American options in the Volterra Heston model*

Etienne Chevalier† Sergio Pulido‡ Elizabeth Zúñiga§

March 23, 2021

Abstract

We price American options using kernel-based approximations of the Volterra Heston model. We choose these approximations because they allow simulation-based techniques for pricing. We prove the convergence of American option prices in the approximating sequence of models towards the prices in the Volterra Heston model. A crucial step in the proof is to exploit the affine structure of the model in order to establish explicit formulas and convergence results for the conditional Fourier-Laplace transform of the log price and an adjusted version of the forward variance. We illustrate with numerical examples our convergence result and the behavior of American option prices with respect to certain parameters of the model.

1 Introduction

Stochastic volatility models whose trajectories are continuous but less regular than Brownian motion, also known as rough volatility models, seem well-adapted to capture stylized features of the time series of realized volatility and of the implied volatility surface. Indeed, recent statistical studies in [14, 25, 24] demonstrate that – under multiple time scales and across many markets – the time series of realized volatility oscillates more rapidly than Brownian motion. In addition, the observed implied volatility smile for short maturities is steeper than the one obtained with classical low-dimensional diffusion models. As

*The authors would like to thank Eduardo Abi Jaber for useful comments and fruitful discussions. The research of Sergio Pulido benefited from the financial support of the chair Deep finance and Statistics of École Polytechnique. Sergio Pulido acknowledges support by the Europlace Institute of Finance (EIF) and the Labex Louis Bachelier, research project: "The impact of information on financial markets". The work of Elizabeth Zúñiga was supported by grants from the Fondation Mathématique Jacques Hadamard (FMJH).

†Université Paris-Saclay, CNRS, Univ Evry, Laboratoire de Mathématiques et Modélisation d'Evry (LaMME), etienne.chevalier@univ-evry.fr

‡Université Paris-Saclay, CNRS, ENSIIE, Univ Évry, Laboratoire de Mathématiques et Modélisation d'Evry (LaMME), sergio.pulidonino@ensiie.fr

§Université Paris-Saclay, CNRS, Univ Evry, Laboratoire de Mathématiques et Modélisation d'Evry (LaMME), elizabeth.zunigariofrio@univ-evry.fr

maturity decreases, the slope at the money of the implied volatility smile obeys a power law that explodes at zero. This power law can be reproduced by rough volatility models with power kernels in the spirit of fractional Brownian motion, cf. [9, 23]. Furthermore, these empirical discoveries are supported by micro-structural considerations because, as explained for instance in [35, 19], rough volatility models appear naturally as scaling limits of micro-structural pricing models with self-exciting features driven by Hawkes processes.

The aforementioned findings have motivated the study of various rough volatility models in the literature. Among these are the rough fractional stochastic volatility model [25], the rough Bergomi model [9], and the fractional and rough Heston models [16, 29, 21]. In these models, due to the absence of the semimartingale and Markov properties, even simple tasks such as pricing European options have proven challenging. Consequently, the theory of stochastic control for rough volatility models is at an early stage. Under the rough volatility paradigm, classical control problems such as linear quadratic and optimal investment problems have only been analyzed recently, for example in [6] and [22, 7, 30, 31], respectively.

In this paper we tackle an optimal stopping problem, namely the problem of pricing American options, in the Volterra Heston model introduced in [3, 5]. This path-dependent problem is difficult because it requires a good understanding of the conditional laws in a model where in general the semimartingale and Markov properties do not hold. Even though we could extend parts of the analysis to more general frameworks, we concentrate on the Volterra Heston model because in this setup – as we will explain below – we can prove the necessary convergence results.

The Volterra Heston model is a generalization of the widely-known Heston model [33]. The dynamics of the spot variance in the Volterra Heston model are described by a stochastic Volterra equation of convolution type. More specifically, the spot variance process is a Volterra square root or CIR process. When the kernel appearing in the convolution is of power-type, one obtains the now well-known rough Heston model [20, 21]. The $\mathcal{L}^2$-regularity of the kernel in the Volterra Heston model controls the Hölder regularity of the trajectories and the steepness of the implied volatility smile for short maturities. Tractability in the Volterra Heston model is a result of a semi-explicit formula for the Fourier-Laplace transform, which resembles the formula in the classical Heston model. More precisely, the Fourier-Laplace transform can be expressed in terms of the solution to a deterministic system of convolution equations of Riccati-type. This phenomenon is a particular instance of a more general law governing the structure of the Fourier-Laplace transform of what is known as Affine Volterra Processes [5, 36, 26, 18]. The knowledge of the Fourier-Laplace transform in the Volterra Heston model facilitates the application of Fourier-based methods in order to price European options. This circumvents the difficulties encountered in the implementation of other popular rough volatility models, such as the rough Bergomi model, where Monte-Carlo techniques [9, 13] or Donsker-type theorems [34] are employed to compute prices of European options.

