

# On Pricing of Discrete Asian and Lookback Options under the Heston Model

Leonardo Perotti<sup>a,\*</sup>, Lech A. Grzelak<sup>a,b</sup>

<sup>a</sup>*Mathematical Institute, Utrecht University, Utrecht, the Netherlands*

<sup>b</sup>*Financial Engineering, Rabobank, Utrecht, the Netherlands*

---

## Abstract

We propose a new, data-driven approach for efficient pricing of – fixed- and float-strike – discrete arithmetic Asian and Lookback options when the underlying process is driven by the Heston model dynamics. The method proposed in this article constitutes an extension of [19], where the problem of sampling from time-integrated stochastic bridges was addressed. The model relies on the Seven-League scheme [16], where artificial neural networks are employed to “learn” the distribution of the random variable of interest utilizing stochastic collocation points [10]. The method results in a robust procedure for Monte Carlo pricing. Furthermore, semi-analytic formulae for option pricing are provided in a simplified, yet general, framework. The model guarantees high accuracy and a reduction of the computational time up to thousands of times compared to classical Monte Carlo pricing schemes.

*Keywords:* Discrete Arithmetic Asian Option, Discrete Lookback Option, Heston Model, Stochastic Collocation (SC), Artificial Neural Network (ANN), Seven-League Scheme (7L).

---

## 1. Introduction

A non-trivial problem in the financial field is the pricing of path-dependent derivatives, as for instance *Asian* and *Lookback* options. The payoffs of such derivatives are expressed as functions of the underlying process monitored over the life of the option. The monitoring can either be *continuous* or *discrete*. Depending on different specifications of the underlying dynamics, only a few case-specific theoretical formulae exists. For example, under the Black-Scholes and Merton log-normal dynamics, closed-form formulae were derived for *continuously-monitored geometric* Asian options (see, for instance, [6]). In the same model framework, [8, 5] derived an analytic formula for the *continuously-monitored* Lookback option, using probabilistic arguments as the *reflection principle*. However, options whose payoffs are *discretely-monitored* are less tractable analytically, and so approximations are developed, as for the discrete Lookback option under the lognormal dynamics [12]. Furthermore, stochastic volatility frameworks are even more challenging for the pricing task, where no applicable closed-form theoretical solutions are known.

Whenever an exact theoretical pricing formula is not available, a rich literature on numerical methods and approximations exists. The three main classes of approaches are Monte Carlo (MC) methods (e.g. [13]), partial differential equations (PDEs) techniques (see the extensive work in [22] and [21]), and Fourier-inversion based techniques (among many relevant works, we report [2, 25]).

Monte Carlo methods are by far the most flexible approaches, since they neither require any particular assumption on the underlying dynamics, nor on the targeted payoff. Furthermore, they benefit of a straightforward implementation based on the discretization of the time horizon as, for instance, the well-known *Euler-Maruyama* scheme. The cost to be paid, however, is typically a significant computational time to get accurate results. PDEs approaches are more

---

\*Corresponding author.

*Email addresses:* L.Perotti@uu.nl (Leonardo Perotti), L.A.Grzelak@uu.nl (Lech A. Grzelak)

problem-specific since they require the derivation of the partial differential equation which describe the evolution of the option value over time. Then, the PDE is usually solved using finite difference methods. Fourier-inversion based techniques exploit the relationship between probability density function (PDF) and characteristic function (ChF) to recover the underlying transition density by means of Fast Fourier Transform (FFT). Thanks to the swift algorithm for FFT, such methods produce high-speed numerical evaluation, but they often result problem-specific, depending on the underlying dynamics.

In this article we propose an extension to MC schemes that allows for efficient pricing of *discretely-monitored Asian and Lookback options*, without losing the flexibility typical of MC methods. We develop the methodology in the complex framework of the stochastic volatility model of Heston [11], with an extensive application to the case of *Feller condition* not satisfied. Under this dynamics, we show how to price fixed- or float-strike discrete Asian and Lookback options. Moreover, the pricing model is applied also for the challenging task of pricing options with both a fixed- and a float-strike component. We underline that the strengths of the method are its speed and accuracy, coupled with a significant flexibility. The procedure is, indeed, independent on the underlying dynamics (it could be applied, for instance, at any stochastic volatility model), and it is not sensitive to the targeted payoff.

Inspired by the works in [16, 19], the method relies on the technique of Stochastic Collocation (SC) [10], which is employed to accurately approximate the targeted distribution by means of piecewise polynomials. Artificial neural networks (ANNs) are used for fast recovery of the coefficients which uniquely determine the piecewise approximation. Given these coefficients, the pricing can be performed in a “MC fashion” sampling from the target (approximated) distribution and computing the numerical average of the discounted payoffs. Furthermore, in a simplified setting, we provide a semi-analytic formula which allows to directly price options without the need of sampling from the desired distribution. In both the situations (MC and semi-analytic pricing) we report a significant computational speed-up, without affecting the accuracy of the result which remains comparable with the one of highly expensive MC methods.

The remainder of the paper is as follows. In Section 2, we formally define *discrete arithmetic Asian and Lookback options*, as well as the model framework for the underlying process. Then, in Section 3, the pricing model is described. Two different cases are considered, in increasing order of complexity, to handle efficiently both unconditional sampling (Section 3.2) and conditional sampling (Section 3.3) for pricing of discrete arithmetic Asian and Lookback options. Section 4 provides theoretical support to the given numerical scheme. The quality of the methodology is also inspected empirically with several numerical experiments, reported in Section 5. Section 6 concludes.

## 2. Discrete arithmetic Asian and Lookback options

In a generic setting, given the present time  $t_0 \geq 0$ , the payoff at time  $T > t_0$  of a discrete arithmetic Asian or Lookback option, with underlying process  $S(t)$ , can be written as:

$$H_\omega(T; S) = \max\left(\omega(A(S) - K_1 S(T) - K_2), 0\right), \quad (2.1)$$

where  $A(S) \equiv A(S(t); t_0 < t \leq T)$  is a deterministic function of the underlying process  $S$ , the constants  $K_1, K_2 \geq 0$  control the float- and fixed-strikes of the option, and  $\omega = \pm 1$ . Particularly, discrete arithmetic Asian and Lookback options are obtained setting the quantity  $A(S)$  respectively as follows:

$$A(S) := \frac{1}{N} \sum_{n \in I} S(t_n), \quad A(S) := \omega \max_{n \in I} \omega S(t_n), \quad (2.2)$$

with  $I = \{1, \dots, N\}$  a set of indexes,  $t_1 < t_2 < \dots < t_N = T$ , a discrete set of future monitoring dates, and  $\omega$  as in (2.1).

Note that in both the cases of discrete arithmetic Asian and Lookback options  $A(S)$  is expressed as a deterministic transformation of the underlying process’ path, which is the only requirement to apply the proposed method. Therefore, in the paper, we refer always to the class of *discrete arithmetic Asian options*, often just called *Asian options*, for simplicity. However,

the theory holds for both the classes of products. Actually, the pricing model applies to any product with a *path-dependent European-type payoff*, requiring only a different definition of  $A(S)$ .

### 2.1. Pricing of arithmetic Asian options and Heston framework

This section focuses on the risk-neutral pricing of arithmetic – fixed- and float strike – Asian options, whose payoff is given in (2.1) with  $A(S)$  in (2.2). By setting  $K_1 = 0$  or  $K_2 = 0$ , we get two special cases: the fixed- or the float-strike arithmetic Asian option. From Equation (2.1), for  $K_1 = 0$ , the simplified payoff of a *fixed-strike arithmetic Asian option* reads:

$$H_{\omega}^{\text{fx}}(T; S) = \max\left(\omega(A(S) - K_2), 0\right), \quad (2.3)$$

therefore, with the risk-neutral present value:

$$V_{\omega}^{\text{fx}}(t_0) = \frac{M(t_0)}{M(T)} \mathbb{E}_{t_0}^{\mathbb{Q}} \left[ \max\left(\omega(A(S) - K_2), 0\right) \right], \quad (2.4)$$

where we assume the money-savings account  $M(t)$  to be defined through the deterministic dynamics  $dM(t) = rM(t)dt$  with constant interest rate  $r \geq 0$ . For  $K_2 = 0$ , however, the payoff in Equation (2.1), becomes the one of a *float-strike arithmetic Asian option*:

$$H_{\omega}^{\text{fl}}(T; S) = \max\left(\omega(A(S) - K_1 S(T)), 0\right). \quad (2.5)$$

The payoff in (2.5) is less tractable when compared to the one in (2.3), because of the presence of two dependent stochastic unknowns, namely  $S(T)$  and  $A(S)$ . However, a similar representation to the one in Equation (2.4) can be achieved, allowing for a unique pricing approach in both cases. By a change of measure from the risk-neutral measure  $\mathbb{Q}$  to the measure  $\mathbb{Q}^S$  associated with the numéraire  $S(t)$ , i.e. the *stock measure*, we prove the following proposition.

**Proposition 2.1** (Pricing of float-strike arithmetic Asian option under the stock measure). *Under the stock measure  $\mathbb{Q}^S$ , the value at time  $t_0 \geq 0$  of an arithmetic float-strike Asian option, with maturity  $T > t_0$  and future monitoring dates  $t_n$ ,  $n \in \{1, \dots, N\}$ , reads:*

$$V_{\omega}^{\text{fl}}(t_0) = S(t_0) \mathbb{E}_{t_0}^S \left[ \max\left(\omega(A^{\text{fl}}(S) - K_1), 0\right) \right], \quad (2.6)$$

with  $A^{\text{fl}}(S)$ , defined as:

$$A^{\text{fl}}(S) := A\left(\frac{S(\cdot)}{S(T)}\right) = \frac{A(S)}{S(T)},$$

where  $A(S)$  is defined in (2.2).

*Proof.* For a proof, see Appendix A.1. ■

Both the representations in Equations (2.4) and (2.6) can be treated in the same way. This means that the present value of both a fixed- and a float-strike Asian option can be computed similarly, as stated in the following proposition.

**Proposition 2.2** (Symmetry of fixed- and float-strike Asian option present value). *Let us consider the process  $S(t)$  and the money-savings account  $M(t)$ , for  $t \geq t_0$ . Then, the same representation holds for the value at time  $t_0$  of both fixed- or float-strike Asian options, with maturity  $T > t_0$ , underlying process  $S(t)$ , and future monitoring dates  $t_n$ ,  $n \in \{1, \dots, N\}$ . The present value is given by:*

$$V_{\omega}^{\lambda}(t_0) = \begin{cases} \frac{M(t_0)}{M(T)} \mathbb{E}_{t_0}^{\mathbb{Q}} \left[ \max\left(\omega(A(S) - K_2), 0\right) \right], & \text{for } \lambda = \text{fx}, \\ S(t_0) \mathbb{E}_{t_0}^S \left[ \max\left(\omega(A^{\text{fl}}(S) - K_1), 0\right) \right], & \text{for } \lambda = \text{fl}. \end{cases} \quad (2.7)$$

*Proof.* The proof follows by direct comparison between Equations (2.4) and (2.6). ■

When  $K_1 \neq 0$  and  $K_2 \neq 0$ , Equation (2.1) is the payoff of a fixed- *and* float-strikes arithmetic Asian option. Its present value does not allow any simplified representation, and we write it as the expectation of the discounted payoff under the risk-neutral measure  $\mathbb{Q}$ :

$$V_\omega(t_0) = \frac{M(t_0)}{M(T)} \mathbb{E}_{t_0}^{\mathbb{Q}} \left[ \max(\omega(A(S) - K_1 S(T) - K_2), 0) \right]. \quad (2.8)$$

By comparing Equations (2.7) and (2.8) a difference in the two settings is unveiled. Equation (2.7) is characterized by a unique unknown stochastic quantity  $A(S)$ , whereas in (2.8) an additional term appears, namely the stock price at final time,  $S(T)$ . Furthermore, the two stochastic quantities in (2.8) are not independent. This might suggest that different procedures should be employed for the different payoffs. In particular, in a MC setting, to value (2.7) we only have to sample from the *unconditional* distribution of  $A(S)$ ; while in (2.8) the MC scheme requires dealing with both the sampling of  $S(T)$  and the *conditional* sampling of  $A(S)|S(T)$ .

Let us define the stochastic volatility dynamics of Heston for the underlying stochastic process  $S(t)$ , with initial value  $S(t_0) = S_0$ , through the following system of stochastic differential equations (SDEs):

$$dS(t) = rS(t)dt + \sqrt{v(t)}S(t)dW_x(t), \quad S(t_0) = S_0, \quad (2.9)$$

$$dv(t) = \kappa(\bar{v} - v(t))dt + \gamma\sqrt{v(t)}dW_v(t), \quad v(t_0) = v_0, \quad (2.10)$$

with  $r, \kappa, \bar{v}, v_0 \geq 0$ ,  $\gamma > 0$  the constant rate, the speed of mean reversion, the long-term mean of the variance process, the initial variance, and the volatility-of-volatility, respectively.  $W_x(t)$  and  $W_v(t)$  are Brownian Motions (BMs) under the risk-neutral measure  $\mathbb{Q}$  with correlation coefficient  $\rho < 0$ , i.e.,  $dW_x(t)dW_v(t) = \rho dt$ .

The dynamics in (2.9) and (2.10) are defined in the risk-neutral framework. However, Proposition 2.1 entails a different measure framework, whose dynamics still fall in the class of stochastic volatility model of Heston, with adjusted parameters, as shown in the following proposition.

**Proposition 2.3** (The Heston model under the underlying process measure). *Using the same notation as in (2.9) and (2.10), under the stock  $S(t)$  measure,  $\mathbb{Q}^S$ , the Heston framework yields the following dynamics for the process  $S(t)$ :*

$$\begin{aligned} dS(t) &= (r + v(t))S(t)dt + \sqrt{v(t)}S(t)dW_x^S(t), & S(t_0) &= S_0, \\ dv(t) &= \kappa^*(\bar{v}^* - v(t))dt + \gamma\sqrt{v(t)}dW_v^S(t), & v(t_0) &= v_0, \end{aligned}$$

with  $\kappa^* = \kappa - \gamma\rho$ ,  $\bar{v}^* = \kappa\bar{v}/\kappa^*$ , and  $W_x^S(t)$  and  $W_v^S(t)$  are BMs under the underlying process measure  $\mathbb{Q}^S$  such that  $dW_x^S(t)dW_v^S(t) = \rho dt$ .

*Proof.* For a proof, see Appendix A.1. ■

### 3. Swift numerical pricing using deep learning

This section focuses on efficient pricing of discrete arithmetic Asian options in a MC setting. The method uses a Stochastic Collocation [10] (SC) based approach to approximate the target distribution. Then, artificial neural networks (ANNs) “learn” the proxy of the desired distribution, allowing for fast recovery [16, 19].

