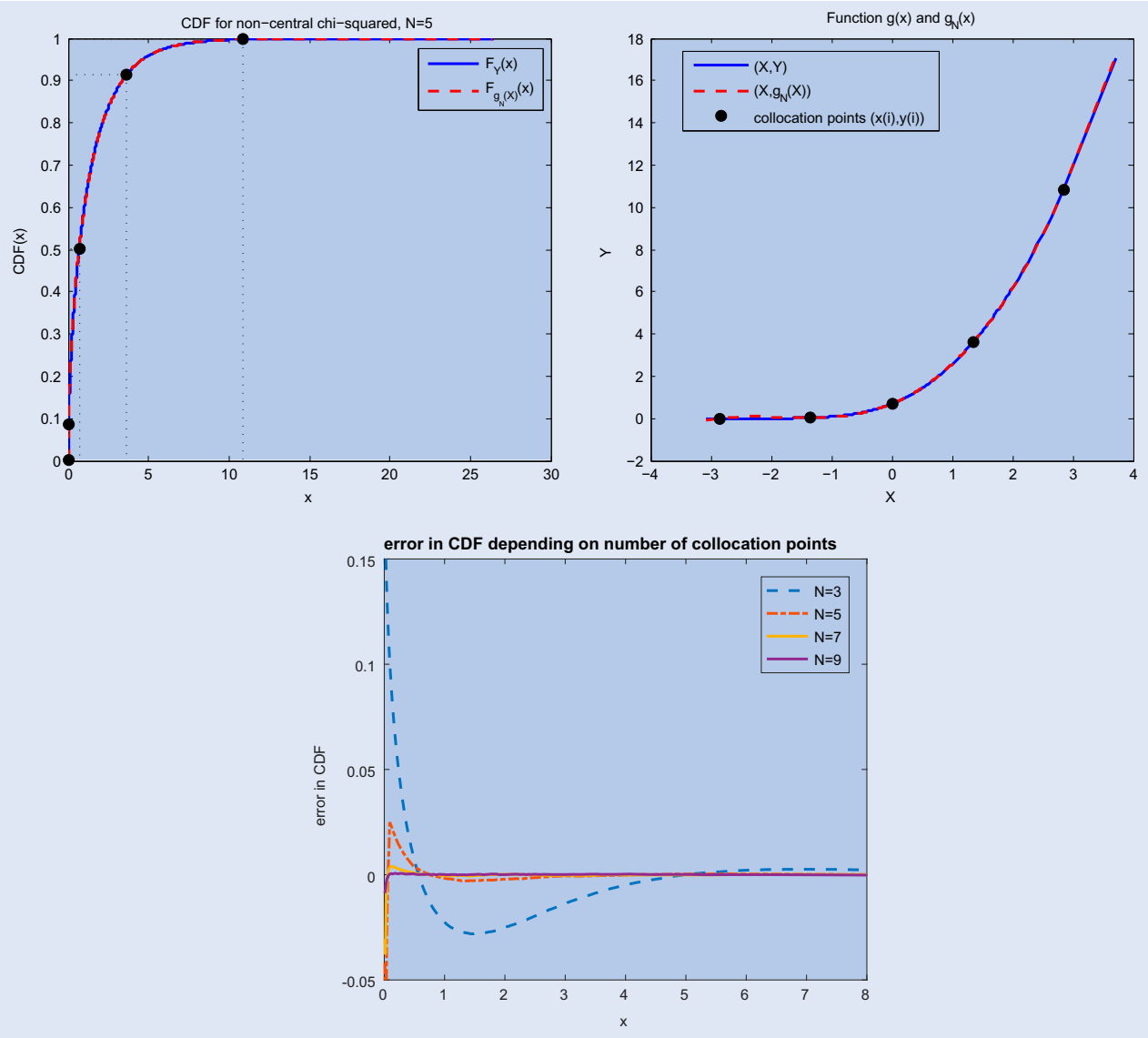

Figure 1. CDF and function $g(x)$ for the non-central chi-squared distribution obtained analytically vs. the collocation method for $N = 5$ collocation points. The third graph shows the error in CDF for a different number of collocation points.

Table 2. Timing results for sampling from $Y \sim \chi^2(d, \lambda)$ by evaluating $g_N(\xi_n)$ with $X \sim \mathcal{N}(0, 1)$ for different numbers of collocation points N and for $M = 1.000.000$ Monte Carlo samples. The results are reported in seconds.

No.Coll.	Coll. points $\boldsymbol{x}$	$F^{-1}_{\chi^2(d,\lambda)}(F_X(\boldsymbol{x}))$	$\xi_n \sim \mathcal{N}(0, 1)$	$g_N(\xi_n)$	Total $[s]$
N=4	0.0002	0.0067	0.0658	0.0261	0.10
N=5	0.0002	0.0074	0.0658	0.0305	0.10
N=6	0.0002	0.0077	0.0658	0.0351	0.11
N=7	0.0003	0.0083	0.0658	0.0394	0.11
N=8	0.0004	0.0089	0.0658	0.0436	0.12
N=9	0.0003	0.0122	0.0658	0.0479	0.13
N=10	0.0004	0.0102	0.0658	0.0521	0.13
MATLAB	—	—	—	—	0.40 $[s]$

$\mathbb{V}\text{ar}[g_2(X)] = \sigma_Y^2$, so that the distributions of Y and $g_2(X)$ are the same, $Y \overset{\text{d}}{=} g_2(X)$.

The essence is to project the expensive stochastic variable Y on a polynomial of X, so a collocation method of sufficiently high degree will yield exact results when a polynomial relation between the variables X and Y exists. Standard examples of such relations include the chi-squared distribution with one degree of freedom, $Y \sim \chi_1^2$, which has the same distribution as the squared standard normal distribution, X^2, the Rayleigh

distribution and the chi-squared distribution, the Maxwell distribution and the chi-squared distribution, the Gamma distribution and the chi-squared distribution, and many others.

To measure the error which results from the collocation method in a more general case, we can consider either the difference between the functions $g(X)$ and $g_N(X)$ or the error associated with the approximated cumulative distribution function. The first type of error is due to the Lagrange interpolation, so the corresponding error estimate is well known.

The relation between Y and X is $y = g(x)$, which is approximated by a Lagrange polynomial, $y \approx g_N(x)$, for N collocation points. The error $e_X(\xi_n)$ is given by the standard Lagrange interpolation error:

$$\begin{aligned} e_X(\xi_n) &= |g(\xi_n) - g_N(\xi_n)| \\ &= \left| \frac{1}{N!} \frac{\mathrm{d}^N g(x)}{\mathrm{d}x^N}\bigg|_{x=\hat{\xi}} \prod_{i=1}^{N} (\xi_n - x_i) \right|, \end{aligned} \tag{3.2}$$

with x_i a collocation point, $\hat{\xi} \in [x_1, x_N]$, and $\boldsymbol{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$, which can be bounded by defining $\hat{\xi}$ as the point where $|\mathrm{d}^N g(x)/\mathrm{d}x^N|$ has its maximum. Equation (3.2) gives the error $e_X(\xi_n)$ in y_n as a function of ξ_n. A small probability of large errors *in the tails* can be observed by deriving the error $e_U(u_n)$, substituting the uniformly distributed random variable u_n in (3.2), using $\xi_n = F_X^{-1}(u_n)$,

$$\begin{aligned} e_U(u_n) &= \left| g(F_X^{-1}(u_n)) - g_N(F_X^{-1}(u_n)) \right| \\ &= \left| \frac{1}{N!} \frac{\mathrm{d}^N g(x)}{\mathrm{d}x^N}\bigg|_{x=\hat{\xi}} \prod_{i=1}^{N} (F_X^{-1}(u_n) - x_i) \right|. \end{aligned} \tag{3.3}$$

The probability to be in the tails is small, as the inverse of the normal distribution, evaluated for uniform variables, determines the error. A 'close-to-linear' relation between the variables X and Y, or a polynomial of X and Y (meaning $\mathrm{d}^N g(x)/\mathrm{d}x^N$ being small) gives a small approximation error. To reduce the error, a variable X which is 'similar', in distributional sense, to variable Y is beneficial.

On the other hand, when approximating the CDF of Y, we have $F_Y(y) = F_Y(g(x)) \approx F_Y(g_N(x))$, which is exact at the collocation points x_i,

$$F_Y(y_i) = F_Y(g(x_i)) = F_Y(g_N(x_i)). \tag{3.4}$$

3.1. Error analysis and Gauss-Hermite quadrature

Using the fact that the collocation points x_i for $i = 1, \ldots, N$ are optimal and correspond to the zeros of orthogonal polynomials, we relate the collocation method to Gauss quadrature where the integral of a real-valued function $\Psi(x)$ can be approximated by a polynomial:

$$\int_{\mathbb{R}} \Psi(x) f_X(x) \mathrm{d}x = \sum_{i=1}^{N} \Psi(x_i)\omega_i + \epsilon_N, \tag{3.5}$$

with $f_X(x)$ the weight function, and ω_i the quadrature weights.

As explained in Golub and Welsch (1969), one can calculate pairs $\{x_i, \omega_i\}_{i=1}^N$ when the three-term recurrence relation is known for orthogonal polynomials generated by $f_X(x)$ and the moments of X are known. The weights ω_i are calculated as the first row of matrix $\widehat{J}$ in theorem 1.2.

The approximation $Y \approx Y_N \equiv g_N(X)$ generates an error which needs to be assessed. We have:

$$\begin{aligned} \mathbb{E}\left[(Y - Y_N)^2\right] &= \mathbb{E}\left[(g(X) - g_N(X))^2\right] \\ &= \int_{\mathbb{R}} (g(x) - g_N(x))^2 f_X(x) \mathrm{d}x, \end{aligned}$$

where $g(x) = F_Y^{-1}(F_X(x))$. The advantage of using optimal collocation points is that the method can be connected to the computation of integrals by quadrature. By theorem 1.2 the collocation points x_i and the corresponding weights w_i can be determined. Since $g(x_i) = g_N(x_i)$, for $i = 1, \ldots, N$, the error reads:

$$\begin{aligned} &\int_{\mathbb{R}} (g(x) - g_N(x))^2 f_X(x) \mathrm{d}x \\ &\quad = \sum_{i=1}^{N} (g(x_i) - g_N(x_i))^2 \omega_i + \epsilon_N = \epsilon_N. \end{aligned} \tag{3.6}$$

Thus, the error in L^2 is determined by the quadrature error.

For $X \sim \mathcal{N}(0, 1)$ a relation exists between the pairs $\{x_i, \omega_i\}_{i=1}^N$ and the Gauss-Hermite quadrature rule. The difference between the Gauss-Hermite quadrature pair $\{x_i^H, \omega_i^H\}_{i=1}^N$ and the Gauss quadrature pair $\{x_i, \omega_i\}_{i=1}^N$, with $X \sim \mathcal{N}(0, 1)$, is the weight function, i.e. we use the normal distribution while Gauss-Hermite quadrature is based on the function e^{-x^2}. For any function $\Psi(x)$ and $X \sim \mathcal{N}(0, 1)$ we have:

$$\begin{aligned} \mathbb{E}[\Psi(X)] &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \Psi(x) \mathrm{d}x \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} \mathrm{e}^{-x^2} \Psi(\sqrt{2}x) \mathrm{d}x \end{aligned} \tag{3.7}$$

So, the relation between the abscissas and weights of these rules is simply $x_i^H = x_i/\sqrt{2}$ and $\omega_i^H = \omega_i \sqrt{\pi}$.

From standard text books, like Abramowitz and Stegun (1972), the error for Gauss-Hermite quadrature, and thus for the collocation method, is given by:

$$\begin{aligned} \epsilon_N &= \frac{N!\sqrt{\pi}}{2^N} \frac{\Psi^{(2N)}(\hat{\xi}_1)}{(2N)!}, \quad \text{with} \quad \Psi(x) = (g(x) - g_N(x))^2 \\ &= \left(\frac{1}{N!} \frac{\mathrm{d}^N g(x)}{\mathrm{d}x^N}\bigg|_{x=\hat{\xi}_2(x)} \prod_{i=1}^{N} (x - x_i) \right)^2, \end{aligned} \tag{3.8}$$

as defined in (3.2) with $-\infty < \hat{\xi}_1 < \infty$ and $x_1 < \hat{\xi}_2(x) < x_N$.

In practical applications, when considering a bounded domain $[a, b]$ and $g(x)$ is $C(a, b)$, by the Weierstrass Theorem we have for any $\epsilon^* > 0$:

$$\left| \int_a^b (g(x) - g_N(x)) f_X(x) \mathrm{d}x \right| \leq \epsilon^*,$$

for N sufficiently large, which is also a classical result.