The numerical resolution of the Riccati convolution equations appearing in the expression of the Fourier-Laplace transform in the Volterra Heston model is, however, cumbersome due to the possibly exploding character of the associated kernel. In order to alleviate these numerical difficulties for the rough Heston model, the author in [1] proposed a kernel-based approximation with a diffusion – high dimensional but parsimonious – model, named the Lifted Heston model. Despite being a semimartingale model, the Lifted Heston model is able to mimic the rough character of the trajectories and to reproduce steep volatility smiles for short maturities. The approximation of the rough Heston model with the Lifted Heston model is an example of a more general approximation technique of Volterra processes via an approximation of the kernel in [4] originally inspired by [17, 15, 32]. The convergence of the approximating processes and the prices of European-type options is guaranteed by stability results proven in [4] and in a more general framework in [2].

To price American options, and inspired by the approach in [1, 4], we draw upon kernel-based approximations of the Volterra Heston model. In the context of the rough Heston model where the kernel is of power-type, and for the approximation scheme in [1], the approximating models are high dimensional-diffusion models where classical simulation-based techniques, such as the Longstaff Schwartz algorithm [38], can be implemented. Within this framework, we can conduct an empirical study of the convergence and behavior of Bermudan put option prices in the approximated sequence of models. The results of our numerical experiments are summarized in Section 5.

Our main theoretical result is Theorem 2.7. In the first part of the theorem, we show convergence of prices of Bermudan options in the approximating sequence of models towards the prices in the original Volterra Heston model. This result is not a direct consequence of previous stability results in [4, 2] because of the path-dependent structure of the option. It is at this stage, and for purely theoretical reasons, that we exploit the affine structure of the model. More precisely, in order to prove the desired convergence results we first need to establish the convergence of the conditional Fourier-Laplace transforms. Once the convergence of the Bermudan option prices is established – and using classical arguments – we can prove, in the second part of Theorem 2.7, the convergence of American option prices by approximating them with Bermudan option prices.

It is important to mention at this point that there exist other studies of optimal stopping and American option pricing in rough or fractional models; see for instance [34, 12, 11, 27, 11]. To understand the novelty of our work it is crucial to point out that in general there are two levels of approximation in the resolution of an optimal stopping problem using a probabilistic approach:

(i) First, the model has to be approximated with simpler models where the trajectories can be simulated or where prices of American options can be computed more easily. For classical diffusion models this could correspond to a classical Euler scheme for simulation or a tree-based discrete approximation. Under rough volatility, simulation is cumbersome due to the non-Markovianity of the model. There is not a unified theory about how this approximation and simulation have to be performed. For instance, in the rough Bergomi model in order to simulate the volatility process one could use hybrid schemes [13]. These schemes correspond to an approximation of the power kernel by concentrating on its behavior around zero and performing a step-wise approximation away from zero. But we could also imagine schemes relying on an approximation of the fractional kernel in terms of a sum of exponentials as in [15, 32]. Other recent studies in this direction are [41, 8]. In this work, for our numerical illustrations, we use the approximation scheme of [1, 4], based on an approximation of the kernel using a sum of exponentials. Regarding the approximation via discrete-type models, in [34] the authors prove a Donsker-type theorem for certain rough volatility models and apply it to perform tree-like approximations. These approximations allow them to develop tree-based algorithms, as opposed to simulation-based techniques, to price American options. The convergence of the American option prices computed on the approximating trees towards prices in the limiting rough models, however, is not the main goal of the study.

(ii) The second approximation occurs at the level of the resolution of the optimal stopping problem for the approximated model. In the approximated model, classical techniques such as the Longstaff Schwarz algorithm, can be difficult to implement because of the high-dimensionality of the model. It is at this stage that recent studies propose novel approaches, including techniques relying on neural networks [37, 27], to ease the implementation. It is also important to mention at this point the study in [10], where the authors propose an approximation of American option prices using penalized versions of the BSPDE satisfied by the value function of the problem. A deep learning-based method is used to approximate the solutions of these penalized BSPDEs.