#### 3.1. “Compressing” distribution with Stochastic Collocation

In the framework of MC methods, the idea of SC – based on the *probability integral transform*<sup>1</sup> – is to approximate the relationship between a “computationally expensive” random variable, say  $A$ , and a “computationally cheap” one, say  $\xi$ . The approximation is then used for sampling. A random variable is “expensive” if its inverse CDF is not known in analytic form,

---

<sup>1</sup>Given the two random variables  $X, Y$ , with CDFs  $F_X, F_Y$ , it holds:  $F_X(X) \stackrel{d}{=} F_Y(Y)$ .

and needs to be computed numerically. With SC, the sampling of  $A$  is performed at the cost of sampling  $\xi$  (see [10]). Formally, the following mapping is used to generate samples from  $A$ :

$$A \stackrel{d}{=} F_A^{-1}(F_\xi(\xi)) =: g(\xi) \approx \tilde{g}(\xi), \quad (3.1)$$

with  $F_A$  and  $F_\xi$  being respectively the CDFs of  $A$  and  $\xi$ , and the function  $\tilde{g}$  a suitable, easily evaluable approximation of  $g$ . The reason why we prefer  $\tilde{g}$  to  $g$ , is that, by definition, every evaluation of  $g$  requires the numerical inversion of  $F_A$ , the CDF of  $A$ .

Many possible choices of  $\tilde{g}$  exist. In [10, 19],  $\tilde{g}$  is an  $(M-1)$ -degree polynomial expressed in Lagrange basis, defined on *collocation points* (CPs)  $\boldsymbol{\xi} := \{\xi_k\}_{k=1}^M$  computed as Gauss-Hermite quadrature nodes, i.e.:

$$\tilde{g}(x) := \sum_{k=1}^M a_k \ell_k(x), \quad \ell_k(x) := \prod_{\substack{1 \leq j \leq M \\ j \neq k}} \frac{x - \xi_j}{\xi_k - \xi_j}, \quad k = 1, \dots, M, \quad (3.2)$$

where the coefficients  $\mathbf{a} := \{a_k\}_{k=1}^M$  of the polynomial in the Lagrange basis representation, called *collocation values* (CVs), are derived by imposing the system of equations:

$$g(\xi_k) = \tilde{g}(\xi_k) =: a_k, \quad k = 1, \dots, M, \quad (3.3)$$

which requires only  $M$  evaluations of  $g$ .

In this work, we define  $\tilde{g}$  as a piecewise polynomial. Particularly, we divide the domain  $\mathbb{R}$  of the random variable  $\xi$  (which for us is standard normally distributed<sup>2</sup>) in three regions. In each region, we define  $\tilde{g}$  as a polynomial. In other words, given the partition of the real line:

$$\Omega_- \cup \Omega_M \cup \Omega_+ := (-\infty, -\bar{\xi}) \cup [-\bar{\xi}, \bar{\xi}] \cup (\bar{\xi}, +\infty), \quad (3.4)$$

for a certain  $\bar{\xi} > 0$ ,  $\tilde{g}$  is specified as:

$$\tilde{g}(\xi) := g_-(\xi) \cdot \mathbb{1}_{\Omega_-}(\xi) + g_M(\xi) \cdot \mathbb{1}_{\Omega_M}(\xi) + g_+(\xi) \cdot \mathbb{1}_{\Omega_+}(\xi), \quad (3.5)$$

where  $\mathbb{1}_{(\cdot)}$  is the indicator function, and  $g_-$ ,  $g_M$ ,  $g_+$  are suitable polynomials. To ensure high accuracy in the approximation,  $g_M$  is defined as a Lagrange polynomial of high-degree  $M-1$ . The CPs  $\boldsymbol{\xi}$ , which identify the Lagrange basis in (3.2), are chosen as Chebyshev nodes in the bounded interval  $\Omega_M = [-\bar{\xi}, \bar{\xi}]$  [7]. Instead, the CVs  $\mathbf{a}$  are defined as in (3.3). The choice of Chebyshev nodes allows to increase the degree of the interpolation (i.e. the number of CPs and CVs), avoiding the *Runge's phenomenon* within the interval  $\Omega_M$  [20]. However, the behaviour of  $g_M$  outside  $\Omega_M$  is out of control. We expect the high-degree polynomial  $g_M$  to be a poor approximation of  $g$  in  $\Omega_-$  and  $\Omega_+$ . Therefore, we define  $g_-$  and  $g_+$  as linear (or at most quadratic) polynomials, with degree  $M_- - 1$  and  $M_+ - 1$ , built on the *extreme* CPs of  $\boldsymbol{\xi}$ .

Summarizing,  $g_-$ ,  $g_M$  and  $g_+$  are all defined as Lagrange polynomials:

$$g_{(\cdot)}(x) := \sum_{k \in I_{(\cdot)}} a_k \ell_k(x), \quad \ell_k(x) := \prod_{\substack{j \in I_{(\cdot)} \\ j \neq k}} \frac{x - \xi_j}{\xi_k - \xi_j}, \quad k \in I_{(\cdot)}, \quad (3.6)$$

where the sets of indexes for  $g_-$ ,  $g_M$  and  $g_+$  are  $I_- = \{1, 2\}$ ,  $I_M = \{1, \dots, M\}$  and  $I_+ = \{M-1, M\}$ , respectively (if a quadratic extrapolation is preferred, we get  $I_- = \{1, 2, 3\}$ , and  $I_+ = \{M-2, M-1, M\}$ ).

**Remark 3.1** (“Compressed” distributions). *The SC technique is a tool to “compress” the information regarding  $A$  in (3.1), into a small number of coefficients, the CVs  $\mathbf{a}$ . Indeed, the relationship between  $A$  and  $\mathbf{a}$  is bijective, provided the distribution of the random variable  $\xi$ , and the corresponding CPs  $\boldsymbol{\xi}$  (or, equivalently, the Lagrange basis in (3.2)), are specified a priori.*

---

<sup>2</sup>The two main reasons of  $\xi$  being standard normal are the availability of such a distribution in most of the computing tools, and the “similarity” between a standard normally r.v. and (the logarithm of)  $A(S)$  (see [10] for more details).

### 3.2. Semi-analytical pricing of fixed- or float-strike Asian options

Let us first consider the pricing of the fixed- or the float-strike Asian options. Both the products allow for the same representation in which the only unknown stochastic quantity is  $A^\lambda(S)$ ,  $\lambda \in \{\text{fx}, \text{fl}\}$ , as given in Proposition 2.2. For the sake of simplicity, in absence of ambiguity, we call  $A^\lambda(S)$  just  $A$ .

For pricing purposes, we can benefit of the SC technique presented in the previous section, provided we know the map  $\tilde{g}$  (or, equivalently, the CVs  $\mathbf{a}$ ), i.e.:

$$\begin{aligned} V_\omega(t_0) &= C \mathbb{E} \left[ \max(\omega(A - K), 0) \right] \\ &\approx C \mathbb{E} \left[ \max(\omega(\tilde{g}(\xi) - K), 0) \right] =: \tilde{V}_\omega(t_0), \end{aligned} \quad (3.7)$$

where  $\xi$  is a standard normally distributed random variable, and  $C$  is a constant coherent with Proposition 2.2. We note,  $\tilde{V}_\omega(t_0)$  is the expectation of (the positive part of) polynomials of a standard normal distribution. Hence, a semi-analytic formula exists, in a similar fashion as the one given in [9].

**Proposition 3.2** (Semi-analytic pricing formula). *Let  $\tilde{V}_\omega(t_0)$  (and  $C$ ) be defined as in Equation (3.7), with  $\tilde{g}$  defined in (3.5). Assume further that  $\alpha_-$ ,  $\alpha_M$  and  $\alpha_+$  are the coefficients in the canonical basis of monomials for the three polynomials  $g_-$ ,  $g_M$  and  $g_+$  respectively of degree  $M_- - 1$ ,  $M - 1$  and  $M_+ - 1$ . Then, using the notation  $a \vee b = \max(a, b)$ , the following semi-analytic pricing approximation holds:*

$$\begin{aligned} \frac{\tilde{V}_\omega(t_0)}{\omega C} &= \left[ \sum_{i=0}^{M_- - 1} \omega^i \alpha_{-\omega, i} m_i(\omega c_K, -\bar{\xi} \vee \omega c_K) \right] \cdot (F_\xi(-\bar{\xi} \vee \omega c_K) - F_\xi(\omega c_K)) \\ &+ \left[ \sum_{i=0}^{M-1} \omega^i \alpha_{M, i} m_i(-\bar{\xi} \vee \omega c_K, \bar{\xi} \vee \omega c_K) \right] \cdot (F_\xi(\bar{\xi} \vee \omega c_K) - F_\xi(-\bar{\xi} \vee \omega c_K)) \\ &+ \left[ \sum_{i=0}^{M_+ - 1} \omega^i \alpha_{\omega, i} m_i(\bar{\xi} \vee \omega c_K, +\infty) \right] \cdot (1 - F_\xi(\bar{\xi} \vee \omega c_K)) - K \cdot (1 - F_\xi(\omega c_K)), \end{aligned}$$

where  $F_\xi$  is the CDF of a standard normal random variable,  $\xi$ ,  $m_i(a, b) := \mathbb{E}[\xi^i | a \leq \xi \leq b]$ <sup>3</sup>,  $c_K$  satisfies  $K = \tilde{g}(c_K)$ , and  $\omega = \pm 1$  according to the call/put case.

*Proof.* For proof of the previous proposition, see Appendix A.2. ■

We note also that Proposition 3.2 uses as input in the pricing formula the coefficients in the canonical basis, not the ones in the Lagrange basis,  $\mathbf{a}$ , in (3.3).

**Remark 3.3** (Change of basis). *Given a Lagrange basis identified by  $M$  collocation points  $\boldsymbol{\xi}$ , any  $(M - 1)$ -degree polynomial  $g(\xi)$  is uniquely determined by the corresponding  $M$  coefficients  $\mathbf{a}$ . A linear transformation exists that connects the  $M$  coefficients in the Lagrange basis with the  $M$  coefficients  $\boldsymbol{\alpha}$  in the canonical basis of monomials. In particular, it holds:*

$$\mathbf{M}\boldsymbol{\alpha} = \mathbf{a},$$

with  $\mathbf{M} \equiv \mathbf{M}(\boldsymbol{\xi})$  a  $M \times M$  Vandermonde matrix with element  $\mathbf{M}_{k, i} := \xi_k^{i-1}$  in position  $(k, i)$ . The matrix  $\mathbf{M}$  admits an inverse; thus, the coefficients  $\boldsymbol{\alpha}$  in the canonical basis are the result of matrix-vector multiplication, provided the coefficients  $\mathbf{a}$  in the Lagrange basis are known. Moreover, since the matrix  $\mathbf{M}$  only depends on  $\boldsymbol{\xi}$ , its inverse can be computed a priori once the CPs  $\boldsymbol{\xi}$  are fixed.

Proposition 3.2 provides a semi-analytic formula for the pricing of fixed- or float-strike Asian options. Indeed, it requires the inversion of the map  $\tilde{g}$  which typically is not available in analytic form. On the other hand, since both the CPs  $\boldsymbol{\xi}$  and the CVs  $\mathbf{a}$  are known, a proxy of  $\tilde{g}^{-1}$  is easily achievable by interpolation on the pairs of values  $(a_k, \xi_k)$ ,  $k = 1, \dots, M$ .

---

<sup>3</sup>A recursive formula for the computation of  $m_i(a, b)$  is given in Appendix A.2.

The last problem is to recover the CVs  $\mathbf{a}$  (which identify  $\tilde{g}$ ) in an accurate and fast way. We recall that, for  $k = 1, \dots, M$ , each CV  $a_k$  is defined in terms of the exact map  $g$  and the CP  $\xi_k$  by the relationship:

$$a_k := g(\xi_k) = F_A^{-1}(F_\xi(\xi_k)).$$

The presence of  $F_A^{-1}$  makes impossible to directly compute  $\mathbf{a}$  in an efficient way. On the other hand, by definition, the CVs  $\mathbf{a}$  are quantiles of the random variable  $A \equiv A(S)$ , which depends on the parameters  $\mathbf{p}$  of the underlying process  $S$ . As a consequence, there must exist some unknown mapping  $H$  which links  $\mathbf{p}$  to the corresponding  $\mathbf{a}$ . We approximate such a mapping from synthetic data setting a regression problem, which is solved with an ANN  $\tilde{H}$  (in the same fashion as in [16, 19]). We have the following mapping:

$$\mathbf{p} \mapsto \mathbf{a} := H(\mathbf{p}) \approx \tilde{H}(\mathbf{p}), \quad \mathbf{p} \in \Omega_{\mathbf{p}}, \quad \mathbf{a} \in \Omega_{\mathbf{a}},$$

with  $\Omega_{\mathbf{p}}$  and  $\Omega_{\mathbf{a}}$  being the spaces of the underlying model parameters and of the CVs, respectively, while the ANN  $\tilde{H}$  is the result of an optimization process on a synthetic training set<sup>4</sup>:

$$\mathbf{T} = \left\{ (\mathbf{p}_i, \mathbf{a}_i) : i \in \{1, \dots, N_{\text{pairs}}\} \right\}. \quad (3.8)$$

The pricing procedure is summarized in the following algorithm.

**Algorithm: Semi-analytic pricing**

1. Fix the  $M$  collocation points  $\xi$ .
2. Given the parameters  $\mathbf{p}$ , approximate the  $M$  collocation values, i.e.  $\mathbf{a} \approx \tilde{H}(\mathbf{p})$ .
3. Given  $\mathbf{a}$ , compute the coefficients  $\alpha_-, \alpha_M$  and  $\alpha_+$  for  $g_-, g_M$  and  $g_+$  (see Remark 3.3).
4. Given  $K$  and  $\mathbf{a}$ , compute  $c_K$  of Proposition 3.2 interpolating  $\tilde{g}^{-1}$  on  $(\mathbf{a}, \xi)$ .
5. Given the coefficients  $\alpha_-, \alpha_M, \alpha_+$ , and  $c_K$ , use Proposition 3.2 to compute  $\tilde{V}_\omega(t_0)$ .

*3.3. Swift Monte Carlo pricing of fixed- and float-strikes Asian options*

Let us consider the case of an option whose payoff has both a fixed- and a float-strike. The present value of such a derivative is given by:

$$V_\omega(t_0) = \frac{M(t_0)}{M(T)} \mathbb{E}_{t_0}^{\mathbb{Q}} \left[ \max(\omega(A(S) - K_1 S(T) - K_2), 0) \right],$$

hence the price of the option is a function of the two dependent quantities  $A(S)$  and  $S(T)$ . This means that, even in a MC setting, the dependency between  $A(S)$  and  $S(T)$  has to be fulfilled. Therefore, a different methodology with respect to the one proposed in the previous section needs to be developed.

Due to the availability of efficient and accurate sampling techniques for the underlying process  $S$  at a given future time  $T$  (we use the COS method [18] enhanced with SC [10], and we call it COS-SC), the main issue is the sampling of the conditional random variable  $A(S)|S(T)$ . This task is addressed in the same fashion as it is done in [19], where ANNs and stochastic collocation are applied for the efficient sampling from time-integral of stochastic bridges<sup>5</sup>, namely  $\int_{t_0}^T S(t)dt$  given the value of  $S(T)$ . The underlying idea here is the same since the random variable  $A(S)$  is conditional to  $S(T)$ . Especially, in the previous sections we pointed out that the distribution of  $A(S)$  has an unknown parametric form which depends on the set of Heston parameters  $\mathbf{p}$ . Similarly, we expect the distribution of  $A|S(T) = \hat{S}$  to be parametric into the “augmented” set of parameters  $\mathbf{p}_{\hat{S}} := \mathbf{p} \cup \{\hat{S}\}$ . Hence, there exists a mapping  $H$  which links  $\mathbf{p}_{\hat{S}}$  with the CVs,  $\mathbf{a}_{\hat{S}}$ , corresponding to the *conditional* distribution  $A(S)|S(T) = \hat{S}$ . We approximate  $H$  by means of a suitable ANN  $\tilde{H}$ , getting the the following mapping scheme:

$$\mathbf{p}_{\hat{S}} \mapsto \mathbf{a}_{\hat{S}} := H(\mathbf{p}_{\hat{S}}) \approx \tilde{H}(\mathbf{p}_{\hat{S}}), \quad \mathbf{p}_{\hat{S}} \in \Omega_{\mathbf{p}_S}, \quad \mathbf{a}_{\hat{S}} \in \Omega_{\mathbf{a}_S},$$

<sup>4</sup>The synthetic data are generated via MC simulation, as explained in Section 5.