Remark $F_X(\cdot)$ is chosen to be the standard normal CDF in the present paper. This works very well, particularly for distributions of Y that are 'close to normal'. The choice to use the normal distribution for X is also motivated by the Cameron–Martin Theorem (Cameron and Martin 1947), which states that polynomial chaos approximations based on the normal distribution converge to any distribution.

Other distributions for X may, of course, be used but the inversion of these may be more expensive.

In the sections to follow, we will show that for 'non-normal' distributions that we encounter in finance, we can still keep the normal CDF for X, in combination however with certain enhancements, like grid stretching and the prescription of an atom at zero. A highly efficient and accurate approximation then results.

4. Stochastic collocation for distributions in finance

In this section, we discuss two enhancements of the basic stochastic collocation method in order to deal with specific features that occur in distributions in computational finance. In particular, we give some details on how to sample from distributions with an atom at zero and present a so-called grid-stretching technique which allows for more accurate handing of leptokurtic distributions.

We will also perform some experiments for some distributions occurring frequently in finance. We start with the Heston model (Heston 1993), and continue with the Stochastic Alpha Beta Rho (SABR Hagan *et al*. 2002) model.

4.1. Stretching of the collocation grid

The collocation method relies on a preferably 'close-to-linear' relation between the variables X and Y. Optimal results are obtained when the densities of X and Y resemble each other.

If the variables X and Y are related in a non-linear way, a large number of collocation points may be required. The collocation points x_i of X are solely determined based on the distribution and the moments of X, thus before the inversions $F_Y^{-1}(\cdot)$ take place.

When a symmetric random variable X is used to approximate an asymmetric variable Y, often a concentration of collocation points is encountered at the domain boundaries. Increasing the degree of the polynomial may not result in a significant increase in accuracy.

Two possible techniques to *accurately approximate* a variable Y with the help of X *before* the inversions $F_Y^{-1}(F_X(x_i))$ for $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)^{\mathrm{T}}$ are:

- *Moment matching:* To ensure that the collocation points determined from X are also based on Y one may choose the parameters of X so that $\mathbb{E}[X] = \mathbb{E}[Y]$, $\mathbb{E}[X^2] = \mathbb{E}[Y^2]$, etc. Unfortunately, the moments of the expensive variable Y are not always easily available or they do not exist.
- *Grid stretching:* Assuming that we have determined a set of collocation points, $\boldsymbol{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$, based on the moments of the variable X, instead of calculating $F_X(\boldsymbol{x})$, we consider another random variable $\hat{X}$ and calculate $F_{\hat{X}}(\boldsymbol{x})$ at the collocation points determined by X. Grid stretching provides us with a highly satisfactory *redistribution of points* at which the inverse $F_Y^{-1}(F_X(\boldsymbol{x}))$ is calculated.

4.1.1. Grid stretching for normal distribution. To illustrate the grid stretching technique we consider $X \sim \mathcal{N}(0, 1)$ and $N = 9$ collocation points. As the computation of the collocation points only requires the moments of X and the eigenvalue calculation of a certain matrix, this can be performed instantly. In table 3 these collocation values are tabulated, for convenience.

The first and last two values of $F_{\mathcal{N}(0,1)}(\cdot)$ accumulate at the boundaries of the interval [0, 1]. This pattern is even more pronounced for a larger number of collocation points N. We need to calculate $F_Y^{-1}(F_{\mathcal{N}(0,1)}(\cdot))$, and numerical instabilities for $F_{\mathcal{N}(0,1)}(\cdot) \to 0$ or $F_{\mathcal{N}(0,1)}(\cdot) \to 1$ may occur in the inversion procedure.

In such a case, we propose to define a new random variable $\hat{X}$ and evaluate its CDF at the collocation points x_i, i.e. $F_{\hat{X}}(x_i)$. A natural choice for $\hat{X}$ is another normally distributed random variable $\hat{X} \sim \mathcal{N}(0, \sigma^2)$ with standard deviation σ chosen such that the CDF of $\hat{X}$ at the first, x_1, or the last collocation point, x_N, is equal to some predefined quantile limits, $p_{\min}$ or $p_{\max}$, respectively, i.e.

$$F_{\mathcal{N}(0,\sigma^2)}(x_1) = p_{\min} \quad \text{or} \quad F_{\mathcal{N}(0,\sigma^2)}(x_N) = p_{\max}.$$

For any x, we have $F_{\mathcal{N}(0,\sigma^2)}(x) = F_{\mathcal{N}(0,1)}(x/\sigma)$, so that,

$$\sigma = \frac{x_1}{F^{-1}_{\mathcal{N}(0,1)}(p_{\min})} \quad \text{or} \quad \sigma = \frac{x_N}{F^{-1}_{\mathcal{N}(0,1)}(p_{\max})}. \tag{4.1}$$

To specify an appropriate value for $p_{\min}$ or $p_{\max}$ one needs some insight in the distribution of interest. If, for example, the distribution has heavy tails one needs to take $p_{\max}$ sufficiently large so that the tail is well approximated by the polynomial- in such a case $p_{\max} = 0.9995$ might be appropriate. In the case of moderate tails, a level of 0.995 should be sufficient to approximate the distribution. The specification of a quantile limit directly implies a minimum/maximum value for $F_X(x_1)$ and $F_X(x_N)$.

For the considered example, if we set $p_{\max} = 0.9995$, with (4.1) $\sigma = x_9/F^{-1}_{\mathcal{N}(0,1)}(0.9995) = 1.3714$. This σ-value leads to the values in table 3. As expected, the maximum value is found when the CDF is equal to the level determined by $p_{\max}$.

With an appropriate σ determined, the sampling equation is given by equation (2.3) with $y_i = F_Y^{-1}(F_{\mathcal{N}(0,\sigma^2)}(x_i))$, where the ξ_n are based on $\mathcal{N}(0, \sigma^2)$ and the collocation points x_i are determined by considering $\mathcal{N}(0, 1)$.

Due to obvious relations between normal CDFs, we can express (2.3) in this case as:

$$y_n \approx g_N(\xi_n) = \sum_{i=1}^{N} F_Y^{-1}\left(F_{\mathcal{N}(0,1)}\left(\boxed{x_i/\sigma}\right)\right)\ell_i(\xi_n),$$

$$\ell_i(\xi_n) = \prod_{j=1,\ i\neq j}^{N} \frac{\boxed{\sigma\xi_n} - x_j}{x_i - x_j},$$

where ξ_n represent samples from the standard normal distribution and x_i are also collocation points from the standard normal distribution.

Remark [When to apply grid stretching with a normal distribution?] To avoid numerically unstable inversions $F_Y^{-1}(a)$ with $a \to 0$ or $a \to 1$ the collocation grid should be stretched when a distribution is heavily tailed and the number of collocation points required is $N > 5$. Moreover, the grid stretching method should be used when there is clear evidence that the distribution Y is of *leptokurtic* type or the distribution is highly skewed.

Table 3. Collocation points and corresponding values of $F_X(x_i)$ and the stretched variant $F_{\mathcal{N}(0,\sigma^2)}(x_i)$ at the same collocation points.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_i	−4.5127	−3.2054	−2.0768	−1.0233	0.0	1.0233	2.0768	3.2054	4.5127
$F_{\mathcal{N}(0,1)}(x_i)$	0.0000	0.0007	0.0189	0.1531	0.5	0.8469	0.9811	0.9993	1.0000
$F_{\mathcal{N}(0,\sigma^2)}(x_i)$	0.0005	0.0097	0.0650	0.2278	0.5	0.7722	0.9350	0.9903	0.9995

Remark [Loss of optimality but not of stability] Gauss quadrature has the optimal polynomial degree of convergence for integration. The location of the Gauss quadrature points also leads to a stable interpolation. The grid stretching approach looses the theoretically optimal degree of convergence. On the other hand, it maintains the same interpolation stability properties, because the grid stretching leads to a linear displacement of quadrature nodes relative to the distribution.

4.2. Exact simulation of the Heston model

In financial applications we often need samples from distributions involving two, or more, dimensions, like in the *exact simulation* of stochastic volatility models, a particular example is the Heston stochastic volatility model (Heston 1993). Realisations of stock $S(t)$ depend on the variance realisation at time t *and* the integrated variance up to time t.

Although the exact simulation of the Heston model is well-known (Broadie and Kaya 2006), it is considered to be computationally expensive (see e.g. Andersen 2008). Variants to improve this exact simulation scheme were proposed in Hong Chan and Joshi (2013), Glasserman and Kim (2011).

Since the SCMC sampler is path independent, i.e. the number of inversions needed does not depend on the number of paths, we can significantly improve the conditional sampling, $S(T_2)|S(T_1)$ for $T_2 > T_1$.

Recall that the dynamics of the Heston stochastic volatility model under the risk-free measure $\mathbb{Q}$ are given by:

$$\begin{aligned} \mathrm{d}S(t)/S(t) &= r\mathrm{d}t + \sqrt{V(t)}\left[\rho \mathrm{d}W_V(t) + \sqrt{1-\rho^2}\mathrm{d}W_x(t)\right], \\ S(t_0) &= S_0, \\ \mathrm{d}V(t) &= \kappa(\bar{V} - V(t))\mathrm{d}t + \gamma\sqrt{V(t)}\mathrm{d}W_V(t), \\ V(t_0) &= V_0, \end{aligned} \tag{4.2}$$

with stock $S(t)$, variance process $V(t)$, correlation ρ, speed of mean reversion κ, volatility of variance γ, mean variance $\bar{V}$ and independent Brownian motions, $\mathrm{d}W_S(t)\mathrm{d}W_V(t) = 0\mathrm{d}t$. In log-space, $\log S(t)$, the model belongs to the class of affine processes with the characteristic function (ChF) given by:

$$\begin{aligned} \Phi_{\log S(T)}(u) = \exp\Bigg[& iu\log S_0 + iurT + \frac{V_0}{\gamma^2}\left(\frac{1-\mathrm{e}^{-DT}}{1-G\mathrm{e}^{-DT}}\right) \\ & \cdot (\kappa - i\rho\gamma u - D)\Bigg] \\ & \cdot \exp\left[\frac{\kappa\bar{V}}{\gamma^2}\left((\kappa - i\rho\gamma u - D)T\right.\right. \\ & \left.\left. -2\log\left(\frac{1-G\mathrm{e}^{-DT}}{1-G}\right)\right)\right], \end{aligned}$$

with

$$\begin{aligned} D &= \sqrt{(\kappa - i\rho\gamma u)^2 + (u^2 + iu)\gamma^2}, \quad \text{and} \\ G &= \frac{\kappa - i\rho\gamma u - D}{\kappa - iu\gamma u + D}. \end{aligned}$$

When the unconditional characteristic function is available we can obtain the corresponding CDF using e.g. the COS method (Fang and Oosterlee 2009).

From samples of $\log S(T)$ we get realisations of $S(T)$ by simply taking the exponent. Cheap random variable X and corresponding random sample ξ_n are standard normally distributed variables. The samples, $y_n = \log(S(T))_n$, can then be expressed as in equation (2.3) with $y_i = F^{-1}_{\log S(T)}(F_{\mathcal{N}(0,1)}(x_i))$.

The unconditional sampling from the Heston model is straightforward, as it requires only a few inversions of the original CDF. In practice, however, when pricing path-dependent options it is crucial to sample *conditionally* on previous realisations of the process, i.e. $\log S(T_{i+1})|\log S(T_i)$. This type of sampling can be performed by the application of the SCMC sampler in combination with the exact simulation by Broadie–Kaya (Broadie and Kaya 2006). Application of the collocation technique may significantly improve the speed of generating samples of the stock price process.

Let us consider two time points $T_1 < T_2$ and $t_0 \le T_1$, so that under the Heston dynamics in (4.2), the solution of the log-stock, $\log S(t)$, is given by:

$$\begin{aligned} \log\left(\frac{S(T_2)}{S(T_1)}\right) = & \int_{T_1}^{T_2}\left(r - \frac{1}{2}V(s)\right)\mathrm{d}s \\ & + \rho\int_{T_1}^{T_2}\sqrt{V(s)}\mathrm{d}W_V(s) \\ & + \sqrt{1-\rho^2}\int_{T_1}^{T_2}\sqrt{V(s)}\mathrm{d}W_x(s). \end{aligned} \tag{4.3}$$

Integrating the variance process in (4.2) gives,

$$\begin{aligned} \int_{T_1}^{T_2}\sqrt{V(t)}\mathrm{d}W_V(s) = \frac{1}{\gamma}\Bigg[& V(T_2) - V(T_1) \\ & + \kappa\int_{T_1}^{T_2}V(s)\mathrm{d}s - \kappa\bar{V}(T_2 - T_1)\Bigg], \end{aligned}$$

so that the log-stock process reads (for details, see Broadie and Kaya 2006):

$$\log S(T_2) = \log S(T_1) + \mu_S(T_1, T_2) + \sigma_S(T_1, T_2)\xi, \tag{4.4}$$

with ξ a standard normal, and

$$\begin{aligned} \mu_S(T_1, T_2) &= r(T_2 - T_1) + \left[\frac{\kappa\rho}{\gamma} - \frac{1}{2}\right]Y(T_1, T_2) \\ &\quad + \frac{\rho}{\gamma}\left[V(T_2) - V(T_1) - \kappa\bar{V}(T_2 - T_1)\right], \\ \sigma_S^2(T_1, T_2) &= (1-\rho^2)Y(T_1, T_2), \end{aligned} \tag{4.5}$$

and $Y(T_1,T_2)=\int_{T_1}^{T_2}V(s)\mathrm{d}s$.