The present paper does not focus on the second level of the approximation. For this part, in our numerical experiments we employ classical simulation-based techniques and in particular the Longstaff Schwarz algorithm over a low dimensional space of functions. Our study mainly focuses on the first level of the approximation. More precisely, we concentrate on the convergence of the prices in the approximating model towards the prices in the limiting Volterra model. This point has not been addressed in the previous literature and is what distinguishes our paper from other papers on American options under the rough volatility paradigm. To prove this convergence in our framework and with our kernel-based approximation approach, we appeal to the particular affine structure of the Volterra Heston model, which explains our choice of setting. One could extend some of the results to other settings as long as the results regarding the convergence of the conditional Fourier-Laplace transform remain valid. Beyond the affine paradigm, for instance for the rough Bergomi model, this question falls outside the scope of our work and it is an interesting topic for future research.

The rest of the paper is organized as follows. In Section 2 we introduce the setup and state our main result of convergence, namely Theorem 2.7. Section 3 contains the results on the adjusted forward process and the conditional Fourier-Laplace transform necessary for the proof of the main theorem. The proof of the main theorem is presented in Section 4. In Section 5, within the framework of the rough Heston model, we provide numerical illustrations of the convergence and behavior of Bermudan put option prices. Appendix A explains some properties of the Riccati equations appearing in the expression of the conditional Fourier-Laplace transform. In Appendix B we provide results on the kernel approximation which guarantee certain hypotheses appearing in our main theorem.

Notation

We denote by $\mathcal{L}^2_{loc}$ the space of real-valued locally square integrable functions on $\mathbb{R}_+$. Similarly, given $T > 0$, $\mathcal{L}^2(0,T)$ stands for the space of real-valued square integrable function on the interval $(0,T)$. The space $\mathcal{C}(X,Y)$, where $X, Y \subseteq \mathbb{C}$, is the space of continuous functions from X to Y, with the conventions $\mathcal{C}(X,\mathbb{R}) = \mathcal{C}(X)$ and $\mathcal{C} = \mathcal{C}(\mathbb{R}_+)$. We use the same conventions for $\mathcal{C}_b$, $\mathcal{C}_b^2$, $\mathcal{C}_c$, $\mathcal{H}^\beta$, $\mathcal{B}$ and $\mathcal{B}_c$, which are the spaces of bounded continuous functions, bounded continuous functions with bounded and continuous derivatives up to order two, continuous functions with compact support, Hölder continuous functions of any order less than β, bounded functions and bounded functions with compact support, respectively. We write Δ for the shift operator, i.e. $\Delta_\varepsilon f = f(\cdot + \epsilon)$. For a function h on $\mathbb{R}$ we denote its support by $\mathrm{supp}(h)$. Given a function K and a measure L of locally bounded variation, we let $K * L$ be the convolution $(K * L)(t) = \int_{[0,T]} K(t-s)L(\mathrm{d}s)$, whenever the integral is well-defined. If F is a function on $\mathbb{R}_+$, we define $K * F = K * (F\,\mathrm{d}s)$.

2 Setup and main result

2.1 The model

We consider a Volterra Heston stochastic volatility model as in [3, 5]. In this model, under a risk-neutral measure, the asset's log price X and spot variance V are

$$\begin{aligned} X_t &= X_0 + \int_0^t \left(r - \frac{V_s}{2}\right) \mathrm{d}s + \int_0^t \sqrt{V_t} \left(\rho \, \mathrm{d}W_s + \sqrt{1-\rho^2} \, \mathrm{d}W_s^\perp\right), \\ V_t &= v_0(t) - \lambda \int_0^t K(t-s) V_s \, \mathrm{d}s + \eta \int_0^t K(t-s) \sqrt{V_s} \, \mathrm{d}W_s. \end{aligned} \tag{2.1}$$

In these equations, $X_0 \in \mathbb{R}$ is the initial log price, $(W, W^\perp)$ is a two-dimensional Brownian motion, r is the risk-free rate, and $\rho \in [-1,1]$ is a correlation parameter. The variance process V is a Volterra square root process. The constant $\lambda \geq 0$ is a parameter of mean reversion speed and $\eta \geq 0$ is the volatility of volatility. The kernel K is in $\mathcal{L}^2_{loc}$ and the function v_0 is in $\mathcal{C}$. Observe that – for fixed X_0, interest rate r, and correlation parameter ρ – the log price process X is completely determined by the variance process V and the Brownian motion $(W, W^\perp)$. Proposition 2.3 gives sufficient conditions ensuring the existence and uniqueness of weak solutions to the stochastic Volterra equation of the variance process.