<sup>5</sup>By *stochastic bridge*, we mean any stochastic process conditional to both its initial and final values.

where  $\Omega_{\mathbf{p}_S}$  and  $\Omega_{\mathbf{a}_S}$  are respectively the spaces of the underlying model parameters (augmented with  $\hat{S}$ ) and of the CVs (corresponding to the conditional distribution  $A(S)|S(T)$ ), and  $\tilde{H}$  is the result of a regression problem on a suitable training set  $\mathbf{T}$  (see Equation (3.8)). We propose a first *brute force* sampling scheme.

**Algorithm: Brute force conditional sampling and pricing**

1. Fix the  $M$  collocation points  $\boldsymbol{\xi}$ .
2. Given the parameters  $\mathbf{p}$ , for  $j = 1, \dots, N_{\text{paths}}$ , repeat:
  - (a) generate the sample  $\hat{S}_j$  from  $S(T)$  (e.g. with COS-SC method [18] and [10]);
  - (b) given  $\mathbf{p}_{\hat{S}_j}$ , approximate the  $M$  conditional CVs, i.e.  $\mathbf{a}_{\hat{S}_j} \approx \tilde{H}(\mathbf{p}_{\hat{S}_j})$ ;
  - (c) given  $\mathbf{a}_{\hat{S}_j}$ , use SC to generate the conditional sample  $\hat{A}_j$ .
3. Given the pairs  $(\hat{S}_j, \hat{A}_j)$ , for  $j = 1, \dots, N_{\text{paths}}$ , and any desired  $(K_1, K_2)$ , evaluate:

$$V_\omega(t_0) \approx \frac{1}{N_{\text{paths}}} \frac{M(t_0)}{M(T)} \sum_{j=1}^{N_{\text{paths}}} \max(\omega(\hat{A}_j - K_1 \hat{S}_j - K_2), 0).$$

Nonetheless, the brute force sampling proposed above requires  $N_{\text{paths}}$  evaluations of  $\tilde{H}$  (see 2(b) in the previous algorithm). This is a massive computational cost, even if a single evaluation of an ANN is high-speed. We can, however, benefit of a further approximation. We compute the CVs using  $\tilde{H}$  only at specific *reference* values for  $S(T)$ . Then, the intermediate cases are derived utilizing (linear) interpolation. We choose a set of  $Q$  equally-spaced values  $\{S^1, S^2, \dots, S^Q\}$  for  $S(T)$ , defined as:

$$S^q := S_{\min} + \frac{q-1}{Q-1}(S_{\max} - S_{\min}), \quad q = 1, \dots, Q, \quad (3.9)$$

where the boundaries are quantiles corresponding to the probabilities  $p_{\min}, p_{\max} \in (0, 1)$ , i.e.  $S_{\min} := F_{S(T)}^{-1}(p_{\min})$  and  $S_{\max} := F_{S(T)}^{-1}(p_{\max})$ .

Calling  $\mathbf{p}^q = \mathbf{p}_{S^q}$ , and  $\mathbf{a}^q = \mathbf{a}_{S^q}$ ,  $q = 1, \dots, Q$ , we compute the grid  $\mathbf{G}$  of *reference* CPs, with only  $Q$  ANN evaluations, namely:

$$\begin{bmatrix} \mathbf{a}^1 \\ \dots \\ \mathbf{a}^q \\ \dots \\ \mathbf{a}^Q \end{bmatrix} = \begin{bmatrix} H(\mathbf{p}^1) \\ \dots \\ H(\mathbf{p}^q) \\ \dots \\ H(\mathbf{p}^Q) \end{bmatrix} \approx \begin{bmatrix} \tilde{H}(\mathbf{p}^1) \\ \dots \\ \tilde{H}(\mathbf{p}^q) \\ \dots \\ \tilde{H}(\mathbf{p}^Q) \end{bmatrix} =: \mathbf{G}, \quad (3.10)$$

where  $\mathbf{a}^q$ ,  $H(\mathbf{p}^q)$  and  $\tilde{H}(\mathbf{p}^q)$ ,  $q = 1, \dots, Q$ , are row vectors. The interpolation on  $\mathbf{G}$  is much faster than the evaluation of  $\tilde{H}$ . Therefore, the grid-based conditional sampling results more efficient than the brute force one, particularly when sampling a huge number of MC samples.

The algorithm for the grid-based sampling procedure, to be used instead of point 2. in the previous algorithm, is reported here.

**Algorithm: Grid-based conditional sampling**

- 2.1. Fix the *boundary* probabilities  $p_{\min}, p_{\max} \in (0, 1)$  and compute the *boundary* quantiles  $S_{\min} := F_{S(T)}^{-1}(p_{\min})$  and  $S_{\max} := F_{S(T)}^{-1}(p_{\max})$  (e.g. with the COS method [18]).
- 2.2. Compute the *reference* values  $S^q := S_{\min} + \frac{q-1}{Q-1}(S_{\max} - S_{\min})$ ,  $q = 1, \dots, Q$ .
- 2.3. Given the “augmented” parameters  $\mathbf{p}^q$ , evaluate  $Q$  times  $\tilde{H}$  to compute  $\mathbf{G}$  (see (3.10)).
- 2.4. Given the parameters  $\mathbf{p}$  and the grid  $\mathbf{G}$ , for  $j = 1, \dots, N_{\text{paths}}$ , repeat:
  - (a) generate the sample  $\hat{S}_j$  from  $S(T)$  (e.g. with COS-SC method [18] and [10]);
  - (b) given  $\hat{S}_j$ , approximate the  $M$  conditional CVs, i.e.  $\mathbf{a}_{\hat{S}_j}$ , by interpolation in  $\mathbf{G}$ ;
  - (c) given  $\mathbf{a}_{\hat{S}_j}$ , use SC to generate the conditional sample  $\hat{A}_j$ .



## 4. Error analysis

This section is dedicated to the assessment and discussion of the error introduced by the main approximations used in the proposed pricing method. Two primary sources of error are identifiable. The first error is due to the SC technique: in Section 3.1 the exact map  $g$  is approximated by means of the piecewise polynomial  $\tilde{g}$ . The second one is a regression error, which is present in both Sections 3.2 and 3.3. ANNs  $\tilde{H}$  are used instead of the exact mappings  $H$ . For the error introduced by the SC technique, we bound the “ $L^2$ -distance”,  $\epsilon_{SC}$ , between the exact distribution and its SC proxy showing that  $\tilde{g}\mathbb{1}_{\Omega_M} = g_M$  in  $\Omega_M$  is an analytic function.  $\epsilon_{SC}$  is used to provide a direct bound on the option price error,  $\epsilon_P$ . On the other hand, regarding the approximation of  $H$  via  $\tilde{H}$  we provide a general convergence result for ReLU-architecture ANN, i.e. ANN with Rectified Linear Units as activation functions.

### 4.1. Stochastic collocation error using Chebyshev polynomials

Let us consider the error introduced in the methodology using the SC technique (Section 3.1), and investigate how this affects the option price. We restrict the analysis to the case of fixed- or float-strike discrete arithmetic Asian and Lookback options (Section 3.2). We define the error  $\epsilon_P$  as the “ $L^1$ -distance” between the real price  $V_\omega(t_0)$  and its approximation  $\tilde{V}_\omega(t_0)$ , i.e.:

$$\epsilon_P := |\tilde{V}_\omega(t_0) - V_\omega(t_0)|. \quad (4.1)$$

Given the standard normal kernel  $\xi \sim \mathcal{N}(0, 1)$ , we define the SC error as the (squared)  $L^2$ -norm of  $g - \tilde{g}$ , i.e.:

$$\epsilon_{SC} := \mathbb{E}[(g - \tilde{g})^2(\xi)]. \quad (4.2)$$

We decompose  $\epsilon_{SC}$  accordingly to the piecewise definition of  $\tilde{g}$ , namely:

$$\begin{aligned} \epsilon_{SC} &= \mathbb{E}[(g\mathbb{1}_{\Omega_-} - g_-)^2(\xi)] + \mathbb{E}[(g\mathbb{1}_{\Omega_M} - g_M)^2(\xi)] + \mathbb{E}[(g\mathbb{1}_{\Omega_+} - g_+)^2(\xi)] \\ &=: \epsilon_- + \epsilon_M + \epsilon_+, \end{aligned}$$

with the domains  $\Omega_{(\cdot)}$  defined in Equation (3.4), i.e., for  $\bar{\xi} > 0$ :

$$\Omega_- = (-\infty, -\bar{\xi}), \quad \Omega_M = [-\bar{\xi}, \bar{\xi}], \quad \Omega_+ = (\bar{\xi}, +\infty).$$

To deal with the “extrapolation” errors  $\epsilon_-$  and  $\epsilon_+$ , we formulate the following assumption.

**Assumption 4.1.** *The functions  $(g\mathbb{1}_{\Omega_-} - g_-)^2$  and  $(g\mathbb{1}_{\Omega_+} - g_+)^2$  are  $O(\exp x^2/2)$ . Equivalently,  $g^2\mathbb{1}_{\Omega_-}$  and  $g^2\mathbb{1}_{\Omega_+}$  are  $O(\exp x^2/2)$  (since  $g_-$  and  $g_+$  are polynomials).*

Given Assumption 4.1 and the fact that  $\xi \sim \mathcal{N}(0, 1)$ <sup>6</sup>, then the “extrapolation” errors  $\epsilon_-$  and  $\epsilon_+$  vanish, with exponential rate, as  $\bar{\xi}$  tends to infinity, i.e.  $\epsilon_- = \epsilon_-(\bar{\xi})$ ,  $\epsilon_+ = \epsilon_+(\bar{\xi})$ , and:

$$\begin{cases} \epsilon_-(\bar{\xi}) \rightarrow 0, \\ \epsilon_+(\bar{\xi}) \rightarrow 0, \end{cases} \quad \text{for } \bar{\xi} \rightarrow +\infty. \quad (4.3)$$

An illustration of the speed of convergence is reported in Figure 1. Figure 1a shows that the growth of  $g\mathbb{1}_{\Omega_+}$  is (much) less than exponential (consistently with Assumption 4.1), whereas Figure 1b illustrates the exponential decay of  $\epsilon_-$  and  $\epsilon_+$  when  $\bar{\xi}$  increases.

Therefore, if  $\bar{\xi}$  is taken sufficiently big, the error  $\epsilon_{SC}$  in (4.2) is mainly driven by the “interpolation” error  $\epsilon_M$ , whose estimate is connected to error bounds for Chebyshev polynomial interpolation, and it is the focus of the next part.

**Theorem 4.1** (Error bound for analytic function [7, 23]). *Let  $f$  be a real function on  $[-1, 1]$  and  $f_M$  be its  $(M-1)$ -degree polynomial interpolation built on Chebyshev nodes  $\xi_k := \cos \frac{k-1}{M-1}\pi$ ,  $k = 1, \dots, M$ . If  $f$  has an analytic extension in a Bernstein ellipse  $\mathcal{B}$  with foci  $\pm 1$  and major and minor semiaxis lengths summing up to  $\varrho > 1$  such that  $\sup_{\mathcal{B}} |f| \leq \frac{\varrho-1}{4}\bar{C}$  for some constant  $\bar{C} > 0$ , then, for each  $M \geq 1$ , the following bound holds:*

$$\|f - f_M\|_{L^\infty([-1,1])} \leq \bar{C}\varrho^{1-M}.$$

---

<sup>6</sup>The PDF  $f_\xi$  works as a dumping factor in  $\epsilon_- = \mathbb{E}[(g\mathbb{1}_{\Omega_-} - g_-)^2(\xi)]$  and  $\epsilon_+ = \mathbb{E}[(g\mathbb{1}_{\Omega_+} - g_+)^2(\xi)]$ .

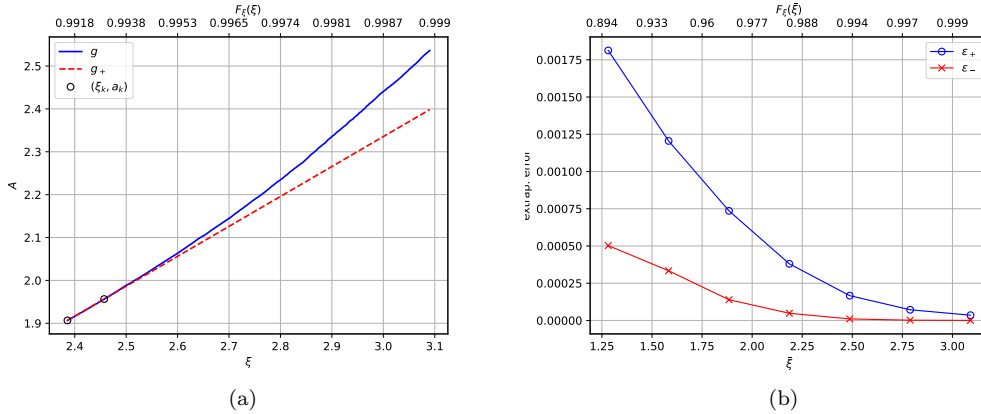


Figure 1: Left: (slow) growth of  $g\mathbb{1}_{\Omega_+}$  compared to the linear extrapolation  $g_+$ . Right: exponential decays of  $\epsilon_-(\bar{\xi})$  and  $\epsilon_+(\bar{\xi})$  in (4.3) when the  $\bar{\xi}$  increases. The upper x-axis represents the probability  $F_\xi(\bar{\xi})$ .

Since  $g\mathbb{1}_{\Omega_M}$  is approximated by means of the  $(M - 1)$ -degree polynomials  $g_M$ , built on Chebyshev nodes, to apply Theorem 4.1, we verify the required assumptions, namely the boundedness of  $g\mathbb{1}_{\Omega_M}$  in  $\Omega_M$  and its analyticity. We recall that:

$$g = F_A^{-1} \circ F_\xi, \quad (4.4)$$

with  $F_A$  and  $F_\xi$  the CDFs of  $A(S)$  and  $\xi$ , respectively. Hence, the boundedness on the compact domain  $\Omega_M$  is satisfied because the map  $g$  is monotone increasing (as composition of monotone increasing functions), and defined everywhere in  $\Omega_M$ .

Furthermore, since the CDF of a standard normal,  $F_\xi$ , is analytic, from (4.4) follows that  $g$  is analytic if  $F_A^{-1}$  is analytic. The analyticity of  $F_A^{-1}$  is fulfilled if  $F_A$  is analytic and  $F'_A = f_A$  does not vanish in the domain  $\Omega_M$ . Observe that, by restricting the domain to  $\Omega_M$ , the latter condition is trivially satisfied because we are “far” from the tails of  $A(S)$  (corresponding to the extrapolation domains  $\Omega_-$  and  $\Omega_+$ ), and  $F'_A$  do not vanish in other regions than the tails.