The expressions above show how to obtain samples for $S(T_2)$ given realisations of $V(T_2)$, $V(T_1)$, $S(T_1)$ and the integrated variance $Y(T_1,T_2)$. Sampling from $Y(T_1,T_2)$ is involved as it depends on realisations of stochastic variable $V(T_2)$ and initial realisations of $V(T_1)$.

Given realisations of $V(T_1)$ and conditional samples of $V(T_2)$ a result by Broadie and Kaya (2006) can be used with a closed-form expression of the characteristic function of $Y:=Y(T_1,T_2)$†:

$$\begin{aligned}&\Phi_{Y|V(T_1)=v,V(T_2)=w}(u)\\&:=\mathbb{E}\left[\mathrm{e}^{iuY(T_1,T_2)}\middle|V(T_1)=v,V(T_2)=w\right]\\&=\frac{\psi(u)\mathrm{e}^{-0.5(\psi(u)-\kappa)\tau}(1-\mathrm{e}^{-\kappa\tau})}{\kappa(1-\mathrm{e}^{-\psi(u)\tau})}\\&\quad\cdot\exp\left[\frac{v+w}{\gamma^2}\left(\frac{\kappa(1+\mathrm{e}^{-\kappa\tau})}{1-\mathrm{e}^{-\kappa\tau}}-\frac{\psi(u)(1+\mathrm{e}^{-\psi(u)\tau})}{1-\mathrm{e}^{-\psi(u)\tau}}\right)\right]\\&\quad\cdot\frac{I_b\left(\sqrt{vw}\frac{4}{\gamma^2}\frac{\psi(u)\mathrm{e}^{-0.5\psi(u)\tau}}{1-\mathrm{e}^{\psi(u)\tau}}\right)}{I_b\left(\sqrt{vw}\frac{4\kappa}{\gamma^2}\frac{\mathrm{e}^{-0.5\kappa\tau}}{1-\mathrm{e}^{\kappa\tau}}\right)},\end{aligned}\tag{4.6}\tag{4.7}$$

with $\tau=T_2-T_1$, $b=2\kappa\bar{V}/\gamma^2-1$, $\psi(u)=\sqrt{\kappa^2-2iu\gamma^2}$, and $I_b(\cdot)$ the modified Bessel function of the first kind. As noted in Lord and Kahl (2010), equation (4.7) involves a complex-valued argument for $I_b(\cdot)$ which, if not treated carefully, may give rise to discontinuities of the characteristic function, $\Phi_{Y|V_1=v,V_2=w}(u)$. To guarantee its continuity one needs to evaluate the following, algebraically equivalent, representation:

$$\Phi_{Y|V_1=b,V_2=w}(u)\frac{\exp\left(b\log q(u)\right)}{q^b(u)},\quad\text{with}$$
$$q(u)=\frac{\psi(u)\mathrm{e}^{-0.5\psi(u)\tau}}{1-\mathrm{e}^{\psi(u)\tau}}.\tag{4.8}$$

With the CDF for $Y|V_1=v,V_2=w$ determined, we can obtain samples by inverting the CDF, i.e. for $u_n\sim\mathcal{U}([0,1])$, $y_n=F^{-1}_{Y|v,w}(u_n)$. When this inversion procedure would take place for each sample u_n this technique would be inefficient.

By using the SCMC sampler, however, we can reduce the number of inversions significantly by specifying a collocation variable X for which the inversion is *cheap*.

Let us assume that, with the results of the previous section, we have obtained samples for the non-central chi-squared random variables $V(T_1)$ and‡ $V(T_2)$. We focus on efficient sampling from $Y(T_1,T_2)$, given the samples of $V(T_1)$ and $V(T_2)$. Since the realisations y_n from $Y(T_1,T_2)$ are dependent on the realisations v_n from $V(T_1)$ and w_n from $V(T_2)$, we write $y_n|v_n,w_n=:y_n(v_n,w_n)$. Using the definition of the 2D Lagrange polynomial (Micchelli 1980, Sauer and Xu 1995) we find:

$$y_n|v_n,w_n\approx y_n(v_n,w_n)=\sum_{j=1}^{N_{V_1}}\sum_{k=1}^{N_{V_2}}y_n(v_j,w_k)\ell_j(v_n)\ell_k(w_n).\tag{4.9}$$

For a pair (j,k), we use for the calculation of $y_n(v_j,w_k)$:

$$y_n(v_j,w_k)=\sum_{i=1}^{N_Y}F^{-1}_{Y|V_1=v_j,V_2=w_k}(F_X(x_i))\ell_i(\xi_n),\tag{4.10}$$

so that

$$\begin{aligned}&y_n|v_n,w_n\\&\approx\sum_{i=1}^{N_Y}\sum_{j=1}^{N_{V_1}}\sum_{k=1}^{N_{V_2}}F^{-1}_{Y|V_1=v_j,V_2=w_k}(F_X(x_i))\ell_j(v_n)\ell_k(w_n)\ell_i(\xi_n),\end{aligned}\tag{4.11}$$

with

$$\begin{aligned}\ell_i(\xi_n)&=\prod_{i=1,\ j\neq i}^{N_Y}\frac{\xi_n-x_k}{x_j-x_k},\quad \ell_j(v_n)=\prod_{k=1,\ \neq k}^{N_{V_1}}\frac{v_n-v_k}{v_j-v_k},\\\ell_i(w_n)&=\prod_{k=1,\ 1\neq k}^{N_{V_2}}\frac{w_n-w_k}{w_i-w_k},\end{aligned}\tag{4.12}$$

where v_n, w_n and ξ_n are samples of $V(T_1)$, $V(T_2)$ and X, respectively, and v_j,w_k,x_i are collocation points of $V(T_1)$, $V(T_2)$ and X.

With the collocation points $v_1,\ldots,v_{N_{V_1}},w_1,\ldots,w_{N_{V_2}}$ and $x_1,\ldots,x_{N_Y}$ determined and by $N_{V_1}\cdot N_{V_2}\cdot N_Y$ inversions of $F^{-1}_{Y|V(T_1)=v_j,V(T_2)=v_k}(F_X(x_i))$, we can generate vectors of samples for variables X, $V(T_1)$ and $V(T_2)$ and evaluate polynomial (4.11) to obtain the samples $Y|V_1,V_2$.

In order to facilitate efficient evaluation of the Lagrange polynomial in 3D it is recommended to use the barycentric polynomial representation given by§:

$$y_n|v_n,w_n\approx\sum_{i,j,k}y_{i,j,k}\frac{\lambda_i\lambda_j\lambda_k\ell(\xi_n)\ell(v_n)\ell(w_n)}{(\xi_n-x_i)(v_n-v_j)(w_n-w_k)},$$

where for $z\in\{x,v,w\}$ we have $\ell(z_n)=\prod_{i=1}^{N_z}(z_n-z_i)$ and for $l\in\{i,j,k\}$ we have $\lambda_l=1/\prod_{i=1,i\neq l}^{N_z}(z_l-z_i)$.

In the next subsection an example is presented.

4.2.1. Integrated variance experiment. We present an example of how the collocation technique can be used for efficient calculation of samples from the integrated variance, $Y(T_1,T_2)|v_n,w_n$, given the samples of the variance process at times T_1 and T_2, i.e. $Y(T_1,T_2)$. We show how to efficiently obtain M samples from the integrated variance given the marginal variables $V(T_1)$ and $V(T_2)$.

In the experiment $T_1=5y$ and $T_2=10y$ are used with the following set of parameters:

$$\gamma=0.2,\quad\kappa=0.5,\quad\bar{V}=V_0=0.1.\tag{4.13}$$

The first ingredient for obtaining samples of $Y(T_1,T_2)$ is to generate samples, v_n, from $V(T_1)$. This is just sampling from the non-central chi-squared random variable, as discussed in

†Where a conditional expectation w.r.t. an event with null probability is defined as $\mathbb{E}[X|Y=y]=\int_{\mathbb{R}}xf_{X|Y}(x|y)\mathrm{d}x$ with X and Y being continuous random variables and where $f_{X|Y}(x|y)$ is the conditional density.

‡Note that by $V(T_2)$ we actually mean $V(T_2)|V(T_1)$

§$\ell(\xi_n)\ell(v_n)\ell(w_n)$ can be taken outside the sum.

Table 4. Inversions $y_{i,j,k} = F^{-1}_{Y|v_j,w_k}(F_X(x_i))$ for all collocation points x_i, v_j, w_k.

v_j	w_k	x_1	x_2	x_3	x_4	x_5
v_1	w_1	0.1295	0.2106	0.3383	0.5391	0.8560
v_1	w_2	0.2040	0.3240	0.5023	0.7615	1.1450
v_1	w_3	0.3619	0.5481	0.7965	1.1267	1.5875
v_2	w_1	0.2387	0.3733	0.5667	0.8403	1.2393
v_2	w_2	0.3362	0.5210	0.7730	1.1081	1.5748
v_2	w_3	0.5347	0.7974	1.1214	1.5264	2.0692

Table 5. Timing results for evaluating the polynomial in (4.11) depending on the number of samples of v_n, w_n and ξ_n. The experiment was performed on a standard PC, i5-2400 with 4 GB of RAM.

#paths	10	100	1.000	10.000	100.000	500.000	1.000.000
Time [s]	0.00010	0.00012	0.00082	0.0075	0.023	0.11	0.21

section 2.1.1. With the samples v_n we need to determine $w_n|v_n$. For this purpose we apply the SCMC sampler given by:

$$w_n|v_n \approx g_{N_{V_2},N_{V_1}}(\xi_n) = \sum_{i=1}^{N_{V_2}}\sum_{j=1}^{N_{V_1}} F^{-1}_{V(T_2)|V(T_1)=v_j}(F_X(x_i))\ell_i(\xi_n)\ell_j(v_n), \tag{4.14}$$

where ξ_n are the independent samples from the normal distribution, v_n are samples from the non-central chi-squared $V(T_1)$, and x_i and v_i are collocation points for approximating variable X and non-central chi-squared variable $V(T_1)$, respectively. These points can be calculated using equation (A9) and the eigenvalues of the corresponding matrix (see theorem 1.2). In order to obtain any number of samples from $w_n|v_n$ we need $N_{V_1} \cdot N_{V_2}$ inversions $F^{-1}_{V(T_2)|V(T_1)=v_j}(F_X(x_i))$, for $i = 1, \ldots, N_{V_2}$ and $j = 1, \ldots, N_{V_1}$.

Then, we sample from $Y|v_n, w_n$, with the following numbers of collocation points, $N_Y = 5$, $N_{V_1} = 2$, $N_{V_2} = 3$, which implies 30 inversions in (4.11).

Under the Heston model both $V(T_1)$ and $V(T_2)|V(T_1)$ are non-centrally chi-squared distributed: $V(T_1) \sim c(t)\chi^2_{d,\lambda(t,V_0)}$, $V(T_2)|V(T_1) \sim c(t)\chi^2_{d,\lambda(T_2-T_1,V(T_1))}$, with d representing the degrees of freedom and with non-centrality parameter $\lambda(t,\cdot)$, with values:

$$c(t) = \frac{\gamma^2}{4\kappa}\left(1-\mathrm{e}^{-\kappa t}\right),\ d = \frac{4\kappa\bar{V}}{\gamma^2},$$
$$\lambda(t,V_0) = \frac{4\kappa V_0 \mathrm{e}^{-\kappa t}}{\gamma^2(1-\mathrm{e}^{-\kappa t})}. \tag{4.15}$$

With the parameter specifications in (4.13) we find:

$$V(T_1) \sim 0.0184\chi^2_{5,0.4471},$$
$$V(T_2)|V(T_1) \sim 0.0199\chi^2_{5,\lambda(5,V(T_1))}. \tag{4.16}$$

The collocation points are determined by the moments of the marginal distributions of $V(T_1)$ and $V(T_2)$. The moments for the non-central chi-squared distribution are known. For $\nu \sim \chi^2_{d,\lambda}$ we have:

Table 6. Heston model parameters sets (Andersen 2008). For all tests, additionally, $S(T) = 100$, $r = 0$ and $K = (50, 75, 100, 125, 150, 200)$.