Following the setting in [5], we introduce a subset $\mathcal{K}$ of $\mathcal{L}^2_{loc}$ in which we will consider the kernels.

Definition 2.1. *Let $K \in \mathcal{L}^2_{loc}$. We write $K \in \mathcal{K}$ if the following holds:*

(i) *There exist a constant $\gamma \in (0, 2]$ and a locally bounded function $c_K : \mathbb{R}_+ \to \mathbb{R}_+$ such that*

$$\int_0^\varepsilon |K(t)|^2 \,\mathrm{d}t + \int_0^{T-\varepsilon} |K(t+\varepsilon) - K(t)|^2 \,\mathrm{d}t \le c_K(T)\varepsilon^\gamma \tag{2.2}$$

for every $T > 0$ and $0 < \varepsilon \le T$.

(ii) *K is non-identically zero, non-negative, non-increasing, continuous on $(0, \infty)$ and admits a so-called resolvent of first kind L.*[1] *In addition, L is non-negative and*

$$\text{the function } s \mapsto L([s, s+t]) \text{ is non-increasing on } \mathbb{R}_+$$

for every $t > 0$.

Inspired by [3], we specify the space of functions in which we will take the functions v_0. For a given kernel $K \in \mathcal{K}$, with associated constant γ as in (2.2), let

$$\mathcal{G}_K = \{g \in \mathcal{H}^{\gamma/2} :\ g(0) \ge 0,\ \Delta_\varepsilon g - (\Delta_\varepsilon K * L)(0)g - \mathrm{d}(\Delta_\varepsilon K * L) * g \ge 0 \text{ for all } \varepsilon \ge 0\}. \tag{2.3}$$

The space $\mathcal{G}_K$ is stochastically invariant with respect to the adjusted version of the forward variance defined in Section 3.1, and it plays a crucial role in our arguments.

Throughout our study we will make the following assumption.

Assumption 2.2. *The kernel K and the function v_0 satisfy:*

(i) *$K \in \mathcal{K}$ and $\Delta_\varepsilon K$ satisfies (ii) in Definition 2.1 for all $\varepsilon \ge 0$.*

(ii) *$v_0 \in \mathcal{G}_K$.*

The existence and uniqueness in law for the stochastic Volterra equation of the variance process in (2.1) is guaranteed by the following proposition.

Proposition 2.3. *Suppose that Assumption 2.2 holds. Then the stochastic Volterra equation for the variance process V in (2.1) has a unique $\mathbb{R}_+$–valued weak solution. Furthermore, the trajectories of V belong to $\mathcal{H}^{\gamma/2}$ and given $p \ge 1$*

$$\sup_{t \in [0,T]} \mathbb{E}[|V_t|^p] \le c, \quad T > 0, \tag{2.4}$$

where $c < \infty$ is a constant that only depends on $p, T, \lambda, \eta, \gamma, c_K$ and $\|v_0\|_{\mathcal{C}[0,T]}$.

[1] This is a real-valued measure L of locally bounded variation on $\mathbb{R}_+$ such that $K * L = 1$.

Proof. This result follows from [3, Theorems 2.1 and 2.3], with the exception of the last assertion on the bound (2.4). Following the argument in the proof of [5, Lemma 3.1], this bound can be shown to depend on p, T, λ, η, $\|v_0\|_{\mathcal{C}[0,T]}$ and L^2–continuously on $K|_{[0,T]}$. Note that, thanks to the Fréchet-Kolmogorov theorem, the set of restrictions $K|_{[0,T]}$ of non-increasing kernels satisfying the property (2.2) for a given c_K and γ is relatively compact in $L^2(0,T)$. Maximizing the bounds over all such K yields a bound $c < \infty$ that only depends on $p, T, \lambda, \eta, \gamma, c_K$ and $\|v_0\|_{\mathcal{C}[0,T]}$. ☐

The theoretical results of this study are stated for general kernels K and functions v_0 satisfying Assumption 2.2. This is convenient in order to keep the notation simple. It is also in tune with forward-type stochastic volatility models, such as the rough Bergomi model [9]. Indeed, thanks to (2.4), taking expectations in the equation for the variance process in (2.1) yields the following relation between the function v_0 and the initial forward-variance curve $(\mathbb{E}[V_t])$

$$v_0(t) = \mathbb{E}[V_t] + \lambda \int_0^t K(t-s)\mathbb{E}[V_s]\,\mathrm{d}s.$$

For the numerical illustrations in Section 5 we will use the setting of the rough Heston model [21], which we summarize in the following example.