On the contrary, proving that  $F_A$  is analytic is not trivial because of the lack of an explicit formula for  $F_A$ . However, it is beneficial to represent  $F_A$  through the *characteristic function* (ChF) of  $A(S)$ ,  $\phi_A$ . For that purpose, we use a well-known inversion result.

**Theorem 4.2** (ChF Inversion Theorem). *Let us denote by  $F$  and  $\phi$  the CDF and the ChF of a given real-valued random variable defined on  $\mathbb{R}$ . Then, it is possible to retrieve  $F$  from  $\phi$  according to the inversion formula:*

$$F(x) - F(0) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \phi(u) \frac{1 - e^{-iux}}{iu} du,$$

with the integral being understood as a principal value.

*Proof.* For detailed proof, we refer to [14]. ■

Thanks to Theorem 4.2, we have that if  $\phi_A$  is analytic, so it is  $F_A$  (as long as the integral in (4.2) is well defined). Thus, the problem becomes to determine if  $\phi_A$  is analytic. We rely on a characterization of *entire*<sup>7</sup> ChFs, which can be used in this framework to show that in the cases of – fixed- or float-strike – discrete arithmetic Asian and Lookback options, the (complex extension of the) function  $\phi_A$  is analytic in a certain domain.

**Theorem 4.3** (Characterization of entire ChFs [3]). *Let  $Y$  be a real random variable. Then, the complex function  $\phi(z) := \mathbb{E}[e^{izY}]$ ,  $z \in \mathbb{C}$ , is entire if and only if the absolute moments of  $Y$  exist for any order, i.e.  $\mathbb{E}[|Y|^k] < +\infty$  for any  $k \in \mathbb{N}$ , and the following limit holds:*

$$\lim_{k \rightarrow +\infty} \left( \frac{|\mathbb{E}[Y^k]|}{k!} \right)^{\frac{1}{k}} = 0. \quad (4.5)$$

<sup>7</sup>Entire functions are *complex analytic* functions in the whole complex plan  $\mathbb{C}$ .

*Proof.* A reference for proof is given in [3]. ■

When dealing with the Heston model, there is no closed-form expression for the moments of the underlying process  $S(t)$ , as well as for the moments of its transform  $A(S)$ . Nonetheless, a *conditional* case can be studied and employed to provide a starting point for a convergence result.

**Proposition 4.4** (Conditional ChF  $\phi_{A|\mathbf{V}}$  is entire). *Let us define the  $N$ -dimensional random vector  $\mathbf{V}$ , with values in  $\Omega_{\mathbf{V}} := \mathbb{R}_+^N$ , as:*

$$\mathbf{V} := [I_v(t_1), I_v(t_2), \dots, I_v(t_N)]^T, \quad I_v(t_n) := \int_{t_0}^{t_n} v(\tau) d\tau, \quad n = 1, \dots, N.$$

*Let the complex conditional characteristic function  $\phi_{A|\mathbf{V}}(z) := \mathbb{E}[e^{izA}|\mathbf{V}]$ ,  $z \in \mathbb{C}$ , be the extended ChF of the conditional random variable  $A|\mathbf{V}$ , with  $A \equiv A(S)$  as given in Equation (2.2).*

*Then,  $\phi_{A|\mathbf{V}}(z)$  is entire.*

*Proof.* See Appendix A.4. ■

From now on, using the notation of Proposition 4.4, we consider satisfied the following assumption on the tail behaviour of the random vector  $(\mathbf{V}, A)$ . Informally, we require that the density of the joint distribution of  $(\mathbf{V}, A)$  has uniform (w.r.t.  $\mathbf{V}$ ) exponential decay for  $A$  going to  $+\infty$ .

**Assumption 4.2.** *There exists a point  $z^* \in \mathbb{C}$ ,  $z^* = x^* - iy^*$ , with  $x^*, y^* \in \mathbb{R}$  and  $y^* > 0$ , such that:*

$$\int_{\Omega_{\mathbf{V}} \times \mathbb{R}^+} e^{y^* a} dF_{\mathbf{V}, A}(\mathbf{v}, a) < +\infty,$$

*with  $F_{\mathbf{V}, A}(\cdot, \cdot)$  the joint distribution of the random vector  $\mathbf{V}$  and the random variable  $A$ .*

Thanks to Assumption 4.2, the ChF  $\phi_A(z)$  is well defined for any  $z \in \mathcal{S}_{y^*} \subset \mathbb{C}$ , with the strip  $\mathcal{S}_{y^*} := \mathbb{R} + i \cdot [-y^*, y^*]$ . Moreover, applying Fubini's Theorem, for any  $z \in \mathcal{S}_{y^*}$ , we have:

$$\phi_A(z) = \int_{\Omega_{\mathbf{V}}} \phi_{A|\mathbf{V}=\mathbf{v}}(z) dF_{\mathbf{V}}(\mathbf{v}). \quad (4.6)$$

Thus, we can show that the ChF  $\phi_A(z)$  is analytic in the strip  $\mathcal{S}_{y^*}$  (the details are given in Appendix A.2).

**Proposition 4.5** (ChF  $\phi_A$  is analytic). *Let  $\phi_A(z) := \mathbb{E}[e^{izA}]$ ,  $z \in \mathcal{S}_{y^*}$ , with  $A \equiv A(S)$ . Then,  $\phi_A(z)$  is analytic in  $\mathcal{S}_{y^*}$ .*

*Proof.* A proof is given in Appendix A.2. ■

Thanks to Proposition 4.5, and consistently with the previous discussion, we conclude that the map  $g$  in (4.4), is analytic on the domain  $\Omega_M$ . Therefore, we can apply Theorem 4.1, which yields the following error estimate:

$$\|g \mathbb{1}_{\Omega_M} - g_M\|_{L^\infty(\Omega_M)} \leq \bar{C} \rho^{1-M},$$

for certain  $\rho > 1$  and  $\bar{C} > 0$ . As a consequence, the following bound for the  $L^2$ -error  $\epsilon_M$  holds:

$$\epsilon_M = \mathbb{E}[(g \mathbb{1}_{\Omega_M} - g_M)^2(\xi)] \leq \bar{C}^2 \rho^{2-2M}. \quad (4.7)$$

Furthermore, the exponential convergence is also confirmed numerically, as reported in Figure 2. In Figure 2a we can appreciate the improvement in the approximation of  $\tilde{g} \mathbb{1}_{\Omega_M}$  by means of  $g_M$ , when  $M$  is increased, whereas Figure 2b reports the exponential decay of  $\epsilon_M$ .

Using (4.7), the  $L^2$ -norm of  $g - \tilde{g}$ ,  $\epsilon_{SC}$  in (4.2), is bounded by:

$$\epsilon_{SC} \equiv \epsilon_{SC}(\bar{\xi}, M) = \mathbb{E}[(g(\xi) - \tilde{g}(\xi))^2] \leq \epsilon_-(\bar{\xi}) + \bar{C}^2 \rho^{2-2M} + \epsilon_+(\bar{\xi}),$$

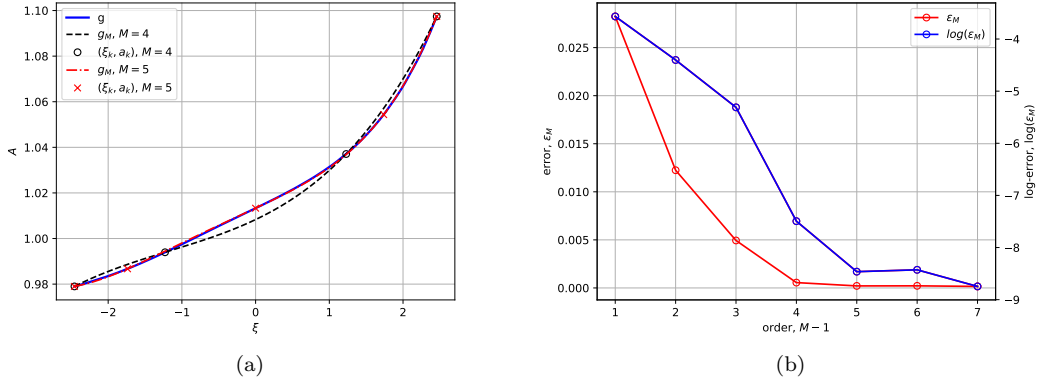


Figure 2: Left: exact map  $g\mathbb{1}_{\Omega_M}$  (blue) compared to the interpolation  $g_M$  in the domain  $\Omega_M$ , for  $M = 4, 5$ . Right:  $L^2$ -error  $\epsilon_M$  in (4.7) exponential decay in the order of the polynomial  $g_M$ .

which goes to zero when  $\bar{\xi} \in \mathbb{R}^+$  and  $M \in \mathbb{N}$  tend to  $+\infty$ . Therefore, for any  $\epsilon > 0$  there exist  $\bar{\xi}^* \in \mathbb{R}^+$  and  $M^* \in \mathbb{N}$  such that:

$$\epsilon_{SC}(\bar{\xi}^*, M^*) < \epsilon^2, \quad (4.8)$$

and because of the exponential decay, we expect  $\bar{\xi}^*$  and  $M^*$  do not need to be taken too big.

Eventually, we can benefit of the bound in (4.8) to control the pricing error,  $\epsilon_P$  in (4.1). By employing the well-known inequality  $\max(a + b, 0) \leq \max(a, 0) + \max(b, 0)$  and the Cauchy-Schwarz inequality, we can write:

$$\begin{aligned} \tilde{V}_\omega(t_0) &= \mathbb{E} \left[ \left( \omega(\tilde{g}(\xi) - K) \right)^+ \right] \leq \sqrt{\mathbb{E} \left[ (\tilde{g}(\xi) - g(\xi))^2 \right]} + \mathbb{E} \left[ \left( \omega(g(\xi) - K) \right)^+ \right] \\ &\leq \sqrt{\epsilon_{SC}(\bar{\xi}, M)} + V_\omega(t_0), \end{aligned}$$

and using the same argument twice (exchanging the roles of  $g$  and  $\tilde{g}$ ), we end up with the following bound for the option price error:

$$\epsilon_P \leq \sqrt{\epsilon_{SC}(\bar{\xi}^*, M^*)} \leq \epsilon,$$

with  $\bar{\xi}^*$  and  $M^*$  as in (4.8).

#### 4.2. Artificial neural network regression error

As the final part of the error analysis, we investigate when ANNs are suitable approximating maps. In particular, we focus on ANNs with ReLU-architectures, namely ANNs whose activation units are all Rectified Linear Units defined as  $\phi(x) = x\mathbb{1}_{x>0}(x)$ .

Consider the Sobolev space  $(\mathcal{W}^{n,\infty}([0, 1]^d), \|\cdot\|_{n,d}^\infty)$ , with  $n, d \in \mathbb{N} \setminus \{0\}$ , namely the space of functions  $C^{n-1}([0, 1]^d)$  whose derivatives up to the  $(n-1)$ -th order are all Lipschitz continuous, equipped with the norm  $\|\cdot\|_{n,d}^\infty$  defined as:

$$\|f\|_{n,d}^\infty = \max_{|\mathbf{n}| \leq n} \text{ess sup}_{\mathbf{x} \in [0,1]^d} |D^{\mathbf{n}} f(\mathbf{x})|,$$

with  $\mathbf{n} := (n_1, \dots, n_d) \in \mathbb{N}^d$ ,  $|\mathbf{n}| = \sum_{i=1}^d n_i$ , and  $D^{\mathbf{n}}$  the weak derivative operator. Furthermore, we define the unit ball  $B_{n,d} := \{f \in \mathcal{W}^{n,\infty}([0, 1]^d) : \|f\|_{n,d}^\infty \leq 1\}$ . Then, the following approximation result holds:

**Theorem 4.6** (Convergence for ReLU ANN). *For any choice of  $d, n \in \mathbb{N} \setminus \{0\}$  and  $\epsilon \in (0, 1)$ , there exists an architecture  $H(\mathbf{x}|\cdot)$  based on ReLU (Rectified Linear Unit) activation functions  $\phi$ , i.e.  $\phi(x) = x\mathbb{1}_{x>0}(x)$ , such that:*

1.  $H(x|\cdot)$  is able to approximate any function  $f \in B_{n,d}$  with an error smaller than  $\epsilon$ , i.e., there exists a matrix of weights  $W$  such that  $\|f(\cdot) - H(\cdot|W)\|_{n,d}^\infty < \epsilon$ ;
2.  $H$  has at most  $c(\ln 1/\epsilon + 1)$  layers and at most  $c\epsilon^{-d/n}(\ln 1/\epsilon + 1)$  weights and neurons, with  $c = c(d, n)$  an appropriate constant depending on  $d$  and  $n$ .

*Proof.* A proof is available in [24]. ■

Substantially, Theorem 4.6 states that there always exists a ReLU-architecture (with finite number of layers and activation units) suitable to approximate at any desired precision functions with a certain level of regularity (determined by  $(W^{n,\infty}([0, 1]^d), \|\cdot\|_{n,d}^\infty)$ ).

**Remark 4.7** (Input scaling). *We emphasize that although Theorem 4.6 applies to (a subclass of sufficiently regular) functions with domain the  $d$ -dimensional hypercube  $[0, 1]^d$ , this is not restrictive. Indeed, as long as the regularity conditions are fulfilled, Theorem 4.6 holds for any function defined on a  $d$ -dimensional hyperrectangle since it is always possible to linearly map its domain into the  $d$ -dimensional hypercube.*

Furthermore, we observe that all the results of convergence for ANN rely on the assumption that the training is performed successfully, and so the final error in the optimization process is negligible. Under this assumption, Theorem 4.6 provides a robust theoretical justification for using ReLU-based ANNs as regressors. The goodness of the result can also be investigated empirically, as we will show in the next section (see, for instance, Figure 4).

## 5. Numerical experiments

In this part of the paper, we detail some numerical experiments. We focus on applying the methodology given in Section 3.2 for the numerical pricing of fixed-strike discrete arithmetic Asian and Lookback options. We address the general case of discrete arithmetic Asian options described in Section 3.3. For each pricing experiment errors and timing results are given. The pricing errors are reported in basis points (bps) of the underlying process initial value, and the timing results are given in seconds. The benchmarks are computed via MC using the *almost exact simulation* of the Heston models, detailed in Result Appendix A.3.

All the computation are implemented and run on a MacBook Air (M1, 2020) machine, with chip Apple M1 and 16 GB of RAM. The code is written in Python, and torch is the library used for the design and training of the ANN, as in [19].