Heston	γ	κ	ρ	T_2	$\bar{V} = V_0$
Set I	1	0.5	−0.9	10	0.04
Set II	0.9	0.3	−0.5	15	0.04
Set III	1	1	−0.3	5	0.09

$$\mathbb{E}[v^n] = 2^{n-1}(n-1)!(d+n\lambda) + \sum_{i=1}^{n-1}\frac{(n-1)!2^{i-1}}{(n-i)}(d+i\lambda)\mathbb{E}[v^{n-1}]. \tag{4.17}$$

Application of equation (A9) and computation of the eigenvalues, according to theorem 1.2, of the corresponding matrix, give the following sets of collocation points for $V(T_1)$ and $V(T_2)$: $\boldsymbol{v} = (0.0651, 0.2139)^{\mathrm{T}}$ and $\boldsymbol{w} = (0.0488, 0.1524, 0.3388)^{\mathrm{T}}$. In the experiment, the integrated variance Y will be approximated by standard normal X, with collocation points $\boldsymbol{x} = (-2.8570, -1.3556, 0.0, 1.3556, 2.8570)^{\mathrm{T}}$. These collocation points indicate the locations at which the *expensive* inversions from (4.11) need to be calculated. The inversions are presented in table 4, for convenience.

The M samples from $y_n|v_n, w_j$ are determined by evaluating the polynomial in (4.11). The polynomial evaluation can be vectorised and, as presented in table 5, with vectors of samples v_n, w_n and ξ_n for $V(T_1)$, $V(T_2)$ and X, the polynomial evaluation is very fast, e.g. for 100.000 paths the time needed is about 0.02 s. In figure 2 the simulation results are presented. By using only 30 inversions, we generate samples that are very close to those obtained by direct inversion.

4.2.2. Heston Model exact sampling experiment. We use the SCMC sampler for exact sampling for the Heston model, and specify three sets of Heston parameters in table 6, as in Andersen (2008). These model parameters correspond to test cases with leptokurtic densities. All parameter sets in table 6 do not satisfy the so-called Feller condition indicating a high degree of asymmetry (skewness).

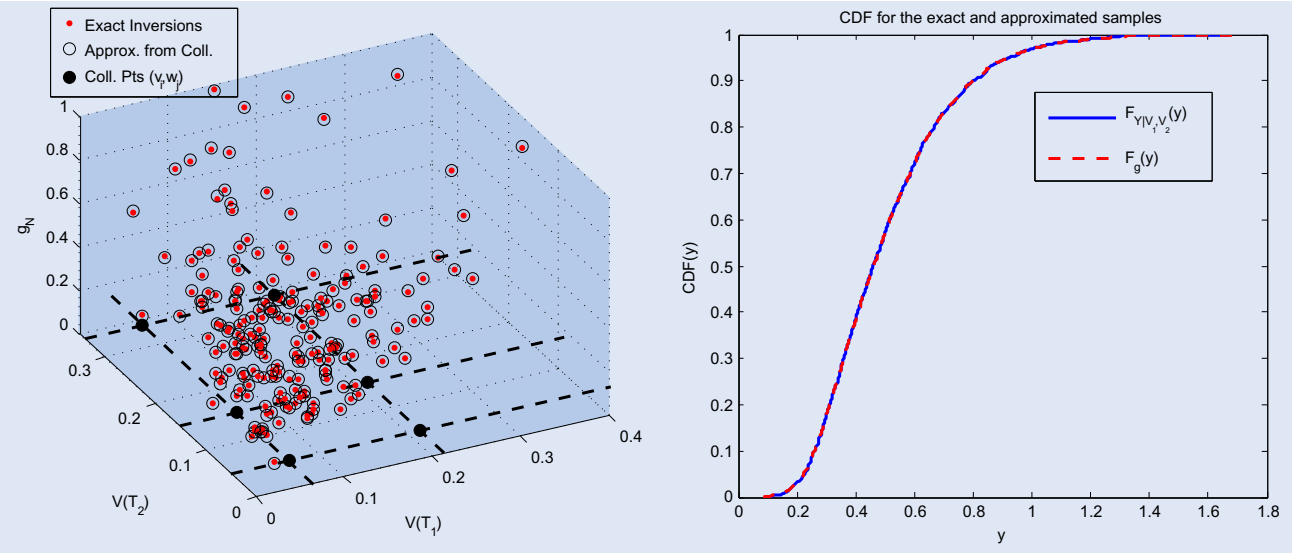

Figure 2. Left: First 200 exact and approximated samples from the conditional integrated variance $Y|V(T_1), V(T_2)$. The projected collocation points on the space $(V(T_1), V(T_2))$ where $V(T_1) = \{v_1, v_2\}$ and $V(T_2) = \{w_1, w_2, w_3\}$ as tabulated in table 4 are indicated by black dots. Right: Empirical CDFs of the integrated variance obtained by the exact simulation and the SCMC sampler.

In all experiments a time grid is defined consisting of $\tau = \{T_1, T_2\}$ with $T_1 = T_2/2$ (T_2 as in table 6). Intermediate time step $T_1 = T_2/2$ is of high interest here as the sampling for stock $S(T_2)$ requires sampling from $Y(t_0, T_1)|V(t_0), V(T_1)$ *and* $Y(T_1, T_2)|V(T_1), V(T_2)$. Sampling for the first time interval is much easier than for any other interval, due to the fact that the initial variance $V(t_0)$ is *known* and sampling from the integrated variance then requires only a two-dimensional collocation grid. In any other case a three-dimensional collocation grid, as in (4.11), is needed.

Once the samples from the variances $V(T_1)$, $V(T_2)$ and the integrated variances: $Y(t_0, T_1)$, $Y(T_1, T_2)$ are calculated by the SCMC sampler, the samples for the stock $S(T_2)$ are obtained via equation (4.4).

For all numerical experiments we specify $N_{V_1} = 7, N_{V_2} = 7$ and vary the number of collocation points, N_Y, for the integrated variance Y. We generate 1.000.000 Monte Carlo paths.

As explained in section 4.1.1, when we use a normal distribution for the collocation it is recommended to use the *stretched collocation grid*, especially in the case of a large number of collocation points.

Table 7 shows the results for pricing European options. The table shows the impact of different numbers of collocation points N_Y on the implied volatility error. For the given extreme sets of parameters the SCMC sampler performs excellently. In the worst case presented, it is necessary to perform $7 \cdot 7 \cdot 8$ inversions of the original distribution to obtain an accuracy at the level 0.1%. Based on these inversions we are able to generate any number of Monte Carlo paths. In figure 3 the final implied volatility results for the given sets of parameters are depicted.

4.3. *Distributions with mass at zero, virtual collocation points*

In this section we consider the problem of efficiently sampling from distributions that have a significant probability mass at 0. Such distributions are typically associated with stochastic processes with so-called *absorbing boundary conditions*. The basic example in finance is the Bessel process associated with Constant Elasticity of Variance (CEV) dynamics, given by:

$$\mathrm{d}S(t) = \sigma S^{\beta}(t)\mathrm{d}W(t), \tag{4.18}$$

with some initial condition $S_0 > 0$. The invertible transformation, $Z_1(t) = S^{1-\beta}(t)/(1-\beta)$, for $\beta \neq 1$, and Itô's lemma, give the time-changed Bessel process:

$$\mathrm{d}Z_1(t) = -\frac{\beta\sigma^2}{2(1-\beta)Z_1(t)}\mathrm{d}t + \sigma\mathrm{d}W(t). \tag{4.19}$$

The process $Z_2(t) = Z_1^2(t)$ is a time-changed squared Bessel process of dimension $\delta := (1-2\beta)/(1-\beta)$, which satisfies the following SDE:

$$\mathrm{d}Z_2(t) = \delta\sigma^2\mathrm{d}t + 2\sigma\sqrt{|Z_2(t)|}\mathrm{d}W(t), \tag{4.20}$$

and with $\nu(t) = \sigma^2 t$, we find $Z_2(t) \equiv Z_{\nu(t)}$ where $Z(t)$ is a δ-dimensional squared Bessel process,

$$\mathrm{d}Z(t) = \delta\mathrm{d}t + 2\sqrt{|Z(t)|}\mathrm{d}W(t), \tag{4.21}$$

with degrees of freedom δ. With $\delta \leq 0$, $Z(t)$ has a boundary condition at zero which is absorbing. One can show that for $\frac{1}{2} \leq \beta < 1$ a unique solution of SDE (4.18) exists, and that the boundary value at zero is absorbing. Moreover, the probability density function *does not integrate to unity* for $t > 0$.

For $\frac{1}{2} \leq \beta < 1$ the CDF of the CEV process in (4.18) is given by (Schroder 1989):

$$\mathbb{P}[S(T) \leq y|S_0] = 1 - F_{\chi^2(b,c(y))}(a), \tag{4.22}$$

with

$$a = \frac{S_0^{2(1-\beta)}}{(1-\beta)^2\sigma^2 T}, \quad b = \frac{1}{1-\beta}, \quad c(y) = \frac{y^{2(1-\beta)}}{(1-\beta)^2\sigma^2 T}, \tag{4.23}$$

where $F_{\chi^2(b,c(y))}(a)$ is the CDF of the non-central chi-squared distribution with b degrees of freedom and non-centrality parameter $c(y)$. Probability mass at zero is found for the follow-

Table 7. Implied volatilities and errors calculated for different numbers of collocation points N_Y with fixed $N_{V_1} = N_{V_2} = 7$ calculated for different strike levels, K, in the evaluation of European option prices. 'Fourier IV' stands for the reference implied volatilities obtained by a Fourier technique (Fang and Oosterlee 2009). Superscript * indicates the cases in which the stretched grid technique with $q = 0.995$ was used. Parameters sets I, II and III are presented in table 6. 'Collocation errors' represent the difference of the implied volatilities between the reference, Fourier implied volatility, and volatilities obtained from the collocation method. In the performed experiments the standard error was < 0.0001.

Implied Volatilities and Errors [%]								
	Strike [K]	50	75	100	125	150	175	200
	Fourier IV [%]	20.21	14.98	10.42	6.54	5.83	6.15	6.53
	# Coll. points				Collocation Errors [%]			
Set I	$N_Y = 3$	2.09	1.00	−0.24	−1.38	−0.99	−0.68	−0.58
	$N_Y = 4$	0.42	−0.18	−0.53	−0.82	−0.40	−0.25	−0.28
	$N_Y = 5$	0.55	0.45	0.40	0.06	0.06	0.00	−0.04
	$N_Y = 6^*$	0.21	0.15	0.11	0.14	0.06	0.00	−0.04
	$N_Y = 7^*$	0.18	0.10	0.07	0.13	0.07	0.00	−0.05
	$N_Y = 8^*$	0.15	0.06	0.05	0.10	0.06	0.00	−0.05
	Strike [K]	50	75	100	125	150	175	200
	Fourier IV [%]	17.65	13.64	10.85	9.99	10.55	11.35	12.11
	# Coll. points				Collocation Errors [%]			
Set II	$N_Y = 3$	0.04	−1.32	−2.59	−2.91	−2.45	−2.01	−1.68
	$N_Y = 4$	−0.43	−1.11	−1.96	−2.12	−1.67	−1.32	−1.12
	$N_Y = 5$	0.27	0.24	−0.15	−0.30	−0.08	0.01	−0.01
	$N_Y = 6^*$	0.05	0.05	0.12	0.13	0.01	−0.04	−0.09
	$N_Y = 7^*$	0.03	0.05	0.16	0.17	0.04	−0.04	−0.10
	$N_Y = 8^*$	0.01	0.03	0.13	0.13	0.02	−0.067	−0.11
	Strike [K]	50	75	100	125	150	175	200
	Fourier IV [%]	30.84	26.92	24.74	23.94	24.02	24.50	25.12
Set III	# Coll. points				Collocation Errors [%]			
	$N_Y = 3$	−0.17	−0.29	0.04	0.24	−0.01	−0.20	−0.28
	$N_Y = 4$	0.07	0.06	0.04	0.04	0.06	0.06	0.06