Example 2.4. In the rough Heston model, the kernel K is a fractional kernel

$$K(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)}, \tag{2.5}$$

with $\alpha \in \left(\frac{1}{2}, 1\right]$, and the function v_0 is of the form

$$v_0(t) = V_0 + \lambda \overline{\nu} \int_0^t K(s)\,\mathrm{d}s, \tag{2.6}$$

where $V_0 \geq 0$ is an initial variance and $\overline{\nu} \geq 0$ is a long term mean reversion level. Assumption (2.2), with $\gamma = 2\alpha - 1$, holds in this framework thanks to [5, Examples 2.3 and 6.2] and [3, Example 2.2].

Assume that Assumption 2.2 holds. Let $\mathbb{P}$ be the probability measure and $\mathbb{F} = (\mathcal{F}_t)$ be the filtration of the stochastic basis associated to the weak solution (X, V) to (2.1). Suppose that $f \in \mathcal{C}_b(\mathbb{R})$. Our goal is to determine the value process $(P_t)_{0 \leq t \leq T}$ of the American option with payoff process $(f(X_t))_{0 \leq t \leq T}$. We know that P is given by

$$P_t = \operatorname{ess\,sup}_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\left[\mathrm{e}^{-r(\tau-t)} f(X_\tau) | \mathcal{F}_t\right], \quad 0 \leq t \leq T, \tag{2.7}$$

where $\mathbb{E}$ is the expectation with respect to $\mathbb{P}$ and $\mathcal{T}_{t,T}$ denotes the set of $\mathbb{F}$-stopping times taking values in $[t, T]$. In order to compute American option prices, the financial model has to be approximated by more tractable models. In this work, we will consider approximations of the Volterra Heston model resulting from $\mathcal{L}^2$-approximations of the kernel. In the next section, we describe the approximation procedure.

2.2 Approximation of the kernel and the Volterra Heston model

We consider a sequence of kernels $(K^n)_{n\geq 1}$ in $\mathcal{L}^2_{loc}$ and functions $(v_0^n)_{n\geq 1}$ in $\mathcal{C}$. We make the following assumption.

Assumption 2.5. *The kernels $(K^n)_{n\geq 1}$ and the functions $(v_0^n)_{n\geq 1}$ satisfy:*

(i) *There exist a constant $\gamma \in (0,2]$ and a locally bounded function $c_K : \mathbb{R}_+ \to \mathbb{R}_+$ such that K^n satisfies (i) in Definition 2.1, for all $n \geq 1$.*

(ii) *$\Delta_\varepsilon K^n$ satisfies (ii) in Definition 2.1 for all $\varepsilon \geq 0$ and $n \geq 1$.*

(iii) *K^n converges to K in $\mathcal{L}^2_{loc}$.*

(iv) *$v_0^n \in \mathcal{G}_{K^n}$, with the constant γ of (i), for all $n \geq 1$ and v_0^n converges to v_0 in $\mathcal{C}$.*

According to Proposition 2.3, under Assumption 2.5, for each $n \geq 1$ there exists a unique weak solution (X^n, V^n) to

$$\begin{aligned} X_t^n &= X_0 + \int_0^t \left(r - \frac{V_s^n}{2}\right) \mathrm{d}s + \int_0^t \sqrt{V_t^n}\left(\rho\, \mathrm{d}W_s^n + \sqrt{1-\rho^2}\, \mathrm{d}W_s^{n,\perp}\right), \\ V_t^n &= v_0^n(t) - \lambda \int_0^t K^n(t-s) V_s^n\, \mathrm{d}s + \eta \int_0^t K^n(t-s)\sqrt{V_s^n}\, \mathrm{d}W_s^n, \end{aligned} \tag{2.8}$$