### 5.1. Experiments' specifications

Among the three examples of applications presented, two of them rely on the technique given in Section 3.2, while the third is based on the theory in Section 3.3. The first experiment is the pricing of *fixed-strike discrete arithmetic Asian options* (FxA) with underlying a stock price process following the Heston dynamics. The second example, instead, is connected to the “interest rate world”, and is employed for the pricing of *fixed-strike discrete Lookback swaptions* (FxL). We assume the underlying swap rate is driven by a *displaced* Heston model with driftless dynamics, typically used for interest rates. The last one is an application to the pricing of *fixed- and float-strikes discrete arithmetic Asian options* (FxFIA) on a stock price driven by the Heston dynamics. In the first (FxA) and last experiment (FxFIA),  $A(S)$  in (2.2) is specified as:

$$A(S) = \frac{1}{5} \sum_{n=1}^5 S(t_n), \quad t_n := T - (5 - n)\tau_A, \quad n = 1, \dots, 5, \quad (5.1)$$

with  $\tau_A = \frac{1}{12}$  as monitoring time lag, and  $T > 4\tau_A$  as option maturity. Differently, in the second experiment (FxL)  $A(S)$  is given by:

$$A(S) = \min_n S(t_n), \quad t_n := T - (30 - n)\tau_L, \quad n = 1, \dots, 30, \quad (5.2)$$

with  $\tau_L = \frac{1}{120}$  as monitoring time lag, and  $T > 29\tau_L$  as option maturity. Observe that, assuming the unit is 1 year, with 12 identical months and 360 days, the choices of  $\tau_A$  and  $\tau_L$  correspond respectively to 1 month and 3 days of time lag in the monitoring dates.

## 5.2. Artificial neural network development

In this section we provide the details about the generation of the training set, for each experiment, and the consequent training of the ANN used in the pricing model.

### 5.2.1. Training set generation

The training sets are generated through MC simulations, using the almost exact sampling in Result Appendix A.3. In the first two applications (FxA and FxL) the two training sets are defined as in (3.8), and particularly they read:

$$\begin{aligned}\mathbf{T}_{\text{FxA}} &= \left\{ \left( \{r, \kappa, \gamma, \rho, \bar{v}, v_0, T\}_i, \{a_1, \dots, a_{21}\}_i \right) : i \in \{1, \dots, N_{\text{pairs}}^{\text{FxA}}\} \right\}, \\ \mathbf{T}_{\text{FxL}} &= \left\{ \left( \{\kappa, \gamma, \rho, \bar{v}, v_0, T\}_i, \{a_1, \dots, a_{21}\}_i \right) : i \in \{1, \dots, N_{\text{pairs}}^{\text{FxL}}\} \right\}.\end{aligned}$$

The Heston parameters, i.e.  $\mathbf{p} \setminus \{T\} - \{r, \kappa, \gamma, \rho, \bar{v}, v_0\}$  and  $\{\kappa, \gamma, \rho, \bar{v}, v_0\}$  for FxA and FxL, respectively – are sampled using *Latin Hypercube Sampling* (LHS), to ensure the best filling of the parameter space [16, 19]. For each set  $\mathbf{p} \setminus \{T\}$ ,  $N_{\text{paths}}$  paths are generated, with a time step of  $\Delta t$  and a time horizon up to  $T_{\text{max}}$ . The underlying process  $S$  is monitored at each time  $T$  for which there are enough past observations to compute  $A(S)$ , i.e.:

$$\begin{aligned}T &\geq 4\tau_A + \Delta t, & \text{for FxA,} \\ T &\geq 29\tau_L + \Delta t, & \text{for FxL.}\end{aligned}$$

Consequently, the product between the number of Heston parameters' set and the number of available maturities determines the magnitude of the two training sets (i.e.,  $N_{\text{pairs}}^{\text{FxA}}$  and  $N_{\text{pairs}}^{\text{FxL}}$ ).

For each  $\mathbf{p}$ , the CVs  $\mathbf{a}$  corresponding to  $A(S)$  are computed as:

$$a_k := F_A^{-1}(F_\xi(\xi_k)) \approx Q_A(F_\xi(\xi_k)), \quad k \in \{1, \dots, 21\},$$

where  $Q_A$  is the empirical quantile function of  $A(S)$ , used as numerical proxy of  $F_A^{-1}$ , and  $\xi_k$  are the CPs computed as Chebyshev nodes:

$$\xi_k := -\bar{\xi} \cos \frac{k-1}{20} \pi, \quad k \in \{1, \dots, 21\},$$

with  $\bar{\xi} := F_\xi^{-1}(0.993) \approx 2.46$ . We note that, the definition of  $\bar{\xi}$  avoid any CV  $a_k$  to be “deeply” in the tails of  $A(S)$ , which are more sensitive to numerical instability in a MC simulation.

The information about the generation of the two training sets are reported in Table 1. Observe that  $\mathbf{T}_{\text{FxA}}$  is richer in elements than  $\mathbf{T}_{\text{FxL}}$  because of computational constraints. Indeed, the higher number of monitoring dates of  $A(S)$  in FxL makes the generation time of  $\mathbf{T}_{\text{FxL}}$  more than twice the one of  $\mathbf{T}_{\text{FxA}}$  (given the same number of pairs).

Since in the general procedure (see Section 3.3) ANNs are used to learn the *conditional distribution*  $A(S)|S(T)$  (not just  $A(S)$ !), the third experiment requires a training set which contains also information about the conditioning value,  $S(T)$ . We define  $\mathbf{T}_{\text{FxF1A}}$  as:

$$\mathbf{T}_{\text{FxF1A}} = \left\{ \left( \{r, \kappa, \gamma, \rho, \bar{v}, v_0, T, S^q, p^q\}_i, \{a_1, \dots, a_{14}\}_i \right) : i \in \{1, \dots, N_{\text{pairs}}\} \right\},$$

Table 1: Training sets  $\mathbf{T}_{\text{FxA}}$  and  $\mathbf{T}_{\text{FxL}}$  generation details.

$\mathbf{p}$	FxA			FxL			MC		
	$N_{\text{val}}$	range	met.	$N_{\text{val}}$	range	met.	$N_{\text{paths}}$	$\Delta t$	
$r$	700	[0.00, 0.05]	LHS				$10^6$		
$\kappa$	700	[0.20, 1.10]	LHS	300	[0.80, 1.60]	LHS		$\frac{1}{120}$	
$\gamma$	700	[0.80, 1.10]	LHS	300	[0.40, 1.00]	LHS			
$\rho$	700	[-0.95, -0.20]	LHS	300	[-0.80, -0.30]	LHS			
$\bar{v}$	700	[0.02, 0.15]	LHS	300	[0.10, 0.20]	LHS			
$v_0$	700	[0.02, 0.15]	LHS	300	[0.10, 0.20]	LHS			
$T$	160	[0.34, 1.67]	EQ-SP	170	[0.26, 1.67]	EQ-SP			
$N_{\text{pairs}}$	112000 (700 × 160)			51000 (300 × 170)					
							SC		
							$M$	21	
							$\bar{\xi}$	2.46	
							$F_\xi(\bar{\xi})$	0.993	

Table 2: Training set  $\mathbf{T}_{\text{FxFLA}}$  generation details.

$\mathbf{p}$	$N_{\text{val}}$	range	met.	MC	
				$N_{\text{tot}}$	$N_{\text{paths}}$
$r$	180	[0.00, 0.05]	LHS	$3 \cdot 10^6$	
$\kappa$	180	[0.20, 1.10]	LHS	$2 \cdot 10^5$	
$\gamma$	180	[0.80, 1.10]	LHS	$\Delta t$	$\frac{1}{120}$
$\rho$	180	[-0.92, -0.28]	LHS		
$\bar{v}$	180	[0.03, 0.10]	LHS	SC	
$v_0$	180	[0.03, 0.10]	LHS	$Q$	15
$T$	160	[0.34, 1.67]	EQ-SP	$M$	14
$S^q$	15	[0.35, 2.01]	EQ-SP	$\bar{\xi}$	2.46
$p^q$	15	[0.05, 0.85]	IMPL	$F_{\xi}(\bar{\xi})$	0.993
$N_{\text{pairs}}$	432000 (180 × 160 × 15)				

where  $p^q$  is the probability corresponding to the quantile  $S^q$ , given as in (3.9), i.e.:

$$S^q = S_{\min} + \frac{q-1}{14}(S_{\max} - S_{\min}), \quad q \in \{1, \dots, 15\},$$

with the  $S_{\min} = F_{S(T)}^{-1}(p_{\min})$  and  $S_{\max} = F_{S(T)}^{-1}(p_{\max})$ . Heuristic arguments drove the choice of adding in the input set  $\mathbf{p}$  the probability  $p^q := F_{S(T)}(S^q)$ , i.e. the probability *implied* by the final value  $p^q$ . Indeed, the ANN training process results more accurate when both  $S^q$  and  $p^q$  are included in  $\mathbf{p}$ .

Similarly as before, the sets of Heston parameters are sampled using LHS. For each set,  $N_{\text{tot}}$  paths are generated, with a time step of  $\Delta t$  and a time horizon up to  $T_{\max}$ . The underlying process  $S$  is monitored at each time  $T$  for which there are enough past observations to compute  $A(S)$ , i.e.:

$$T \geq 4\tau_A + \Delta t.$$

For any maturity  $T$  and any realization  $S^q$ , the inverse CDF of the conditional random variable  $A(S)|S(T) = S^q$  is approximated with the empirical quantile function,  $Q_{A|S^q}$ . The quantile function  $Q_{A|S^q}$  is built on the  $N_{\text{paths}}$  “closest” paths to  $S^q$ , i.e. those  $N_{\text{paths}}$  paths whose final values  $S(T)$  are the closest to  $S^q$ .

Eventually, for any input set  $\mathbf{p} = \{r, \kappa, \gamma, \rho, \bar{v}, v_0, T, S^q, p^q\}$ , the CVs  $\mathbf{a}$  corresponding to  $A(S)|S(T) = S^q$  are computed as:

$$a_k := F_{A|S^q}^{-1}(F_{\xi}(\xi_k)) \approx Q_{A|S^q}(F_{\xi}(\xi_k)), \quad k \in \{1, \dots, 14\},$$

with  $\xi_k$  the Chebyshev nodes:

$$\xi_k := -\bar{\xi} \cos \frac{k-1}{13}\pi, \quad k \in \{1, \dots, 14\},$$

and  $\bar{\xi} := F_{\xi}^{-1}(0.993) \approx 2.46$ .

The information about the generation of the training set  $\mathbf{T}_{\text{FxFLA}}$  are reported in Table 2.

### 5.2.2. Artificial neural network training

Each training set  $\mathbf{T}_{(\cdot)}$  store a finite amount of pairs  $(\mathbf{p}, \mathbf{a})$ , in which each  $\mathbf{p}$  and each corresponding  $\mathbf{a}$  are connected by the mapping  $H$ . The artificial neural network  $\tilde{H}$  is used to approximate and generalize  $H$  for inputs  $\mathbf{p}$  not in  $\mathbf{T}_{(\cdot)}$ . The architecture of  $\tilde{H}$  was initially chosen accordingly to [16, 19], then suitably adjusted by heuristic arguments.

$\tilde{H}$  is a *fully connected (or dense)* ANN with five *layers* – one *input*, one *output*, and three *hidden* (HidL), as the one illustrated in Figure 3. Input and output layers have a number

Table 3: Artificial neural network and optimization details.

ANN architecture						ANN optimization					
	FxA	FxL	FxF1A	HidL	3	E	3000	InitLR	$10^{-3}$	Opt	Adam
InS	7	6	9	HidS	200	B	1024	DecR	0.1	LossFunc	MSE
OutS	21	21	14	ActUn	ReLU			DecS	1000		

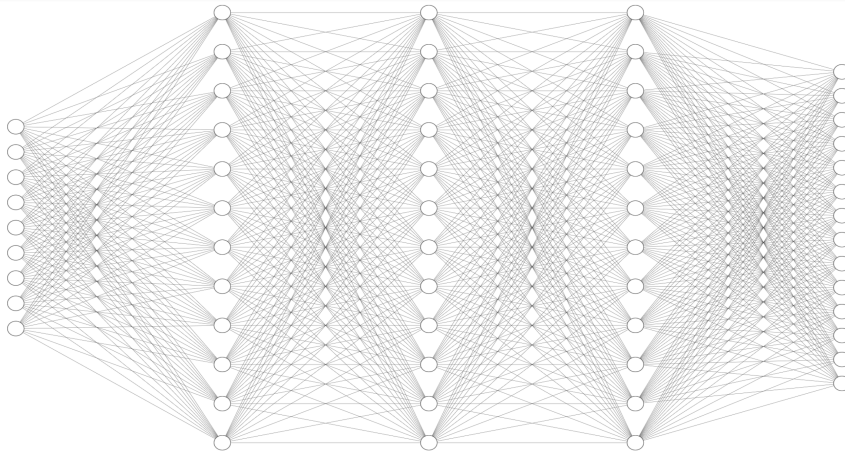


Figure 3: Illustration of a dense ANN with (from the left) one input layer, three hidden layers and one output layer. Each white node represents an activation unit.

of *units (neurons)* – input size (InS) and output size (OutS) – coherent with the targeted problem (FxA, FxL, or FxFIA). Each hidden layer has the same hidden size (HidS) of 200 neurons, selected as the optimal one among different settings. ReLU (Rectified Linear Unit) is the *non-linear activation unit (ActUn)* for each neuron, and it is defined as  $\phi(x) := \max(x, 0)$  [17]. The loss function (LossFunc) is the *Mean Squared Error (MSE)* between the actual outputs,  $\mathbf{a}$  (available in  $\mathbf{T}$ ), and the ones predicted by the ANN,  $\hat{H}(\mathbf{p})$ .

The optimization process is composed of 3000 epochs (E). During each epoch, the major fraction (70%) of the  $\mathbf{T}$  (the actual *training set*) is “back-propagated” through the ANN in batches of size 1024 (B). The stochastic gradient-based optimizer (Opt) Adam [15] is employed in the optimization. Particularly, the optimizer updates the ANN weights based on the gradient computed on each random batch (during each epoch). The initial learning rate (InitLR) is  $10^{-3}$ , with a decaying rate (DecR) of 0.1 and a decaying step (DecS) of 1000 epochs. The details are reported in Table 3.

Furthermore, during the optimization routine, the 20% of  $\mathbf{T}$  is used to *validate* the result (namely, to avoid the overfit of the training set). Eventually, the remaining 10% of  $\mathbf{T}$  is used for *testing* the quality of the ANN. Figure 4 provides a visual insight into the high accuracy the ANN reaches at the end of the training process. Figure 4a shows the scatter plot of the real CVs  $a_k$ ,  $k = 1, \dots, 21$ , against the ones predicted using the ANN, for the experiment FxA; in Figure 4b a zoom is applied to the “worst” case, namely the CV  $a_{10}$ , for which anyway is reached the extremely high  $R^2$  score of 0.9994.

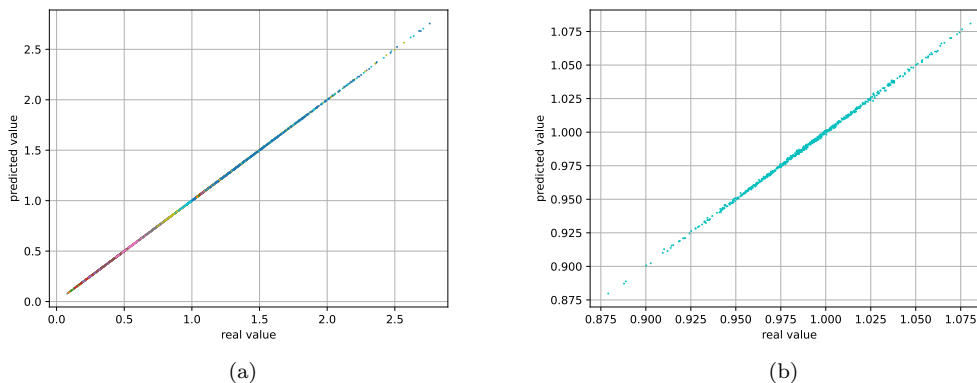


Figure 4: First experiment: FxA. Left: scatter plot of the real CVs ( $a_k$ ,  $k = 1, \dots, 21$ , with different colors) against the predicted ones. Right: zoom on the “worst” case, namely  $a_{10}$ .