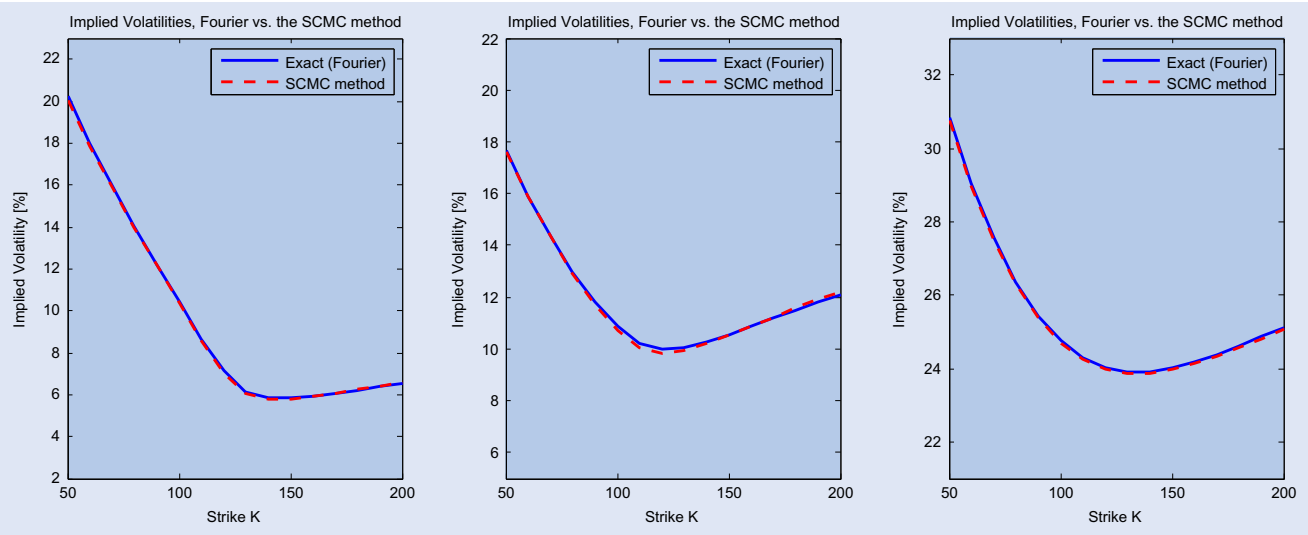

Figure 3. Implied volatilities for Heston Sets I, II and III from table 6, evaluated at T_2. The number of collocation points used corresponds to the last N_Y in table 7. Simulation was performed with one intermediate time, i.e. $\tau = \{T_1, T_2\}$.

ing (example) set of parameters, $T = 2$, $S_0 = 0.07$, $\beta = 0.5$ and $\sigma = 0.4$, for which we obtain $\mathbb{P}[S(T = 2) = 0] = 0.6456$.

We wish to approximate the (expensive) distribution for $S(t)$, so, referring to the general description $Y(t) := S(t)$ in this case. Following the strategy from the previous sections, for some predetermined collocation points $\boldsymbol{x}$ of approximating variable X with CDF $F_X(\boldsymbol{x})$, we may encounter cases for which $F_X(x_i) \leq F_{S(T)}(0)$. As $S(T)$ has an atom at zero there is *no bijective mapping* and the inversion $y_i = F^{-1}_{S(T)}(F_X(x_i))$ is not well-defined. One way to deal with this is to choose a stochastic variable X which also has an atom at 0. However, this would not be beneficial as X would very likely need to follow an expensive distribution as well.

Result 4.1 *For any nonnegative random variable θ such that $\mathbb{P}[\theta = 0] = p$ and $\mathbb{P}[\theta > 0] = 1-p$, a random variable $\zeta \in \mathbb{R}$*

exists with $\mathbb{P}[\zeta < 0] = p$, *so that* $\mathbb{P}[\theta = 0] = \mathbb{P}[\max(\zeta, 0) = 0] = p$.

The proposed technique here is the usage of so-called *virtual collocation points*. The main idea is to define a mixture of CDFs for $S(T)$, i.e.

$$\hat{F}_{S(T)}(y) = \begin{cases} F_{S(T)}(y), & \text{for } y > 0, \\ f(y), & \text{for } y < 0, \end{cases} \tag{4.24}$$

where $f(y)$ is a monotonically increasing function of y such that $f(-\infty) = 0$ and $f(0) = F_Y(0+\epsilon)$, $\epsilon \to 0$. The function $f(y)$ can be seen as a (*virtual*) monotonic extrapolation of the original CDF, $F_{S(T)}(y)$ for $y < 0$.

In principle, $f(y)$ in (4.24) can be chosen arbitrarily. Here, a linear extrapolation of the CDF is used for $f(y)$. This is to be preferred over a linear extrapolation of the polynomial $g_N(x)$.

With the new, continuous, CDF $\hat{F}_{S(T)}(y)$ defined, we relate its inversions to the original distribution $F_{S(T)}(y)$. The samples from $S(T)$ can be obtained from:

$$\begin{aligned} s_n &= \max(\hat{F}^{-1}_{S(T)}(F_X(\xi_n)), 0) \\ &= \begin{cases} F^{-1}_{S(T)}(F_X(\xi_n)), & \text{for } F_X(\xi_n) > F_{S(T)}(0), \\ 0, & \text{for } F_X(\xi_n) < F_{S(T)}(0). \end{cases} \end{aligned} \tag{4.25}$$

The samples s_n for $S(T)$ can be generated, employing the SCMC sampler, via

$$s_n \approx g_N(\xi_n) = \max\left(\sum_{i=1}^{N} \hat{F}^{-1}_{S(T)}(F_X(x_i))\ell_i(\xi_n), 0 \right). \tag{4.26}$$

In order to determine n samples $s_{2,n}|s_{1,n}$ from $S(T_2)|S(T_1)$ for times $T_2 > T_1$ one needs the 2D variant of the SCMC sampler where the sampling is performed via:

$$\begin{aligned} &s_{2,n}|s_{1,n} \\ &\approx \max\left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} F^{-1}_{S(T_2)|S(T_1)=s_{1,j}}(F_X(x_i))\ell(\xi_n)\ell(s_{1,n}), 0 \right), \end{aligned}$$

with x_i and $s_{1,j}$ the collocation points from X and $S(T_1)$, respectively, ξ_n the samples from X, $s_{1,n}$ the samples from $S(T_1)$ and where $s_{2,n}|s_{1,n}$ indicates samples from $S(T_2)$ given realisations from $S(T_1)$. The collocation points of $S(T_1)$ can be calculated from the moments using (4.22).

4.3.1. Sampling from the CEV process. We consider the following set of CEV parameters, $T = 2$, $S_0 = 0.07$, $\beta = 0.5$ and $\sigma = 0.4$. Equation (4.22) indicates that the CEV model in (4.18) has a mass at zero, i.e.

$$\mathbb{P}[S(2) \leq 0 | S_0 = 0.07] = 1 - F_{\chi^2(2,0)}(0.8750) = 0.6456. \tag{4.27}$$

We take $N = 5$ collocation points and $X \sim \mathcal{N}(0, 1)$. In table 8 we tabulate the collocation points and corresponding inversions, for convenience. The boxed points x_i for which $F_X(x_i) < \mathbb{P}[S(T) = 0] = 0.6456$ are the virtual nodes. For this set of collocation points $f(x)$ is the linear extrapolation of $F_{S(T)}(x)$ for $x < 0$. Our experiments show that linear extrapolation of the CDF provides a sufficiently *smooth* transition towards the virtual part of the CDF. The linear extrapolation requires two additional function evaluations to determine its value and its gradient at $F_Y(0 + \epsilon)$ using finite differences.

The analytic and approximate CDFs of $S(T)$ are presented in figure 4. The left-hand figure shows the approximating CDF (red line) before the samples are capped at 0 and the right-hand figure depicts the resulting approximate CDF. The experiment is performed with $M = 1.000.000$ samples from the standard normal variable. As the computation of virtual points is extremely cheap (i.e. extrapolation in only three points) the timing results are similar to those obtained for sampling from the non-central chi-squared distribution in section 2.1.1.

Remark [Grid stretching for the squared Bessel process] The quality of the results may be further improved by increasing the number of collocation points and by applying grid stretching as in section 4.1.1.

In section 4.4 we will show how the results from this section can be used in the exact sampling from the SABR model.

4.4. The SABR model

The Stochastic Alpha Beta Rho (SABR) model by Hagan *et al.* (2002) is a popular model in the financial industry because of the availability of an analytic asymptotic implied volatility formula. Practical applications of the SABR model include interpolation of volatility surfaces and the hedging of volatility risk. In the context of pricing interest rate derivatives, a combination of the SABR model and the market standard Libor Market Models (Rebonato 2007) is of particular interest. Relevant references on this topic include Mercurio and Morini (2009), Hagan and Leśniewski (2008) and Labordere (2007).

The SABR model is given by the following system of stochastic differential equations (SDEs) with constant parameters α and β:

$$\begin{aligned} \mathrm{d}S(t) &= \sigma(t)S^{\beta}(t)\mathrm{d}W_S(t), \quad S(t_0) = S_0, \\ \mathrm{d}\sigma(t) &= \alpha\sigma(t)\mathrm{d}W_\sigma(t), \quad \sigma(t_0) = \sigma_0, \end{aligned} \tag{4.28}$$

where $\mathrm{d}W_S(t)\mathrm{d}W_\sigma(t) = \rho\mathrm{d}t$ and the processes are defined under the T-forward measure $\mathbb{Q}^T$. The process $\sigma(t)$ follows a log-normal distribution (just as the standard Black–Scholes model). Further, since for a constant $\sigma(t) = \sigma$ the asset forward price, $S(t)$, follows a CEV process, one can expect that the conditional SABR process, $S(T)$ given the paths of $\sigma(t)$ on the interval $0 \leq t \leq T$, is a CEV process as well. The next step will be to combine the conditional CEV process with the joint distribution of $\sigma(T)$ and† $\int_0^T \sigma^2(s)\mathrm{d}s$.

As in Islah (2009) with the conditional samples $\int_0^T \sigma^2(t)\mathrm{d}t \big| \sigma(T)$ established, we are able to perform *exact*‡ sampling from $S(T)$ with the help of the following proposition.

Proposition 4.2 *[Cumulative distribution for conditional SABR process] For some* $S(0) > 0$, *the conditional cumulative distribution of* $S(T)$ *with an absorbing boundary at* $S(T) = 0$, *given* $\sigma(T)$ *and* $\zeta(T) := \int_0^T \sigma^2(t)\mathrm{d}t$, *reads*§

$$\begin{aligned} &\mathbb{P}\left[S(T) \leq y \middle| S(0) > 0, \sigma(T), \zeta(T) \right] \\ &\quad = 1 - F_{\chi^2(d,\lambda(y))}(a(T)), \end{aligned} \tag{4.29}$$

†The dependence on this integral will become clear later in the text.
‡The conditional probability is exact for $\rho = 0$ and for $\rho \neq 0$ constitutes an approximation.
§We use $\mathbb{P}[\cdot]$ as the SABR model is defined under the T-forward measure.

Table 8. Collocation points and appropriate inversions for the CEV model, boxed numbers indicate *virtual* points obtained by linear extrapolation of (x_i, y_i).

	$\boxed{x_1}$	$\boxed{x_2}$	$\boxed{x_3}$	x_4	x_5
x_i	−2.8570	−1.3556	0.0	1.3556	2.8570
$F_{\mathcal{N}(0,1)}(x_i)$	0.0021	0.0876	0.5000	0.9124	0.9979
$F^{-1}_{S(T)}(F_X(x_i))$	—	—	—	0.2770	0.9901
y_i	$\boxed{-0.3646}$	$\boxed{-0.3162}$	$\boxed{-0.0825}$	0.2770	0.9901

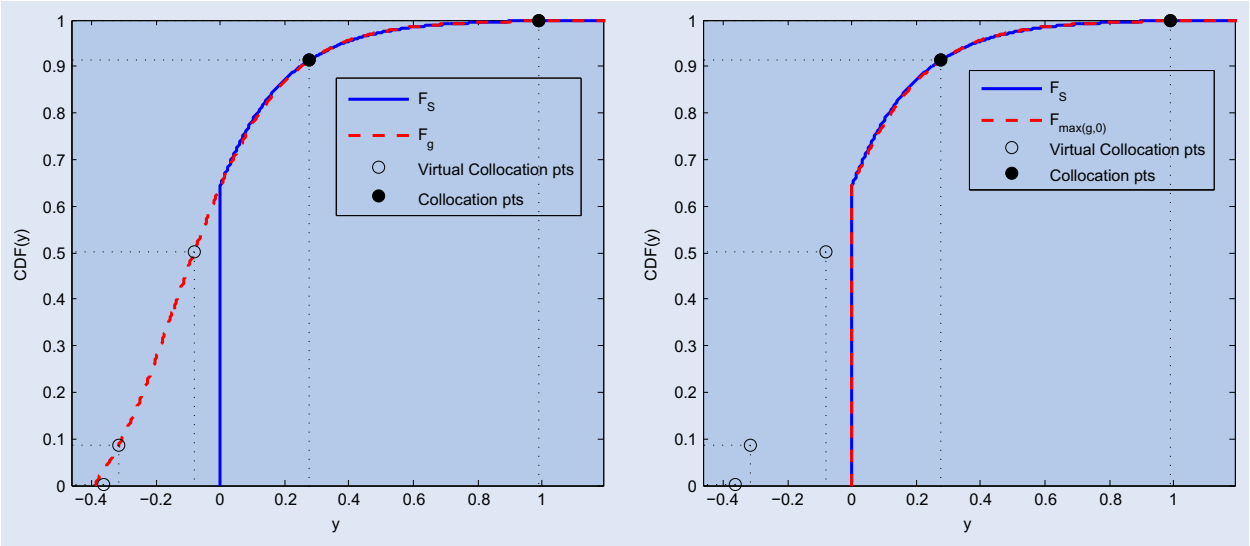

Figure 4. Analytic and approximate CDFs for the CEV process. The red dashed line corresponds to the samples obtained by the SCMC sampler. Left: without capping the samples at 0, Right: including capping at 0.

where

$$
\begin{aligned}
a(T) &= \frac{1}{\nu(T)}\left[\frac{S(0)^{1-\beta}}{(1-\beta)} + \frac{\rho}{\alpha}\big(\sigma(T)-\sigma(0)\big)\right]^2,\\
d &= 2 - \frac{1-2\beta-\rho^2(1-\beta)}{(1-\beta)(1-\rho^2)},\\
\lambda(y) &= \frac{y^{2(1-\beta)}}{(1-\beta)^2\nu(T)},\\
\nu(T) &= (1-\rho^2)\zeta(T).
\end{aligned}
\tag{4.30}
$$

$F_{\chi^2(d,\lambda(y))}(a(T))$ is the non-central chi-squared cumulative distribution function evaluated at $a(T)$, with d degrees of freedom and $\lambda(y)$ the non-centrality parameter.