where $(W^n, W^{n,\perp})$ is a Brownian motion in the corresponding stochastic basis. Furthermore, given $p \geq 1$

$$\sup_{n\geq 1} \sup_{t\in[0,T]} \mathbb{E}^n[|V_t^n|^p] \leq c, \quad T > 0, \tag{2.9}$$

with a constant $c < \infty$ which can be chosen to depend only on $p, T, \lambda, \eta, \gamma, c_K$ and $\sup_{n\geq 1} \|v_0^n\|_{\mathcal{C}[0,T]}$, and where $\mathbb{E}^n$ denotes the expectation in the respective probability space. Moreover, the argument in the proof of [4, Theorem 3.6] shows that

$$(X^n, V^n) \text{ converges in law to } (X, Y) \text{ in } \mathcal{C}(\mathbb{R}_+, \mathbb{R}^2), \quad \text{as } n \to \infty. \tag{2.10}$$

This is a consequence of a more general result proven in Proposition 3.3.

For completely monotone kernels[2], an approximation with a sum of exponentials is natural. We briefly explain this procedure below.

2.2.1 Approximation with a sum of exponentials

Assume that the kernel K is completely monotone. By Bernstein's theorem this is equivalent to the existence of a non-negative Borel measure μ on $\mathbb{R}_+$ such that

$$K(t) = \int_{\mathbb{R}_+} \mathrm{e}^{-xt} \mu(\mathrm{d}x). \tag{2.11}$$

[2] K is completely monotone if $(-1)^m \frac{\mathrm{d}^m}{\mathrm{d}t^m} K(t) \geq 0$ for all non-negative integers m.

As in [4] and [15, 32], an approximation of the measure μ in (2.11) with a weighted sum of Dirac measures

$$\mu^n = \sum_{i=1}^{n} c_i^n \delta_{x_i^n} \tag{2.12}$$

yields a candidate approximation of the kernel

$$K^n(t) = \int_{\mathbb{R}_+} \mathrm{e}^{-xt} \mu^n(\mathrm{d}x) = \sum_{i=1}^{n} c_i^n \mathrm{e}^{-x_i^n t}. \tag{2.13}$$

The kernels $(K^n)_{n\geq 0}$ are completely monotone. If in addition they are not identically zero, as explained in [5, Example 6.2], condition (ii) in Assumption 2.5 holds.

The representation (2.13) yields the following factor-representation for the Volterra equation (2.8) satisfied by the variance process V^n

$$\begin{aligned} V_t^n &= v_0^n(t) + \sum_{i=1}^{n} c_i^n Y_t^{n,i}, \\ Y_t^{n,i} &= \int_0^t (-x_i^n Y_s^{n,i} - \lambda V_s^n)\,\mathrm{d}s + \int_0^t \eta \sqrt{V_s^n}\,\mathrm{d}W_s^n, \quad i = 1, \ldots, n. \end{aligned} \tag{2.14}$$

This representation is convenient because the process $(Y^{n,i})_{i=1}^n$ is an n-dimensional Markov process with an affine structure. This observation, together with the convergence in (2.10), was exploited in [4] in order to approximate European option prices in the rough Heston models employing Fourier methods. The affine structure will also play a crucial role in our study.

We now describe a natural way to determine the weights c_i^n and the points x_i^n. Let $(\eta_i^n)_{i=0}^n$ be a strictly increasing sequence in $[0,\infty)$ and define c_i^n and x_i^n as the mass and the center of mass of the interval $[\eta_{i-1}^n, \eta_i^n)$, i.e.

$$\begin{aligned} c_i^n &= \int_{[\eta_{i-1}^n, \eta_i^n)} \mu(\mathrm{d}x) = \mu([\eta_{i-1}^n, \eta_i^n)), \\ c_i^n x_i^n &= \int_{[\eta_{i-1}^n, \eta_i^n)} x \mu(\mathrm{d}x), \quad i = 1, \ldots, n. \end{aligned} \tag{2.15}$$

In Appendix B we provide sufficient conditions on the measure μ and the partitions $(\eta_i^n)_{i=0}^n$ that imply condition (i) in Assumption 2.5.

For the numerical illustrations in Section 5 we will use a fractional kernel and a geometric partition which we present in the following example.