### 5.3. Sampling and pricing

Given the trained model from the previous section, we can now focus on the actual sampling and/or pricing of options. In particular, for the first two experiments we consider the following payoffs:

$$\text{FxA: } \max\left(\omega(A(S) - K_2), 0\right), \quad \frac{K_2}{S(t_0)} \in [0.2, 1.7], \quad (5.3)$$

$$\text{FxD: } \max\left(\omega(A(S) - K_2), 0\right), \quad \frac{K_2}{S(t_0)} \in [0.2, 1.7], \quad (5.4)$$

whereas for the third, FxFIA, we have:

$$\max\left(\omega(A(S) - K_1 S(T) - K_2), 0\right), \quad \left(\frac{K_1}{S(t_0)}, \frac{K_2}{S(t_0)}\right) \in [0.10, 0.85] \times [0.10, 0.85], \quad (5.5)$$

with  $A(S)$  defined as in (5.1) for FxA and FxFIA, and as in (5.2) for FxD.

All the results in the following sections are compared to a MC benchmark obtained using the almost exact simulation described in Appendix Appendix A.3.

#### 5.3.1. Numerical results for FxA

The procedure described in Section 3.2 is employed to solve the problem of pricing fixed-strike discrete Asian options with payoff as in (5.3), with underlying stock price initial value  $S(t_0) = 1$ . In this experiment, the ANN is trained on Heston model parameters' ranges, which include the examples proposed in [1] representing some real applications. Furthermore, we note the following aspect.

**Remark 5.1** (Scaled underlying process and (positive) homogeneity of  $A$ ). *The unit initial value is not restrictive. Indeed, the underlying stock price dynamics in Equations (2.9) and (2.10) are independent of  $S(t_0)$ , with the initial value only accounting as a multiplicative constant (this can be easily proved by means of Itô's lemma). Moreover, since  $A(S)$  is (positive) homogeneous in  $S$  also  $A(S)$  can be easily "scaled" according to the desired initial value. Particularly, given the constant  $c > 0$ ,  $cA(S) \stackrel{d}{=} A(cS)$ .*

The methodology is tested on different randomly chosen sets of Heston parameters. We report the details for two specific sets, Set I and Set II, available in Table 4. For Set I, in Figure 5, we compare the population from  $A(S)$  obtained employing SC with the MC benchmark (both with  $N_{\text{paths}} = 3 \cdot 10^6$  paths each). Figure 5a shows the highly accurate approximation of the exact map  $g = F_A^{-1} \circ F_\xi$  by means of the piecewise polynomial approximation  $\tilde{g}$ . As a consequence, both the PDF (see Figure 5b) and the CDF (see Figure 5c) perfectly match. Moreover, the methodology is employed to value fixed strike arithmetic Asian options (calls and puts) for two sets of parameters (Set I and Set II) and 100 different strikes (from deep *out of the money* to deep *in the money*). The resulting prices and absolute errors (in bps) are reported in Figure 6. The same figure shows also the results obtained using the semi-analytic formula

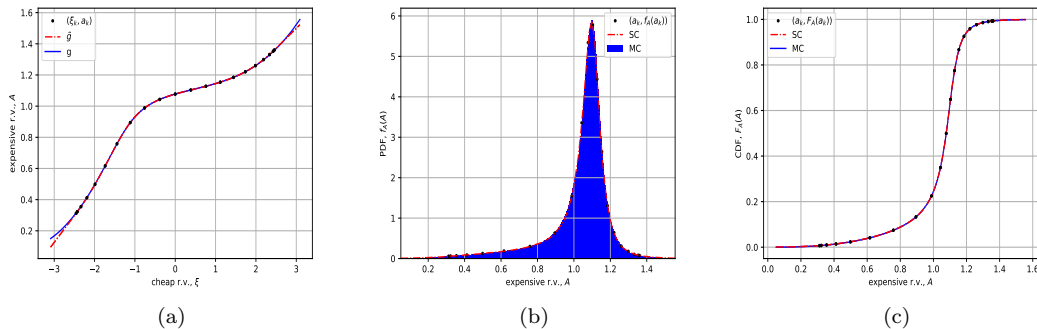


Figure 5: Left: comparison between maps  $g$  from MC and  $\tilde{g}$  from SC (with linear extrapolation). Center: comparison between MC histogram of  $A(S)$  and the numerical PDF from SC. Right: comparison between MC numerical CDF of  $A(S)$  and the numerical CDF from SC.

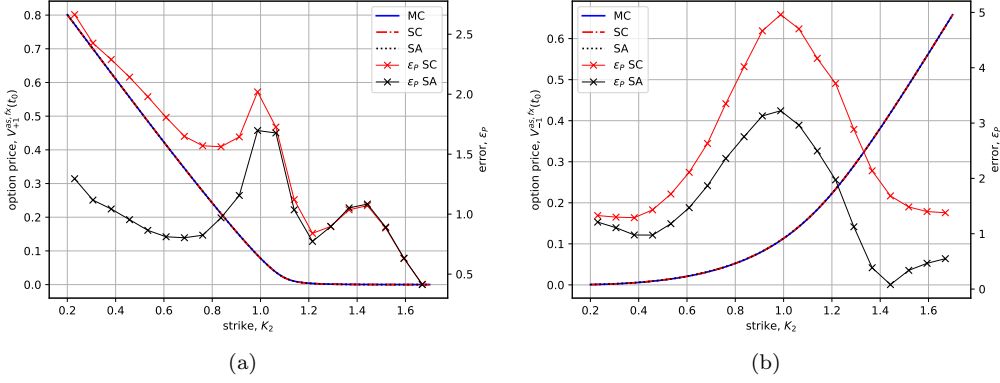


Figure 6: Options for 100 different strikes and maximum absolute error in bps (the SC method (SC) and the semi-analytic formula (PA) are compared with a Monte Carlo benchmark (MC)).  $N_{\text{paths}} = 4 \times 10^6$  for both MC and SC. The extrapolation is quadratic. Left: call, with parameters' Set I. Right: put, with parameters' Set II.

from Proposition 3.2. The timing results are reported (in seconds) in Table 4, together with the maximum absolute error in bps of  $S(t_0)$  (given the different strikes).

The semi-analytic formula requires a constant evaluation time, as well as the SC technique (if  $N_{\text{paths}}$  is fixed), whereas the MC simulation is dependent on the parameter  $T$  (since we decided to keep the same MC step in every simulation). Therefore, the methodology becomes more convenient the longer is the maturity of the option  $T$ . The option pricing computational time is reduced of tens of times when using SC to generate the population from  $A$ , while it is reduced of thousands of times if the semi-analytic (SA) formula is employed.

Eventually, the error distribution of 10000 different option prices (one call and one put with 100 strikes each for 50 randomly chosen Heston parameters' sets and maturities) is given in Figure 7a. The MC semi-analytic price (assuming a linear extrapolation) is compared with MC benchmark. The outcome is satisfactory, and shows the robustness of the methodology proposed. Indeed, the absolute error is smaller than 7.42 bps in the 90% of the experiments, and an error of less than 5 bps is achieved in the 76.4% of the cases.

### 5.3.2. Numerical results for FxL

In this section, we use the procedure to efficiently value the *pipeline risk* typically embedded in mortgages. The pipeline risk (in mortgages) is one of the risk a financial institution is exposed to any time a client buys a mortgage. Indeed, when a client decides to buy a mortgage there is a *grace period* (from one to three months in The Netherlands), during which (s)he is allowed to pick the most convenient rate, namely the minimum.

Observe now that a suitable Lookback option on a payer swap, namely a Lookback payer swaption, perfectly replicates the optionality offered to the client. In other words, the “cost” of the pipeline risk is assessed by evaluating a proper Lookback swaption. In particular, we price fixed-strike discrete Lookback swaptions with a monitoring period of 3 months and 3-day frequency (see 5.2).

We assume the underlying swap rate  $S(t)$ ,  $0 \leq t \leq T$ , is driven by the dynamics given in Equations (2.9) and (2.10) with  $S(t_0) = 0.05$  and parallel shifted of  $\theta = 0.03$ . By introducing

Table 4: Tested Heston parameters and results for FxA.

Set I							Set II								
$r$	$\kappa$	$\gamma$	$\rho$	$\bar{v}$	$v_0$	$T$	$r$	$\kappa$	$\gamma$	$\rho$	$\bar{v}$	$v_0$	$T$		
0.04	0.5	1.0	-0.8	0.08	0.05	1.0	0.02	1.0	0.9	-0.6	0.10	0.13	1.5		
Call (with Set I)							Put (with Set II)								
	time	speed-up	max. err.		time	speed-up	max. err.		time	speed-up	max. err.		time	speed-up	max. err.
MC	60.0 s	/	/	MC	98.2 s	/	/	MC	98.2 s	/	/	MC	98.2 s	/	/
SC	0.784 s	77	2.82 bps	SC	0.782 s	126	4.96 bps	SC	0.782 s	126	4.96 bps	SC	0.782 s	126	4.96 bps
SA	0.013 s	4694	2.04 bps	SA	0.014 s	7283	3.27 bps	SA	0.014 s	7283	3.27 bps	SA	0.014 s	7283	3.27 bps

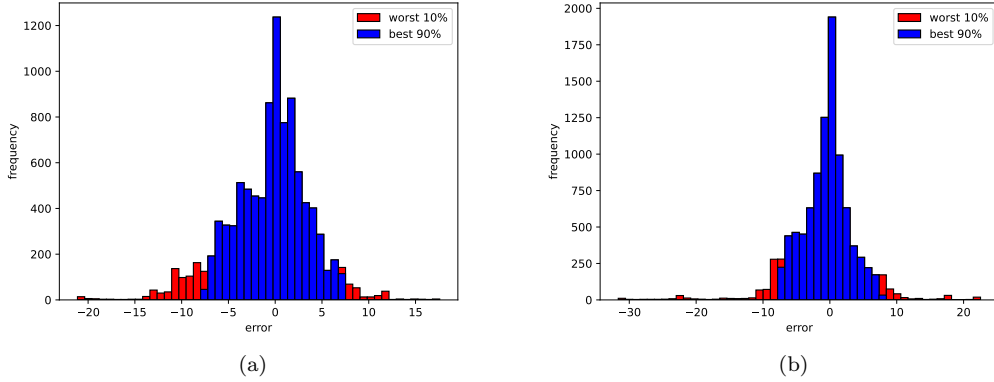


Figure 7: SA error distribution of 10000 option prices: 2 call & put  $\times$  100 strikes  $\times$  50 parameters' set. Blue and red represent the best 90% and the worst 10% errors. The MC benchmark is obtained with  $N_{\text{paths}} = 3 \times 10^6$  paths. Left: FxA. Right: FxL.

a shift, we handle also the possible situation of *negative rates*, which otherwise would require a different model specification.

**Remark 5.2** (Parallel shift of  $S(t)$  and  $A(S)$ ). *A parallel shift  $\theta$  does not affect the training set generation. Indeed, since  $A(S) = \min_n S(t_n)$ , it holds  $A(S - \theta) \stackrel{d}{=} A(S) - \theta$ . Then, it is enough to sample from  $A(S)$  (built from the paths of  $S(t)$  without shift), and perform the shift afterward, to get the desired distribution.*

The timing results from the application of the procedure are comparable to the ones in Section 5.3.1 (see Table 4). Furthermore, in Figure 5b, we report the error distribution obtained by pricing call and put options for 50 randomly chosen Heston parameters' sets and 100 different strikes. The 90% quantile of the (absolute) error distribution is 8.99 bps for the option prices obtained using SC, and 8.18 bps when the semi-analytic formula is employed. The thresholds obtained here are comparable to the ones of the previous example, but we can notice some more extreme outliers which are mainly due to the quality of the training set  $\mathbf{T}_{\text{FxL}}$ . Indeed, the MC computation of  $A(S) = \min_n S(t_n)$  (in the generation of the training set) is more sensitive to numerical instability than the computation of  $A(S) = \sum_n S(t_n)$ . This reason might explain the slightly less accurate results.

### 5.3.3. Numerical results for FxFIA

The third and last experiment consists in the conditional sampling of  $A(S)|S(T)$ . The samples are then used, together with  $S(T)$ , for pricing of fixed- and float-strikes discrete Asian options.

The procedure is tested on 30 randomly chosen Heston parameters' sets. Both the MC benchmark and the SC procedure are based on populations with  $N_{\text{paths}} = 3 \cdot 10^3$  paths. The process  $S(T)$  is sampled using the COS method [18] combined with SC (COS-SC) to avoid a huge number of CDF numerical inversions [10], and so increase efficiency. Then, we apply the grid-based algorithm of Section 3.3. We evaluate the ANN at a reduced number of reference quantiles, and we compute the CVs corresponding to each sample of  $S(T)$  by means of linear interpolation. The CVs identify the map  $\tilde{g}$ , which is employed for the conditional sampling.

Table 5: Tested Heston parameters and results for FxFIA.

Set III						
$r$	$\kappa$	$\gamma$	$\rho$	$\bar{v}$	$v_0$	$T$
0.01	0.46	0.99	-0.79	0.09	0.11	1.0
Call (with Set III)						
	tot. time	speed-up	max. err.			
MC	63.9 s	/	/			
SC	1.654 (0.193) s	39	2.37 bps			

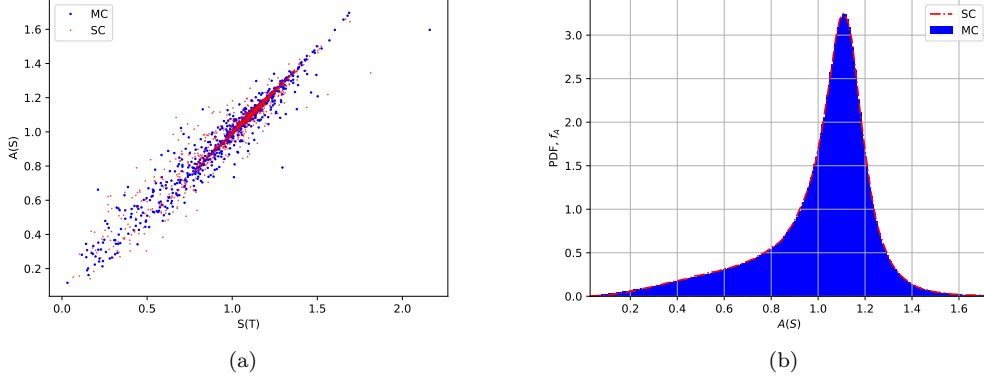


Figure 8: Left: joint distribution of  $(S(T), A(S))$ , for the Heston set of parameters in Set III (see Table 5). Right: marginal distribution of  $A(S)$ , for the same set of Heston parameters.