Proof. Proof can be found in Islah (2009). □

The main objective here is to show how the sampling from the SABR model can be performed efficiently where we assume that conditional samples $\int_0^T \sigma^2(t)\mathrm{d}t\,|\sigma(T)$ have already been obtained.

We generate samples from S by using a collocation variable X, where the samples of $\sigma(T)$, σ_n, and of $\zeta(T)$, ζ_n are already given. Naturally the samples s_n are functions of σ_n and ζ_n, and we find, defined recursively as in the previous section, the following interpolation formula:

$$
\begin{aligned}
&s_n|\sigma_n,\zeta_n\\
&\approx \sum_{i=1}^{N_\sigma}\sum_{j=1}^{N_\zeta}\sum_{k=1}^{N} F^{-1}_{S|\sigma=\sigma_i,\zeta=\zeta_j}(F_X(x_i))\ell_i(\sigma_n)\ell_j(\zeta_n)\ell_k(\xi_n),
\end{aligned}
\tag{4.31}
$$

with

$$
\begin{aligned}
\ell_i(\sigma_n) &= \prod_{l=1,\ l\neq j}^{N_\sigma}\frac{\sigma_n-\sigma_l}{\sigma_i-\sigma_l}, \quad \ell_j(\zeta_n) = \prod_{l=1,\ l\neq j}^{N_\zeta}\frac{\zeta_n-\zeta_l}{\zeta_j-\zeta_l},\\
\ell_k(\xi_n) &= \prod_{l=1,\ l\neq k}^{N}\frac{\xi_n-x_l}{x_k-x_l},
\end{aligned}
\tag{4.32}
$$

where σ_n, ζ_n and ξ_n are samples of $\sigma(T)$, $\zeta(T)$ and X, respectively, and $\sigma_i, \sigma_l, \zeta_j, \zeta_l, x_k, x_l$ are the collocation points of $\sigma(T)$, $\zeta(T)$ and X, the variable used for approximating $S(T)$. In order to facilitate efficient evaluation of the Lagrange polynomial in 3D it is recommended to use the barycentric polynomial representation.

The equations given above rely on the availability of samples of the integrated log-normal ζ_n and the collocation points ζ_i for $i = 1, \ldots N_\zeta$. With the samples from the integrated log-normal variable given, we can calculate the collocations points, ζ, from the empirical moments, $\mathbb{E}[\zeta_n]$, $\mathbb{E}[\zeta_n^2]$, etc. The moments of the log-normal random variable $\sigma(T)$ may also be calculated from the corresponding characteristic function, $\phi_{\log(\sigma(T))}(-ik)$ for $k = 1, \ldots, N_\sigma$.

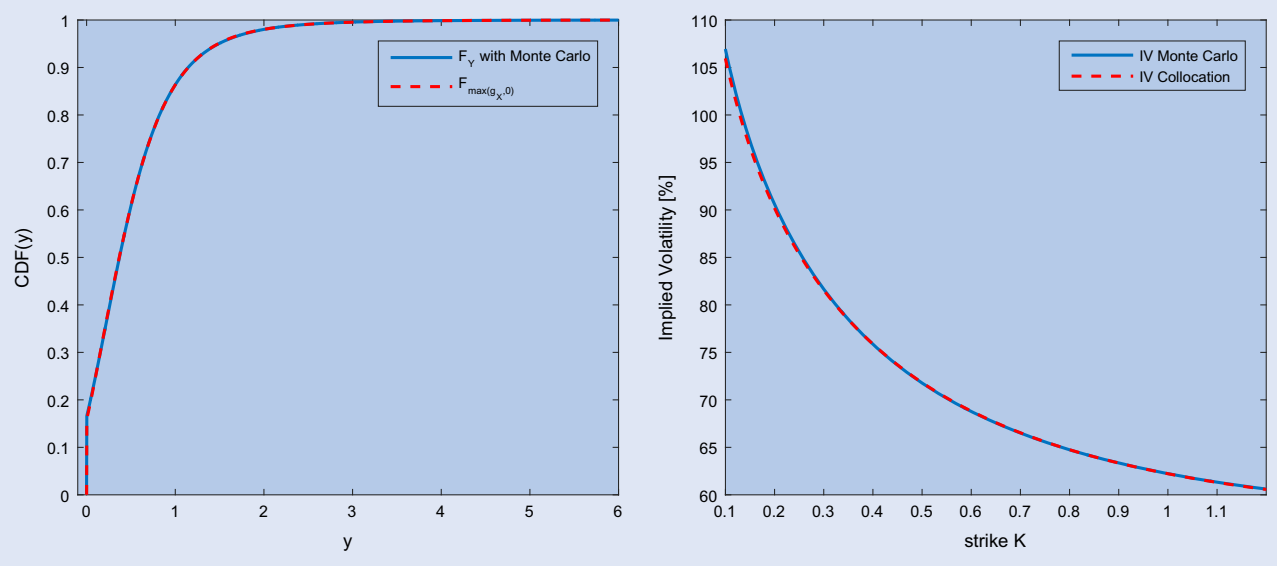

Figure 5. Results for Set I; Left: CDFs obtained by Monte Carlo (exact inversion of equation (4.31)) vs. results from the SCMC sampler. Right: corresponding implied volatilities.

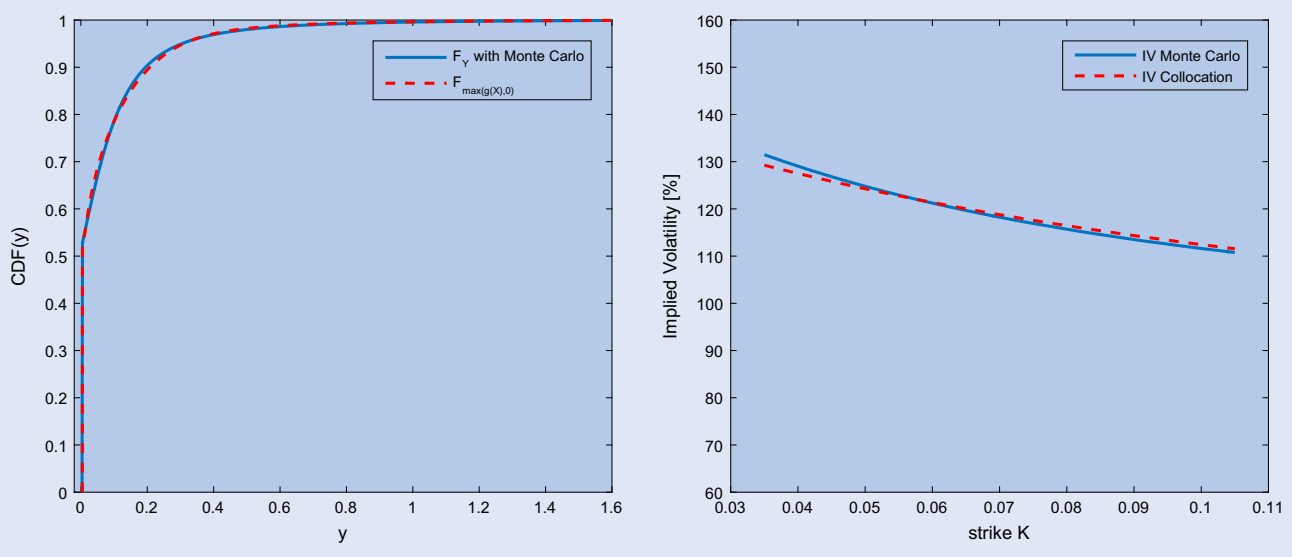

Figure 6. Results for Set II; Left: CDFs obtained by Monte Carlo (exact inversion of equation (4.31) vs. results from the SCMC sampler. Right: corresponding implied volatilities.

Once the collocation points $\sigma_1, \ldots, \sigma_{N_\sigma}$, $\zeta_1, \ldots, \zeta_{N_\zeta}$ and $x_1, \ldots, x_N$ are determined and after $N_\sigma \cdot N_\zeta \cdot N$ inversions of $F^{-1}_{S|\sigma=\sigma_i,\zeta=\zeta_i}(F_X(x_i))$ we can generate vectors of samples for X, $\sigma(T)$ and $\zeta(T)$ and evaluate polynomial (4.31) to obtain the samples $s_n|\sigma_n, \zeta_n$.

So far, we have shown how to obtain the samples from the SABR model at any time T_i. When dealing with *multiple Monte Carlo time steps*, one needs to include an *additional dimension* in the collocation method, i.e. for $S(T_2)|S(T_1), \sigma(T_2), \zeta(T_2)$ the samples by the SCMC sampler can be generated via:

$$s_n(T_2)|\sigma_n, \zeta_n, s_n(T_1) \approx \sum_{i=1}^{N_\sigma}\sum_{j=1}^{N_\zeta}\sum_{k=1}^{N_{S_2}}\sum_{l=1}^{N_{S_1}} y_{i,j,k,l}\ell_i(\sigma_n)\ell_j(\zeta_n)\ell_k(\xi_n)\ell_l(s_n(T_1)).$$

4.4.1. Exact sampling from the SABR model. We test the SCMC sampler for the exact sampling from the SABR model. In this experiment we will perform a simulation based on the conditional CDF given in proposition 4.2. Since this CDF is exact only for $\rho = 0$ we concentrate on this case here, with the following two SABR parameter sets (table 9): a moderate set, set I, with a high initial stock S_0 and moderate vol-vol parameter α and more extreme parameters in set II (as studied in Chen *et al.* (2012)). We take $T = 2$ which is standard in the FX market. In the experiment we also assume that the samples from the log-normal volatility $\sigma(T)$ and the integrated variance $\xi(T)$ are given. The collocation points for these two variables will be calculated based on the empirical moments. As the approximation for $S(T)$, we take 1.000.000 samples from $X \sim \mathcal{N}(0, 1)$. The samples, ξ_n, are then used in (4.31).

Table 9. SABR model parameters chosen in the experiments.

SABR	S_0	σ_0	α	β	T
Set I	0.5	0.5	0.4	0.5	2
Set II	0.07	0.4	0.8	0.5	2

In the experiment we have taken $N_\sigma = 6$, $N_\zeta = 6$ and $N = 9$ which results in $6 \cdot 6 \cdot 9$ expensive inversions of the CDF in (4.29). In the experiments we have used the concept of virtual points from section 4.3 and the grid stretching with $\sigma = 1.3$ from section 4.1. The results from the experiment are shown in figures 5 and 6. For both sets we have obtained excellent results.

5. Conclusions

We have presented the Stochastic Collocation Monte Carlo (SCMC) sampler for highly efficient sampling from computationally expensive distributions. We have shown that even for extreme distributions a few inversions of the cumulative distribution function are sufficient for obtaining any number of Monte Carlo samples. We have also shown that, although our method allows any distribution for approximation of the *expensive* distribution, high-quality results are obtained by using standard normals, even for a distribution with an atom at zero. In our numerical experiment section we have shown that the exact simulation scheme by Broadie–Kaya for sampling from the Heston model can be performed highly efficiently, as well as the simulation of the SABR model.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Abramowitz, M. and Stegun, I.A., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th ed., Vol. 55, Applied Mathematics Series, 1972 (National Bureau of Standards: Washington, DC).