Example 2.6. The fractional kernel (2.5) is completely monotone and in this case

$$\mu(\mathrm{d}x) = \frac{x^{-\alpha}}{\Gamma(1-\alpha)\Gamma(\alpha)}\,\mathrm{d}x.$$

Following [1], we consider the geometric partition $(\eta_i^n)_{i=0}^n$ given by $\eta_i^n = r_n^{i-\frac{n}{2}}$, for $r_n > 1$ such that

$$r_n \downarrow 1 \quad \text{and} \quad n \log r_n \to \infty, \quad \text{as } n \to \infty.$$

In this setting, the vectors (c_i^n) and (x_i^n) in (2.15) take the form

$$c_i^n = \frac{(r_n^{1-\alpha} - 1)}{\Gamma(\alpha)\Gamma(2-\alpha)} r_n^{(1-\alpha)(i-1-n/2)}, \qquad x_i^n = \frac{1-\alpha}{2-\alpha} \frac{r_n^{2-\alpha} - 1}{r_n^{1-\alpha} - 1} r_n^{i-1-n/2}, \qquad i = 1, \ldots, n. \tag{2.16}$$

Like in Example 2.4, along with the kernels $(K^n)_{n\geq 1}$, we consider functions $(v_0^n)_{n\geq 1}$ of the form

$$v_0^n(t) = V_0 + \lambda \overline{v} \int_0^t K^n(s)\, \mathrm{d}s.$$

Under this framework Assumption 2.5 holds[3]. Indeed, Remark B.3 in Appendix B shows that condition (i) holds. As explained in [5, Example 6.2], condition (ii) is a consequence of the complete monotonicity of K^n, $n \geq 1$. Condition (iii) is shown in [1, Lemma A.3]. This convergence and the considerations in Example 2.4 imply condition (iv).

With the setup of Example 2.6, since K^n is a $\mathcal{C}^1$-kernel and Assumption 2.5 holds, [4, Proposition B.3] implies that, for each $n \geq 1$, there exists a unique *strong solution* (X^n, V^n) to (2.8). Since in addition the factor process $(Y^{n,i})_{i=1}^n$ in (2.14) is a diffusion, classical discretization schemes can be used in order to simulate the trajectories of the variance and log price. Relying on this observation, the numerical study in Section 5 uses a simulation-based method in order to approximate American option prices in the rough Heston model. The convergence of the approximated prices is a consequence of the main theoretical findings of our study, which we present in the next section.

2.3 Main convergence result

We start by approximating the American option value process P in (2.7) with Bermudan option prices. More precisely, given a non-negative integer N, $T \geq 0$, a partition $(t_i)_{i=0}^N$ of $[0,T]$ with mesh π_N, and $t \in [0,T]$, we denote by $\mathcal{T}_{t,T}^N$ the set of $\mathbb{F}$-stopping times taking values in $[t,T] \cap \{t_0, ..., t_N\}$. For any $N \geq 0$, the Bermudan value process is then defined by

$$P_t^N = \operatorname{ess\,sup}_{\tau \in \mathcal{T}_{t,T}^N} \mathbb{E}\left[\mathrm{e}^{-r(\tau - t)} f(X_\tau) | \mathcal{F}_t\right], \quad 0 \leq t \leq T. \tag{2.17}$$

In addition, given $(X^n, V^n)_{n\geq 1}$ weak solutions to (2.8), we define the corresponding American option prices

$$P_t^n = \operatorname{ess\,sup}_{\tau \in \mathcal{T}_{t,T}^n} \mathbb{E}^n\left[\mathrm{e}^{-r(\tau - t)} f(X_\tau^n) | \mathcal{F}_t^n\right], \quad 0 \leq t \leq T \tag{2.18}$$

[3] In the case of a uniform partition $\eta_i^n = i\pi_n$, conditions that ensure (i)-(iii) in Assumption 2.5 are studied in [4].

and Bermudan option prices

$$P_t^{N,n} = \operatorname{ess\,sup}_{\tau \in \mathcal{T}_{t,T}^{N,n}} \mathbb{E}^n \left[\mathrm{e}^{-r(\tau - t)} f(X_\tau^n) | \mathcal{F}_t^n \right], \qquad 0 \leq t \leq T. \tag{2.19}$$

In the previous definitions, $(\mathcal{F}_t^n)$ is the filtration and $\mathbb{E}^n$ is the expectation on the stochastic basis associated to the weak solution to (2.8). The sets $\mathcal{T}_{t,T}^n$, $\mathcal{T}_{t,T}^{N,n}$ are defined similarly to $\mathcal{T}_{t,T}$, $\mathcal{T}_{t,T}^N$ on this stochastic basis.