Figure 8a shows the cloud of points (for parameters' Set III in Table 5) of the bivariate distribution  $(S(T), A(S))$  generated using the procedure against the MC benchmark, while Figure 8b only focuses on the marginal distribution of  $A(S)$ . We can appreciate a good matching between the two distributions.

For each set we price  $30 \times 30$  call options with equally-spaced strikes  $K_1$  and  $K_2$  in the ranges of (5.5). The results for the particular case corresponding to Set III are reported in Table 5, where the timing results keep into account of the pricing of all the 900 different call options (according to each combination of  $K_1$  and  $K_2$ ). Furthermore, in the procedure time, it is reported in brackets what amount of it was spent for applying the COS-SC method, i.e. 0.193 seconds. Figure 9a represents the option price given a fixed-strike  $K_2 = 0.5$  and varying the float-strike  $K_1$ , while in Figure 9b the error distribution is given. We report a 90% quantile (for the absolute error) of 5.75 bps.

It might look surprising that the performance of the general procedure is better than the special one, but actually it is not. Indeed, an important aspect needs to be accounted. The high correlation between  $S(T)$  and  $A(S)$  makes the task of the ANN easier, in the sense that the distribution of  $A(S)|S(T)$  typically has a low variance around  $S(T)$ . In other words, the ANN has to “learn” only a small correction to get  $A(S)|S(T)$  from  $S(T)$  ( $S(T)$  is an input for the ANN!), whereas the ANN in the special procedure learns the unconditional distribution of  $A(S)$  with no information on the final value  $S(T)$ , and so only on the Heston parameters. The result is that a small lack in accuracy due to a not perfect training process, or most likely to a not perfect training set, is less significant in the conditional case rather than in the unconditional.

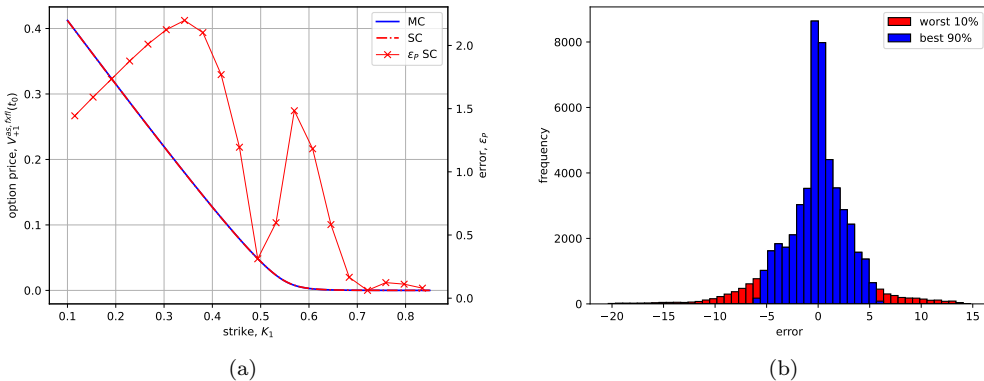


Figure 9: Left: Call price and error for Heston parameters Set III and fixed-strike  $K_2 = 0.5$ . Right: SC error distribution of 54000 option prices: 2 call & put  $\times 30 K_1 \times 30 K_2 \times 30$  parameters' set. Blue and red represent the best 90% and the worst 10% errors. The MC benchmark is obtained with  $N_{\text{paths}} = 3 \times 10^6$  paths.

## 6. Conclusion

In this work, we presented a robust, data-driven procedure for the pricing of fixed- and float-strike discrete Asian and Lookback options, under the stochastic volatility model of Heston. The usage of Stochastic Collocation techniques combined with deep artificial neural networks allows the methodology to reach a high level of accuracy, while reducing the computational time of tens of times, when compared to Monte Carlo benchmarks. Furthermore, a semi-analytic pricing formula for European-type option on a piecewise polynomial of standard normal is provided. Such a result allows to even increase the speed-up up to thousands of times, without deterioration on the accuracy. An analysis of the error provides theoretical justification for the proposed scheme, and the problem of sampling from both unconditional and conditional distributions is further investigated from a numerical perspective. Eventually, the numerical results provide a clear evidence of the quality of the method.

## References

- [1] L. B. Andersen. Efficient Simulation of the Heston Stochastic Volatility Model. *Journal of Computational Finance*, 11, 2007.
- [2] E. Benhamou et al. Fast Fourier Transform for Discrete Asian Options. *Journal of Computational Finance*, 6(1):49–68, 2002.
- [3] S. V. Berezin. On Analytic Characteristic Functions and Processes Governed by SDEs. *St. Petersburg Polytechnical University Journal: Physics and Mathematics*, 2(2):144–149, 2016.
- [4] M. Broadie and Ö. Kaya. Exact Simulation of Stochastic Volatility and other Affine Jump Diffusion Processes. *Operation Research*, 54(2):217–231, 2006.
- [5] A. Conze. Path Dependent Options: The Case of Lookback Options. *The Journal of Finance*, 46(5):1893–1907, 1991.
- [6] J. Devreese, D. Lemmens, and J. Tempere. Path Integral Approach to Asian Options in the Black-Scholes Model. *Physica A: Statistical Mechanics and its Applications*, 389(4):780–788, 2010.
- [7] M. Gaß, K. Glau, M. Mahlstedt, and M. Mair. Chebyshev Interpolation for Parametric Option Pricing. *Finance and Stochastics*, 22:701–731, 2018.
- [8] M. B. Goldman, H. B. Sossin, and M. A. Gatto. Path Dependent Options: “Buy at the low, sell at the high”. *The Journal of Finance*, 34(5):1111–1127, 1979.
- [9] L. A. Grzelak and C. W. Oosterlee. From Arbitrage to Arbitrage-free Implied Volatilities. *Journal of Computational Finance*, 20(3):1–19, 2016.
- [10] L. A. Grzelak, J. A. S. Witteveen, M. Suárez-Taboada, and C. W. Oosterlee. The Stochastic Collocation Monte Carlo Sampler: Highly Efficient Sampling from Expensive Distributions. *Quantitative Finance*, 19(2):339–356, 2019.
- [11] S. L. Heston. A Closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*, 6(2):327–343, 1993.
- [12] R. C. Heynen and H. M. Kat. Lookback Options with Discrete and Partial Monitoring of the Underlying Price. *Applied Mathematical Finance*, 2(4):273–284, 1995.
- [13] A. Kemna and A. Vorst. A Pricing Method for Options based on Average Asset Values. *Journal of Banking & Finance*, 14(1):113–129, 1990.
- [14] M. G. Kendall. *The Advanced Theory of Statistics*. Wiley, 1945.
- [15] D. P. Kingma and J. L. Ba. Adam: A Method for Stochastic Optimization. *Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego*, 2015.
- [16] S. Liu, L. A. Grzelak, and C. W. Oosterlee. The Seven-League Scheme: Deep Learning for Large Time Step Monte Carlo Simulations of Stochastic Differential Equations. *Risks*, 10(3):47, 2022.
- [17] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall. Activation Functions: Comparison of Trends in Practice and Research for Deep Learning. *arXiv e-prints*, 2018.
- [18] C. W. Oosterlee and L. A. Grzelak. *Mathematical Modeling and Computation in Finance*. World Scientific Publishing Europe Ltd., 57 Shelton Street, Covent Garden, London WC2H 9HE, 2019.
- [19] L. Perotti and L. A. Grzelak. Fast Sampling from Time-integrated Bridges using Deep Learning. *Journal of Computational Mathematics and Data Science*, 5, 2022.
- [20] L. N. Trefethen. *Approximation Theory and Approximation Practice, Extended Edition*. Society for Industrial and Applied Mathematics, 2019.
- [21] J. Vecer. Unified Pricing of Asian Options. *Risk*, 15(6):113–116, 2002.
- [22] P. Wilmott, J. Dewynne, and S. Howison. *Option Pricing: Mathematical Models and Computation*. Oxford Financial Press: Oxford, 1993.
- [23] S. Xiang, X. Chen, and H. Wang. Error Bounds for Approximation in Chebyshev Points. *Numerische Mathematik*, 116:463–491, 2010.
- [24] D. Yarotsky. Error Bounds for Approximations with Deep ReLU Networks. *Neural Networks*, 94:103–114, 2017.
- [25] B. Zhang and C. W. Oosterlee. Efficient Pricing of European-style Asian Options under Exponential Lévy Processes based on Fourier Cosine Expansions. *SIAM Journal of Financial Mathematics*, 4:399–426, 2013.

## Appendix A. Proofs and lemmas

### Appendix A.1. Underlying process measure for float-strike options

*Proof of Proposition 2.1.* Under the risk-neutral measure  $\mathbb{Q}$  the value at time  $t_0 \geq 0$  of a float-strike Asian Option, with maturity  $T > t_0$ , underlying process  $S(t)$ , and future monitoring dates  $t_n$ ,  $n \in \{1, \dots, N\}$ , is given by:

$$V_\omega^{\text{f1}}(t_0) = \mathbb{E}_{t_0}^{\mathbb{Q}} \left[ \frac{M(t_0)}{M(T)} \max \left( \omega(A(S) - K_1 S(T)), 0 \right) \right].$$

We define a Radon-Nikodym derivative to change the measure from the risk-neutral measure  $\mathbb{Q}$  to the stock measure  $\mathbb{Q}^S$ , namely the measure associated with the numéraire  $S$ :

$$\frac{d\mathbb{Q}^S}{d\mathbb{Q}} = \frac{S(T) M(t_0)}{S(t_0) M(T)},$$

which yields the following present value, expressed as an expectation under the measure  $\mathbb{Q}^S$ :

$$\begin{aligned} V_\omega^{\text{f1}}(t_0) &= \mathbb{E}_{t_0}^S \left[ \frac{M(t_0)}{M(T)} \max \left( \omega(A(S) - K_1 S(T)), 0 \right) \frac{S(t_0) M(T)}{S(T) M(t_0)} \right] \\ &= S(t_0) \mathbb{E}_{t_0}^S \left[ \max \left( \omega \left( \frac{A(S)}{S(T)} - K_1 \right), 0 \right) \right]. \end{aligned}$$

■

*Proof of Proposition 2.3.* Under the stock measure  $\mathbb{Q}^S$ , implied by the stock  $S(t)$  as numéraire, all the assets discounted with  $S$  must be martingales. Particularly, this entails that  $M(t)/S(t)$  must be a martingale, where  $M(t)$  is the money-savings account defined as  $dM(t) = rM(t)dt$ .

From (2.9) and (2.10), using Cholesky decomposition, the Heston model can be expressed in terms of independent Brownian motions,  $\widetilde{W}_x(t)$  and  $\widetilde{W}_v(t)$ , through the following system of SDEs:

$$\begin{aligned} dS(t) &= rS(t)dt + \sqrt{v(t)}S(t)d\widetilde{W}_x(t), \\ dv(t) &= \kappa(\bar{v} - v(t))dt + \gamma\sqrt{v(t)} \left[ \rho d\widetilde{W}_x(t) + \sqrt{1 - \rho^2} d\widetilde{W}_v(t) \right]. \end{aligned}$$

After application of Itô's Lemma we find:

$$d \frac{M(t)}{S(t)} = \frac{1}{S(t)} rM(t)dt - \frac{M(t)}{S^2(t)} \left( rS(t)dt + \sqrt{v(t)}S(t)d\widetilde{W}_x(t) \right) + \frac{M(t)}{S^3(t)} v(t)S^2(t)dt,$$

which implies the following measure transformation:

$$d\widetilde{W}_x(t) = d\widetilde{W}_x^S(t) + \sqrt{v(t)}dt.$$

Thus, under the stock measure  $\mathbb{Q}^S$ , the dynamics of  $S(t)$  reads:

$$\begin{aligned} dS(t) &= rS(t)dt + \sqrt{v(t)}S(t) \left( d\widetilde{W}_x^S(t) + \sqrt{v(t)}dt \right) \\ &= (r + v(t))S(t)dt + \sqrt{v(t)}S(t)d\widetilde{W}_x^S(t), \end{aligned}$$

while for the dynamics of  $v(t)$  we find:

$$\begin{aligned} dv(t) &= \kappa(\bar{v} - v(t))dt + \gamma\sqrt{v(t)} \left[ \rho \left( d\widetilde{W}_x^S(t) + \sqrt{v(t)}dt \right) + \sqrt{1 - \rho^2} d\widetilde{W}_v(t) \right] \\ &= [\kappa(\bar{v} - v(t)) + \gamma\rho v(t)]dt + \gamma\sqrt{v(t)} \left[ \rho d\widetilde{W}_x^S(t) + \sqrt{1 - \rho^2} d\widetilde{W}_v(t) \right]. \end{aligned}$$

Setting  $\kappa^* := \kappa - \gamma\rho$ ,  $\bar{v}^* := \kappa\bar{v}/\kappa^*$ ,  $W_x^S(t) := \widetilde{W}_x^S(t)$ , and  $W_v^S(t) := \rho\widetilde{W}_x^S(t) + \sqrt{1 - \rho^2}\widetilde{W}_v(t)$  the proof is complete. ■

Appendix A.2. Semi-analytic pricing formula

**Result Appendix A.1** (Moments of truncated standard normal distribution). Let  $\xi \sim \mathcal{N}(0, 1)$  and  $a, b \in [-\infty, +\infty]$ ,  $a < b$ . Then, the recursive expression for:

$$m_i(a, b) := \mathbb{E}[\xi^i | a \leq \xi \leq b],$$

the  $i$ -th moment of the truncated standard normal distribution  $\xi | a \leq \xi \leq b$ , reads:

$$m_i(a, b) = (i - 1)m_{i-2}(a, b) - \frac{b^{i-1}f_\xi(b) - a^{i-1}f_\xi(a)}{F_\xi(b) - F_\xi(a)}, \quad i \in \mathbb{N} \setminus \{0\},$$

where  $m_{-1}(a, b) := 0$ ,  $m_0(a, b) := 1$ , and  $f_\xi$  and  $F_\xi$  are the PDF and the CDF of  $\xi$ , respectively.

**Result Appendix A.2** (Expectation of polynomial of truncated normal distribution). Let  $p(x) = \sum_{i=0}^{M-1} \alpha_i x^i$  be a  $(M - 1)$ -degree polynomial and let  $\xi \sim \mathcal{N}(0, 1)$ , with  $f_\xi$ ,  $F_\xi$  its PDF and CDF, respectively. Then, for any  $a, b \in [-\infty, +\infty]$  with  $a < b$ , the following holds:

$$\int_a^b p(x)f_\xi(x)dx = \sum_{i=0}^{M-1} \alpha_i m_i(a, b) (F_\xi(b) - F_\xi(a)),$$

with  $m_i(a, b)$  as defined in Result Appendix A.1.