Andersen, L.B.G., Simple and efficient simulation of the Heston stochastic volatility model. *J. Comput. Finance*, 2008, **11**(3), 1–22.

Babuška, I., Nobile, F. and Tempone, R., A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Rev.*, 2010, **52**, 317–355.

Beck, J., Nobile, F., Tamellini, L. and Tempone, R., On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods. *Math. Models Methods Appl. Sci.*, 2012, **22**(9), 1–33.

Berrut, J.-P. and Trefethen, L.N., Barycentric Lagrange interpolation. *SIAM Rev.*, 2004, **46**(3), 501–517.

Bieri, M. and Schwab, C., Sparse high order FEM for elliptic sPDEs. *Comput. Methods Appl. Mech. Eng.*, 2009, **198**, 1149–1170.

Broadie, M. and Kaya, Ö., Exact simulation of stochastic volatility and other affine jump diffusion processes. *Oper. Res.*, 2006, **54**, 217–231.

Cameron, R.H. and Martin, W.T., The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals. *Ann. Math.*, 1947, **48**(2), 385–392.

Chen, B., Oosterlee, C.W. and van der Weide, H., A low-bias simulation scheme for the SABR stochastic volatility model. *Int. J. Theor. Appl. Finance*, 2012, **15**(2), 1–37.

Fang, F. and Oosterlee, C.W., A novel pricing method for european options based on Fourier-cosine series expansions. *SIAM J. Sci. Comput.*, 2009, **31**(2), 826–848.

Favard, J., Sur les polynomes de Tchebicheff. *C. R. Acad. Sci., Paris*, 1935, **200**, 2052–2053.

Fritsch, F.N. and Carlson, R.E., Monotone piecewise cubic interpolation. *SIAM J. Numer. Anal.*, 1980, **17**, 238–246.

Glasserman, P. and Kim, K.-K., Gamma expansion of the Heston stochastic volatility model. *Finance Stoch.*, 2011, **15**, 267–296.

Golub, G.H. and Welsch, J.H., Calculation of Gauss quadrature rules. *Math. Comput.*, 1969, **23**(106), 221–230.

Hagan, P.S., Kumar, D., Leśniewski, A.S. and Woodward, D.E., Managing smile risk. *Wilmott Mag.*, 2002, pp. 84–108.

Hagan, P.S. and Leśniewski, A.S., Libor market model with SABR style stochastic volatility. Working Paper, 2008.

Heston, S., A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.*, 1993, **6**, 327–343.

Hong Chan, J. and Joshi, M., Fast and accurate long-stepping simulation of the Heston stochastic volatility model. *J. Comput. Finance*, 2013, **15**(3), 47–97.

Islah, O., Solving SABR in exact form and unifying it with libor market model. SSRN eLibrary, 2009.

Johnson, N.L., Kotz, S. and Balakrishnan, N., *Continuous Univariate Distributions*, 2nd ed., 1994 (Wiley: New York).

Labordere, H., Combining the SABR and LMM models. *Risk*, 2007, October, 68–74.

Lord, R. and Kahl, C., Complex logarithms in Heston-like models. *Math. Finance*, 2010, **20**, 671–694.

Mercurio, F. and Morini, M., No-arbitrage dynamics for a tractable SABR term structure Libor Model. In *Modelling Interest Rates: Last Advances for Derivatives Pricing*, edited by F. Mercurio, 2009 (Risk Books: London).

Micchelli, C.A., A constructive approach to Kergin interpolation in $\mathbb{R}^k$ multivariate B-splines and Lagrange interpolation. *Rocky Mt. J. Math.*, 1980, **10**, 485–497.

Nobile, F., Tempone, R. and Webster, C.G., An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 2008a, **46**, 2411–2442.

Nobile, F., Tempone, R. and Webster, C.G., A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 2008b, **46**, 2309–2345.

Rebonato, R., A time-homogeneous, SABR-consistent extension of the LMM. Working Paper, Royal Bank of Scotland, 2007.

Sauer, T. and Xu, Y., On multivariate Lagrange interpolation. *Math. Comput.*, 1995, **64**, 1147–1170.

Schroder, M., Computing the constant elasticity of variance option pricing formula. *J. Finance*, 1989, **1**(44), 211–218.

Witteveen, J.A.S. and Iaccarino, G., Simplex stochastic collocation with eno-type stencil selection for robust uncertainty quantification. *J. Comput. Phys.*, 2013, **239**, 1–21.

Xiu, D. and Hesthaven, J.S., High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comput.*, 2005, **27**(3), 1118–1139.

Appendix 1. Lagrange polynomials

In this appendix we give a brief introduction to Lagrange polynomials and their different representations.

For a basis of monomials $\mathbf{m}(x) = (1, x, x^2, \ldots, x^{N-1})^{\mathrm{T}}$, a function g can be decomposed as,

$$g_N(x) = a_0^{(N)} + a_1^{(N)} x + \cdots + a_{N-1}^{(N)} x^{N-1}, \quad \text{with} \quad g_N(x_i) = y_i, \tag{A1}$$

with some coefficients a_i, $i \in \{0, \ldots, N-1\}$. The coefficients $\mathbf{a} = (a_0^{(N)}, \ldots, a_{N-1}^{(N)})^{\mathrm{T}}$ are determined by the interpolation conditions $g_N(x_i) = y_i$, for $i = 1, \ldots, N$. These coefficients can be found as solutions of the following linear system, $V\boldsymbol{a} = \boldsymbol{y}$, i.e.

$$\begin{pmatrix} 1 & x_1^1 & x_1^2 & \dots & x_1^{N-1} \\ 1 & x_2^1 & x_2^2 & \dots & x_2^{N-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_N^1 & x_N^2 & \dots & x_N^{N-1} \end{pmatrix} \begin{pmatrix} a_0^{(N)} \\ a_1^{(N)} \\ \vdots \\ a_{N-1}^{(N)} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix},$$

with matrix V the so-called Vandermonde matrix. With this representation the interpolating polynomial $g_N(x)$ can be presented as $g_N(x) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{m}(x)$.

When interpolating with the Lagrange formula (2.3), for each new x-value we would need to perform $\mathcal{O}(N^2)$ operations. A variant called the *barycentric formula* (Berrut and Trefethen 2004) exists, which requires only $\mathcal{O}(N)$ operations per interpolation point x. The barycentric weight coefficients λ_i are defined as

$$\lambda_i = \frac{1}{\prod_{j=1, i\neq j}^{N}(x_i - x_j)}, \quad i = 1, \dots, N, \tag{A2}$$

and $\ell(x) = (x - x_1)(x - x_2)\dots(x - x_N)$. Each $\ell_i(x)$ can be written as:

$$\ell_i(x) = \ell(x)\frac{\lambda_i}{x - x_i}. \tag{A3}$$

Polynomial $g_N(x)$ is now expressed as,

$$g_N(x) = \sum_{i=1}^{N} y_i \ell_i(x), \quad \ell_i(x) = \frac{\lambda_i \ell(x)}{x - x_i}, \quad i = 1, \dots, N. \tag{A4}$$

The representation in (A4) enables us to express the Lagrange polynomial in terms of a Lagrange basis, as $g_N(x) = \boldsymbol{y}\boldsymbol{\ell}(x)$, with $\boldsymbol{y} = (y_1, y_2, \dots, y_N)$.

A.1. Optimal collocation points, relation to moments

This section recalls this relation and the main properties of orthogonal polynomials.

A sequence of orthogonal polynomials $\{p_i\}_{i=0}^{N}$, with $\deg(p_i) = i$, is said to be orthogonal in L^2 with respect to PDF $f_X(X)$ of X, if the following holds,

$$\begin{aligned} \mathbb{E}\left[p_i(X)p_j(X)\right] &= \int_{\mathbb{R}} p_i(x)p_j(x)f_X(x)\mathrm{d}x \\ &= \delta_{ij}\mathbb{E}\left[p_i^2(X)\right], \quad i, j = 0, \dots, N, \end{aligned} \tag{A5}$$

with $\mathbb{R}$ the support of X, δ_{ij} the Kronecker delta.

An important property of orthogonal polynomials is their definition in terms of a recurrence relation, given in the theorem below.

Theorem 1.1 (Recurrence in orthogonal polynomials) *For any given density function $f_X(\cdot)$, a unique sequence of monic orthogonal polynomials $p_i(x)$ exists, with $deg(p_i(x)) = i$, which can be constructed as follows,*

$$p_{i+1}(x) = (x - \alpha_i)p_i(x) - \beta_i p_{i-1}, \quad i = 0, \dots, N-1, \tag{A6}$$

where $p_{-1}(x) \equiv 0$, $p_0(x) \equiv 1$ and where α_i and β_i are the recurrence coefficients,

$$\begin{aligned} \alpha_i &= \frac{\mathbb{E}[Xp_i^2(X)]}{\mathbb{E}[p_i^2(X)]}, \text{ for } i = 0, \dots, N-1, \\ \beta_i &= \frac{\mathbb{E}[p_i^2(X)]}{\mathbb{E}[p_{i-1}^2(X)]}, \text{ for } i = 1, \dots, N-1, \end{aligned} \tag{A7}$$

with $\beta_0 = 0$.

Proof. The proof can be found in Favard (1935). □

Parameters α_i and β_i are completely determined in terms of the moments of random variable X. For the standard normal distribution they are $\alpha_i = 0$ and $\beta_i = i$.

For many densities (weight functions in the integration in (A8)) the three-term recurrence relation in (A6) of the corresponding orthogonal polynomials has been determined. Sometimes however density $f_X(\cdot)$ is not known explicitly or its evaluation is computationally expensive. In such cases it is desirable to express α_i and β_i in (A7) in terms of the moments of X (Golub and Welsch 1969). Let us consider the monomials $m_i(X) = X^i$, and define $\mu_{i,j}$ as follows,

$$\begin{aligned} \mu_{i,j} = \mathbb{E}\left[m_i(X)m_j(X)\right] &= \int_{\mathbb{R}} x^{i+j} f_X(x)\mathrm{d}x \\ &= \mathbb{E}[X^{i+j}], \quad i, j = 0, \dots, N. \end{aligned} \tag{A8}$$

From all moments $\mu_{i,j}$ we construct the so-called Gram matrix $M = \{\mu_{i,j}\}_{i,j=0}^{N}$, which is symmetric and contains all moments $\{1, \mathbb{E}[X^1], \dots, \mathbb{E}[X^{2N}]\}$. Since matrix M is, by definition, positive definite (Golub and Welsch 1969), we decompose $M = R^{\mathrm{T}}R$, by the Cholesky decomposition of M.

The next step is to relate the Cholesky lower triangular matrix R to the orthogonal polynomials. This relation has been found in Golub and Welsch (1969) and is given by,

$$\begin{aligned} \alpha_j &= \frac{r_{j,j+1}}{r_{j,j}} - \frac{r_{j-1,j}}{r_{j-1,j-1}}, \quad j = 1, \dots, N, \quad \text{and} \\ \beta_j &= \left(\frac{r_{j+1,j+1}}{r_{j,j}}\right)^2, \quad j = 1, \dots, N-1, \end{aligned} \tag{A9}$$

with $r_{0,0} = 1$ and $r_{0,1} = 0$ and where $r_{i,j}$ is the (i, j)th element of matrix R. This relation gives us the expressions for α_j and β_j when the matrix of moments has been computed.

The next step is to relate the coefficients α_i and β_i to the zeros of the orthogonal polynomials $p_n(X)$, $n = 0, \dots, N$. This can be done by the *eigenvalue method*, presented in the theorem below.

Theorem 1.2 (Eigenvalue method) *The zeros x_i, $i = 1, \dots, N$, of the orthogonal polynomial $p_N(X)$ are the eigenvalues of the symmetric tridiagonal matrix,*

$$\widehat{J} := \begin{pmatrix} \alpha_1 & \sqrt{\beta_1} & 0 & 0 & 0 \\ \sqrt{\beta_1} & \alpha_2 & \sqrt{\beta_2} & 0 & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & 0 & \sqrt{\beta_{N-2}} & \alpha_{N-1} & \sqrt{\beta_{N-1}} \\ 0 & 0 & 0 & \sqrt{\beta_{N-1}} & \alpha_N \end{pmatrix}, \tag{A10}$$

i.e. $\boldsymbol{x} = (x_1, x_1, \dots, x_N)^{\mathrm{T}}$ is a vector of eigenvalues satisfying $\widehat{J}\boldsymbol{v} = x_i\boldsymbol{v}$ with $i = 1, \dots, N$ for any real vector $\boldsymbol{v}$, with α_i and β_i being the coefficients of the three-term recurrence relation (A6).

Proof. The proof can be found in Golub and Welsch (1969). □

Based on the coefficients α_i and β_i, *the collocation points x_i are the eigenvalues of matrix (A10)*. Eigenvalue calculation for a tridiagonal matrix in theorem 1.2 is performed by e.g. the Lanczos algorithm.

A basic illustrative example for the choice of optimal collocation points is given below.

A.2. Basic example

We consider a simple, illustrative example, with a gamma distributed random variable $Y \sim \Gamma(5, 2)$, where the first argument is the shape parameter and the second controls the scale. As the second random variable we take a standard normal variable, $X \sim \mathcal{N}(0, 1)$. Our objective is to sample from Y with only a few inversions F_Y^{-1} and by using the samples from the normal distribution X.

Let us choose a set of $N = 2$ collocation points $x_1 = -0.7$, and $x_2 = 0.7$. For these collocation points we calculate the corresponding cumulative probability, $F_{\mathcal{N}(0,1)}(x_1) = 0.2420$, $F_{\mathcal{N}(0,1)}(x_2) = 0.758$ and we calculate the *expensive* inversion $y_1 = F_Y^{-1}(F_{\mathcal{N}(0,1)}(x_1)) = 6.6498$, $y_2 = F_Y^{-1}(F_{\mathcal{N}(0,1)}(x_2)) = 12.6826$. As the final step to obtain M approximated samples of Y we need to generate M samples from X and evaluate polynomial $g_2(X)$ in (2.3). In the experiment we have generated $M = 10^6$ samples and the corresponding CDF is depicted in the left-hand picture of figure A1. In the considered example we obtain a satisfactory result, especially in the region between the evaluated collocation points y_n. To cover the whole domain of X, we take one additional collocation point $x_3 = 1.5$ for which we get $F_{\mathcal{N}(0,1)}(x_3) = 0.9332$ and $y_3 =$

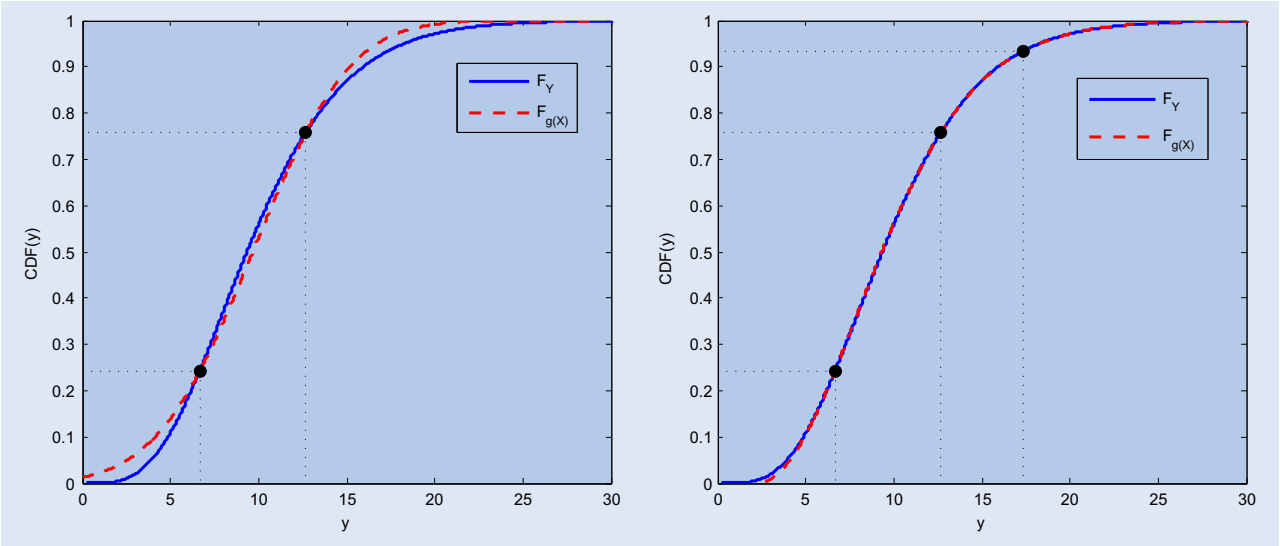

Figure A1. Left-hand side: Exact CDF for $\Gamma(5, 2)$ and approximation with two collocation points. Right-hand side: Exact CDF for $\Gamma(5, 2)$ and approximation with three collocation points.

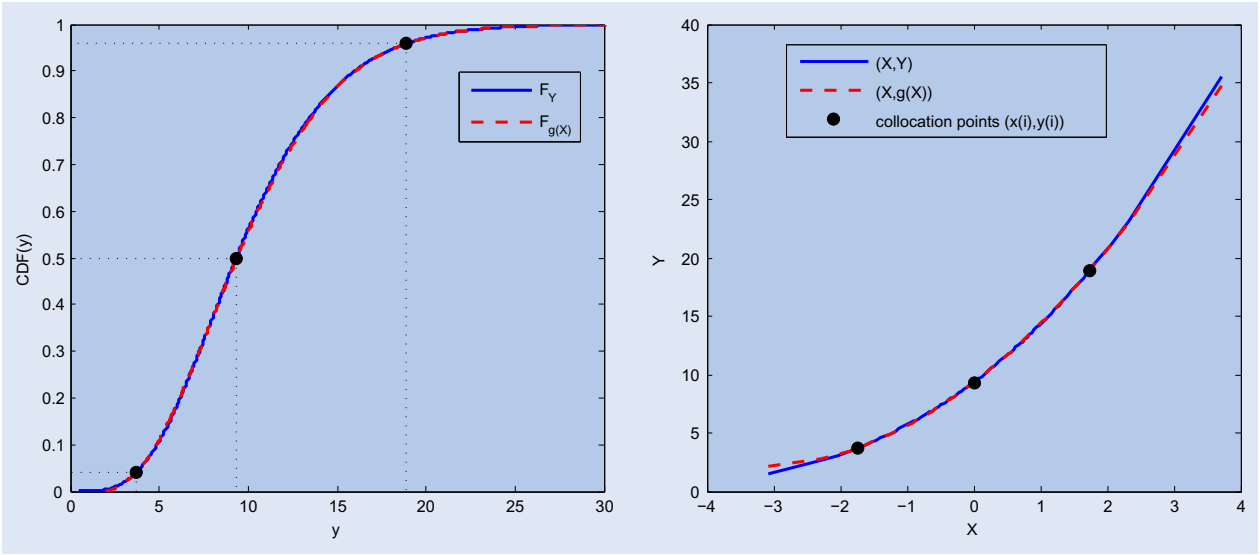

Figure A2. Left-hand side: Exact CDF for $Y \sim \Gamma(5, 2)$ and approximation $g_N(X)$ with $N = 3$ collocation points.

$F_Y^{-1}(F_{\mathcal{N}(0,1)}(x_3)) = 17.3581$. The results for this additional collocation point are shown in the right-hand picture of figure A1, showing a significant improvement of the fit. Based on this illustrative example we see that the method performs very well once an appropriate set of collocation points $\boldsymbol{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$ is chosen.

Now, we use the results from appendix 1 where we have established an algorithm for calculating optimal collocation points. In order to apply the technique, we need to calculate with equation (2.4) $2N$ moments of random variable X. Matrix M and its upper triangular matrix are given by

$$M = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 3 \\ 1 & 0 & 3 & 0 \\ 0 & 3 & 0 & 15 \end{pmatrix} \quad \Rightarrow \quad R = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1.4142 & 0 \\ 0 & 0 & 0 & 2.4495 \end{pmatrix}. \tag{A11}$$

By applying equation (A9) we find, $\boldsymbol{\alpha} = (0, 0, 0)^{\mathrm{T}}$ and $\boldsymbol{\beta} = (1, 2)^{\mathrm{T}}$ so the Jacobian matrix $\widehat{J}$ and the corresponding eigenvalues (the collocation points) are given by the upper triangular matrices, i.e.

$$\widehat{J} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1.4142 \\ 0 & 1.4142 & 0 \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{x} = (-1.7321, 0, 1.7321)^{\mathrm{T}}, \tag{A12}$$

where $\boldsymbol{x}$ stands for the vector of the optimal collocation points.† As before, we calculate the corresponding CDFs $\boldsymbol{F}_{\mathcal{N}(0,1)}(x) = (0.0416, 0.5, 0.9584)^{\mathrm{T}}$ and $\boldsymbol{y} = F_Y^{-1}(\boldsymbol{F}_{\mathcal{N}(0,1)}(\boldsymbol{x})) = (3.7386, 9.3418, 18.8938)^{\mathrm{T}}$, and present the corresponding results in figure A2 (left). figure A2 (right) depicts the collocation points, the relation between the variables X and Y and the approximating function $g_N(X)$. We see that, although the fit at the collocation points is exact and around the collocation points it is highly satisfactory, the quality of the fit deteriorates in the tails. In figure A3 we therefore present the results

†In appendix A.3 the collocation points for a standard normal have been tabulated.

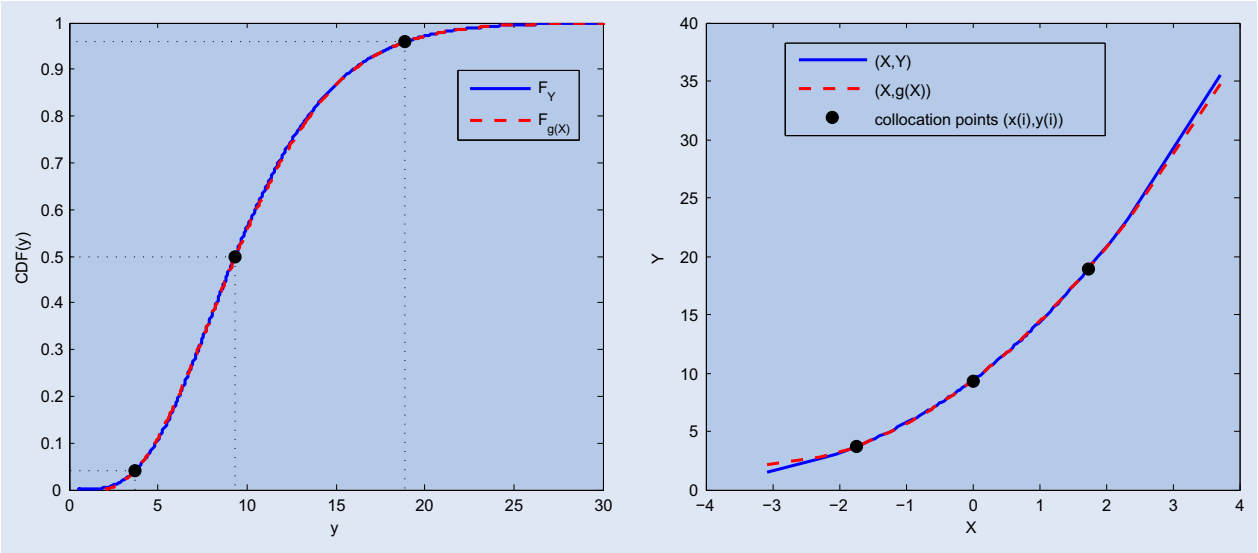

Figure A3. Left-hand side: Exact CDF for $\Gamma(5, 2)$ and approximation with $N = 5$ collocation points.

Table A1. Collocation points for a standard normal $X \sim \mathcal{N}(0, 1)$.

x_i	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$	$N = 7$	$N = 8$	$N = 9$	$N = 10$	$N = 11$
x_1	−1	−1.7321	−2.3344	−2.8570	−3.3243	−3.7504	−4.1445	−4.5127	−4.8595	−5.1880
x_2	1	0.0	−0.7420	−1.3556	−1.8892	−2.3668	−2.8025	−3.2054	−3.5818	−3.9362
x_3		1.7321	0.7420	0.0	−0.6167	−1.1544	−1.6365	−2.0768	−2.4843	−2.8651
x_4			2.3344	1.3556	0.6167	0.0	−0.5391	−1.0233	−1.4660	−1.8760
x_5				2.8570	1.8892	1.1544	0.5391	0.0	−0.4849	−0.9289
x_6					3.3243	2.3668	1.6365	1.0233	0.4849	0.0
x_7						3.7504	2.8025	2.0768	1.4660	0.9289
x_8							4.1445	3.2054	2.4843	1.8760
x_9								4.5127	3.5818	2.8651
x_{10}									4.8595	3.9362
x_{11}										5.1880

for $N = 5$ with the collocation points calculated according to the technique described. The obtained results confirm that by increasing the number of collocation points, we can also improve the fit significantly.

A.3. Tabulated collocation points for a standard normal

Table A1 presents the collocation points for a standard normal distribution.