*Proof.* The proof immediately follows thanks to the following equalities:

$$\begin{aligned} \int_a^b p(x)f_\xi(x)dx &= \sum_{i=0}^{M-1} \alpha_i \int_a^b x^i f_\xi(x)dx = \sum_{i=0}^{M-1} \alpha_i \mathbb{E}[\xi^i \mathbb{1}_{[a,b]}(\xi)] \\ &= \sum_{i=0}^{M-1} \alpha_i \mathbb{E}[\xi^i | a \leq \xi \leq b] \cdot \mathbb{P}[a \leq \xi \leq b]. \end{aligned}$$

■

*Proof of Proposition 3.2.* The approximation  $\tilde{g}$  is strictly increasing in the domain of interest. Then, setting  $c_K = \tilde{g}^{-1}(K)$ , we have:

$$\begin{aligned} \frac{\tilde{V}_\omega(t_0)}{C} &= \int_{-\infty}^{+\infty} \max(\omega(\tilde{g}(x) - K), 0) f_\xi(x)dx \\ &= \int_{\omega c_K}^{+\infty} \omega(\tilde{g}(\omega y) - K) f_\xi(\omega y)dy \\ &= \omega \left( \int_{\omega c_K}^{+\infty} \tilde{g}(\omega y) f_\xi(y)dy - K \mathbb{P}[\xi > \omega c_K] \right), \end{aligned}$$

where the first equality holds by definition of expectation, the second one relies on a suitable change of variable ( $y = x$ ) and the last one holds thanks to the even symmetry of  $f_\xi$ . We define the integral  $I_\omega(c_K)$  as:

$$I_\omega(c_K) := \int_{\omega c_K}^{+\infty} \tilde{g}(\omega x) f_\xi(x)dx,$$

and using the definition of  $\tilde{g}$  as a piecewise polynomial, we get:

$$I_\omega(c_K) = \int_{\omega c_K}^{-\tilde{\xi} \vee c_K} g_{-\omega}(\omega x) f_\xi(x)dx + \int_{-\tilde{\xi} \vee c_K}^{\tilde{\xi} \vee c_K} g_M(\omega x) f_\xi(x)dx + \int_{\tilde{\xi} \vee c_K}^{+\infty} g_\omega(\omega x) f_\xi(x)dx. \quad (\text{A.1})$$

The thesis follows by applying Result Appendix A.2 at each term in (A.1) and exploiting the definition of  $F_\xi$ . ■

*Appendix A.3. Almost Exact Simulation from the Heston Model*

In a MC framework, the most common scheme employed in the industry is the Euler-Maruyama discretization of the system of SDEs which describe the underlying process dynamics. For the the stochastic volatility model of Heston, such scheme can be improved, allowing for an *exact* simulation of the variance process  $v(t)$  (see (2.10)), as shown in [4]. This results in increased accuracy, and avoids numerical issues due to the theoretical non-negativity of the process  $v(t)$ , leading to the so-called *almost exact simulation* of the Heston model [18].

**Result Appendix A.3** (Almost exact simulation from the Heston Model). *Given  $X(t) := \log S(t)$ , its dynamics<sup>8</sup> between the consequent times  $t_i$  and  $t_{i+1}$  is discretized with the following scheme:*

$$\begin{aligned} x_{i+1} &\approx x_i + k_0 + k_1 v_i + k_2 v_{i+1} + \sqrt{k_3 v_i} \xi, \\ v_{i+1} &= \bar{c} \chi^2(\delta, \bar{\kappa} v_i) \end{aligned} \tag{A.2}$$

with the quantities:

$$\Delta t := t_{i+1} - t_i, \quad \delta := \frac{4\kappa\bar{v}}{\gamma^2}, \quad \bar{c} := \frac{\bar{v}}{\delta}(1 - e^{-\kappa\Delta t}), \quad \bar{\kappa} := \bar{c}^{-1}e^{-\kappa\Delta t},$$

the noncentral chi-squared random variable  $\chi^2(\delta, \eta)$  with  $\delta$  degrees of freedom and non-centrality parameter  $\eta$ , and  $\xi \sim \mathcal{N}(0, 1)$ . The remaining constants are defined as:

$$k_0 := \left(r - \frac{\rho}{\gamma}\kappa\bar{v}\right)\Delta t, \quad k_1 := \left(\frac{\rho\kappa}{\gamma} - \frac{1}{2}\right)\Delta t - \frac{\rho}{\gamma}, \quad k_2 := \frac{\rho}{\gamma}, \quad k_3 := (1 - \rho^2)\Delta t.$$

*Derivation.* Given  $X(t) = \log S(t)$ , by applying Itô's Lemma and Cholesky decomposition on the dynamics in (2.9) and (2.10), we get:

$$dX(t) = \left(r - \frac{1}{2}v(t)\right) dt + \sqrt{v(t)} \left[\rho d\widetilde{W}_v(t) + \sqrt{1 - \rho^2} d\widetilde{W}_x(t)\right], \tag{A.3}$$

$$dv(t) = \kappa(\bar{v} - v(t)) dt + \gamma\sqrt{v(t)} d\widetilde{W}_v(t), \tag{A.4}$$

where  $\widetilde{W}_x(t)$  and  $\widetilde{W}_v(t)$  are independent BMs.

By integrating (A.3) and (A.4) in a the time interval  $[t_i, t_{i+1}]$ , the following discretization scheme is obtained:

$$x_{i+1} = x_i + \int_{t_i}^{t_{i+1}} \left(r - \frac{1}{2}v(t)\right) dt + \rho \int_{t_i}^{t_{i+1}} \sqrt{v(t)} d\widetilde{W}_v(t) + \sqrt{1 - \rho^2} \int_{t_i}^{t_{i+1}} \sqrt{v(t)} d\widetilde{W}_x(t), \tag{A.5}$$

$$v_{i+1} = v_i + \kappa \int_{t_i}^{t_{i+1}} (\bar{v} - v(t)) dt + \gamma \int_{t_i}^{t_{i+1}} \sqrt{v(t)} d\widetilde{W}_v(t), \tag{A.6}$$

where  $x_i := X(t_i)$ ,  $x_{i+1} := X(t_{i+1})$ ,  $v_i := v(t_i)$ ,  $v_{i+1} := v(t_{i+1})$ .

Given  $v_i$ , the variance  $v_{i+1}$  is distributed as a suitable scaled noncentral chi-squared distribution [18]. Therefore, we substitute  $\int_{t_i}^{t_{i+1}} \sqrt{v(t)} d\widetilde{W}_v^\lambda(t)$  in (A.5) using (A.6), ending up with:

$$\begin{aligned} x_{i+1} &= x_i + \int_{t_i}^{t_{i+1}} \left(r - \frac{1}{2}v(t)\right) dt + \frac{\rho}{\gamma} \left(v_{i+1} - v_i - \kappa \int_{t_i}^{t_{i+1}} (\bar{v} - v(t)) dt\right) \\ &\quad + \sqrt{1 - \rho^2} \int_{t_i}^{t_{i+1}} \sqrt{v(t)} d\widetilde{W}_x(t). \end{aligned}$$

We approximate the integrals in the expression above employing the left integration boundary values of the integrand, as in the Euler-Maruyama discretization scheme. The scheme (A.2) follows collecting the terms and employing the property  $\widetilde{W}_x(t_{i+1}) - \widetilde{W}_x(t_i) \stackrel{d}{=} \sqrt{\Delta t} \xi$ , with  $\xi \sim \mathcal{N}(0, 1)$  and  $\Delta t := t_{i+1} - t_i$ .  $\blacksquare$

<sup>8</sup>Under the risk-neutral measure,  $\mathbb{Q}$ . However, the same scheme applies under the underlying process measure  $\mathbb{Q}^S$ , with only a minor difference, i.e.,  $k_1 := (\rho\kappa/\gamma + 1/2)\Delta t - \rho/\gamma$ .



*Appendix A.4. SC error analysis for Chebyshev interpolation*

The two following lemmas are useful to show that the conditional complex ChF  $\phi_{A|\mathbf{V}}(z) = \mathbb{E}[e^{izA}|\mathbf{V}]$  is an analytic function of  $z \in \mathbb{C}$ . The first one provides the law of the conditional stock-price distribution, whereas the second one is meant to give algebraic bounds for the target function  $A(S)$ .

**Lemma Appendix A.1** (Conditional distribution under Heston). *Let  $S(t)$  be the solution at time  $t$  of Equation (2.9) and  $I_v(t) := \int_{t_0}^t v(\tau)d\tau$ , with  $v$  driven by the dynamics in Equation (2.10). Then, the following equality in distribution holds:*

$$S(t)|I_v(t_0, t) \stackrel{d}{=} \exp\left(\mu(I_v(t_0, t)) + \sigma(I_v(t_0, t))\xi\right),$$

with  $\xi \sim \mathcal{N}(0, 1)$ ,  $\mu$  and  $\sigma$  defined as  $\mu(y) := \log S(t_0) + r(t - t_0) - y/2$  and  $\sigma(y) := \sqrt{y}$ . Furthermore, for any  $k = 0, 1, \dots$ , the following holds:

$$\mathbb{E}[S(t)^k | I_v(t)] = \exp\left(k\mu(I_v(t)) + \frac{1}{2}k^2\sigma^2(I_v(t))\right) \quad (\text{A.7})$$

In other words, the stock price given the time-integral of the variance process  $I_v$  is log-normally distributed, with parameters dependent on the time-integral  $I_v$ , and its moments up to any order are given as in Equation (A.7).

*Proof.* Writing (2.9) in integral form we get:

$$S(t) = S(t_0) \exp\left(r(t - t_0) - \frac{1}{2} \int_{t_0}^t v(\tau)d\tau + \int_{t_0}^t \sqrt{v(\tau)}dW_x(\tau)\right).$$

By considering the conditional distribution  $S(t)|I_v(t)$  (instead of  $S(t)$ ) the only source of randomness is given by the Itô's integral (and it is due to the presence of the Brownian motion  $W_x(t)$ ). The thesis follows since the Itô's integral of a deterministic argument is normally distributed with zero mean and variance given by the time integral of the argument squared (in the same interval). Therefore,  $S(t)|I_v(t)$  is log-normally distributed, with moments given as in (A.7). ■

**Lemma Appendix A.2** (Algebraic bounds). *Let us consider  $\{s_1, \dots, s_N\}$ , with  $s_n > 0$  for each  $n = 1, \dots, N$ . Then, for any  $k = 1, 2, \dots$ , we have:*

1.  $(\sum_n s_n)^k \leq 2^{(N-1)(k-1)} \sum_n s_n^k$ .
2.  $(\min_n s_n)^k \leq s_{n^*}^k$  for any  $n^* = 1, \dots, N$ .

*Proof.* The second thesis is obvious. We prove here the first one. We recall that in general, given  $a, b > 0$  and any  $k = 1, 2, \dots$ , the following inequality holds:

$$(a + b)^k \leq 2^{k-1}(a^k + b^k). \quad (\text{A.8})$$

Then, applying (A.8)  $N - 1$  times we get:

$$\left(\sum_{n=1}^N s_n\right)^k \leq 2^{k-1} \left(s_1^k + \left(\sum_{n=2}^N s_n\right)^k\right) \leq \dots \leq 2^{(N-1)(k-1)} s_N^k + \sum_{n=1}^{N-1} 2^{n(k-1)} s_n^k,$$

which can be further bounded by:

$$\left(\sum_{n=1}^N s_n\right)^k \leq 2^{(N-1)(k-1)} s_N^k + \sum_{n=1}^{N-1} 2^{(N-1)(k-1)} s_n^k = \sum_{n=1}^N 2^{(N-1)(k-1)} s_n^k,$$

■

We have all the ingredients to prove Proposition 4.4.

*Proof of Proposition 4.4.* To exploit the characterization for entire ChFs in Theorem 4.3 we need to show the finiteness of each absolute moment as well as that Equation (4.5) is satisfied. Both the conditions can be proved using Lemma Appendix A.1 and Lemma Appendix A.2. For  $k = 0, 1, \dots$ , we consider the two cases:

1. If  $A = \frac{1}{N} \sum_n S(t_n)$ , then thanks to Lemma Appendix A.2 we have:

$$\begin{aligned} \mathbb{E}[|A|^k | \mathbf{V}] &= \frac{1}{N^k} \mathbb{E} \left[ \left( \sum_n S(t_n) \right)^k \middle| \mathbf{V} \right] \leq \frac{2^{(N-1)(k-1)}}{N^k} \mathbb{E} \left[ \sum_n S(t_n)^k \middle| \mathbf{V} \right] \\ &= \frac{2^{(N-1)(k-1)}}{N^k} \sum_n \mathbb{E}[S(t_n)^k | \mathbf{V}], \end{aligned}$$

whereas from Lemma Appendix A.1 follows:

$$\begin{aligned} \mathbb{E}[|A|^k | \mathbf{V}] &\leq \frac{2^{(N-1)(k-1)}}{N^k} \sum_n \mathbb{E}[S(t_n)^k | \mathbf{V}] \\ &= \frac{2^{(N-1)(k-1)}}{N^k} \sum_n \exp \left( k\mu_n(\mathbf{V}) + \frac{1}{2}k^2\sigma_n^2(\mathbf{V}) \right), \end{aligned} \tag{A.9}$$

where  $\mu_n(\mathbf{V}) := \mu(I_v(t_n))$  and  $\sigma_n(\mathbf{V}) := \sigma(I_v(t_n))$ ,  $n = 1, \dots, N$ .

2. If  $A = \min_n S(t_n)$ , then we immediately have:

$$\begin{aligned} \mathbb{E}[|A|^k | \mathbf{V}] &\leq \mathbb{E}[S(t_{n^*})^k | \mathbf{V}] \\ &= \exp \left( k\mu_{n^*}(\mathbf{V}) + \frac{1}{2}k^2\sigma_{n^*}^2(\mathbf{V}) \right), \end{aligned} \tag{A.10}$$

for an arbitrary  $n^* = 1, \dots, N$ .

The finiteness of the absolute moments up to any order follows directly from (A.9) and (A.10) respectively, since  $I_v(t_n)$  are finite (indeed, they are time-integrals on compact intervals of continuous paths).

Eventually, thanks to Jensen's inequality we have  $|\mathbb{E}[A | \mathbf{V}]|^k \leq \mathbb{E}[|A|^k | \mathbf{V}]$ . This, together with the at most exponential growth (in  $k$ ) of the absolute moments of  $A | \mathbf{V}$ , ensures that the limit in Equation (4.5) holds. Then, by Theorem 4.3,  $\phi_{A | \mathbf{V}}(z)$  is an entire function of the complex variable  $z \in \mathbb{C}$ .  $\blacksquare$

*Proof of Proposition 4.5.* The goal here is to apply Morera's theorem. Hence, let  $\gamma \in \mathcal{S}_{y^*}$  be any piecewise  $C^1$  closed curve in the strip  $\mathcal{S}_{y^*}$ . Then:

$$\begin{aligned} \int_\gamma \phi_A(z) dz &\stackrel{(4.6)}{=} \int_\gamma \int_{\Omega_{\mathbf{V}}} \phi_{A | \mathbf{V}=\mathbf{v}}(z) dF_{\mathbf{V}}(\mathbf{v}) dz \\ &\stackrel{\text{Fubini}}{=} \int_{\Omega_{\mathbf{V}}} \int_\gamma \phi_{A | \mathbf{V}=\mathbf{v}}(z) dz dF_{\mathbf{V}}(\mathbf{v}) \stackrel{\text{Cauchy}}{=} 0, \end{aligned}$$

where in the first equality we exploited the representation of the unconditional ChF  $\phi_A$  in terms of conditional ChFs  $\phi_{A | \mathbf{V}}$ , in the second equality we use Fubini's theorem to exchange the order of integration, and eventually in the last equation we employ the Cauchy's integral theorem on  $\int_\gamma \phi_{A | \mathbf{V}=\mathbf{v}}(z) dz$ .  $\blacksquare$