Theorem 2.7 below is our main theoretical result. It implies, in particular, that the approximated American option prices P_0^n converge to the prices P_0 in the original Volterra Heston model.

Theorem 2.7. *Suppose that Assumptions 2.2 and 2.5 hold. Let (X, V) and (X^n, V^n) be the unique weak solutions to (2.1) and (2.8), respectively. For a function $f \in \mathcal{C}_b(\mathbb{R})$ define P, P^N, P^n and $P^{N,n}$ as in (2.7), (2.17), (2.18) and (2.19), respectively. Then*

$$P_{t_i}^{N,n} \text{ converges in law to } P_{t_i}^N \text{ as } n \to \infty, \quad N \geq 0,\, 0 \leq i \leq N. \tag{2.20}$$

Moreover, if $f \in \mathcal{C}_b^2(\mathbb{R})$ we have

$$\lim_{\pi_N \to 0} \sup_{n \geq 1} |P_0^{N,n} - P_0^n| = \lim_{\pi_N \to 0} |P_0^N - P_0| = 0 \tag{2.21}$$

and as a result

$$\lim_{n \to \infty} P_0^n = P_0. \tag{2.22}$$

Remark 2.8. *For American put option prices, the convergence stated in (2.22) can be deduced by approximating the payoff function in $\mathcal{C}_b(\mathbb{R})$ with functions $(f_n)_{n\geq 1}$ in the space $\mathcal{C}_b^2(\mathbb{R})$ such that f_n, f_n' and f_n'' are uniformly bounded in n.*

The proof of Theorem 2.7 is based on the study of the adjusted forward variance process and the associated Fourier-Laplace transform, which constitutes the main topic of the next section.

3 Conditional Fourier-Laplace transform

3.1 Adjusted forward process

In this section we study the adjusted forward process. This infinite-dimensional process was studied in [3] to characterize the Markovian structure of the Volterra Heston model (2.1). The adjusted forward process is very useful in order to study path-dependent options such as Bermudan and American options because, as we will see in Section 3.2, it allows us to better understand the conditional laws of the underlying process by means of the conditional Fourier-Laplace transform.

Assume that Assumption 2.2 holds. Let $\mathbb{P}$ be the probability measure and $\mathbb{F} = (\mathcal{F}_t)$ be the filtration of the stochastic basis associated to the weak solution (X, V) to (2.1). The adjusted forward process (v_t) of V is

$$v_t(\xi) = \mathbb{E}\left[V_{t+\xi} + \lambda \int_0^\xi K(\xi - s) V_{t+s}\, \mathrm{d}s | \mathcal{F}_t\right], \quad \xi \geq 0.^{4} \tag{3.1}$$

In particular, the variance process is embedded in the adjusted forward process because $v_t(0) = V_t$. Notice that, thanks to (2.4), the process $\left(\int_0^r K(t+\xi-s)\sqrt{V_s}\, \mathrm{d}W_s\right)_{0\leq r\leq t+\xi}$ is a martingale, and we can rewrite the adjusted forward process as

$$v_t(\xi) = v_0(t+\xi) + \int_0^t K(t+\xi-s)\left[-\lambda V_s\, \mathrm{d}s + \eta\sqrt{V_s}\, \mathrm{d}W_s\right], \quad \xi \geq 0. \tag{3.2}$$

Moreover, as shown in [3, Theorem 3.1], $v_t \in \mathcal{G}_K$ for all $t \geq 0$, i.e. $\mathcal{G}_K$ is stochastically invariant with respect to (v_t).

Similarly, if Assumption 2.5 holds, we can define the adjusted forward process for the approximating sequence $(V^n)_{n\geq 1}$ by

$$\begin{aligned} v_t^n(\xi) &= \mathbb{E}^n\left[V^n_{t+\xi} + \lambda \int_0^\xi K^n(\xi - s) V^n_{t+s}\, \mathrm{d}s | \mathcal{F}^n_t\right] \\ &= v_0^n(t+\xi) + \int_0^t K^n(t+\xi-s)\left[-\lambda V^n_s\, \mathrm{d}s + \eta\sqrt{V^n_s}\, \mathrm{d}W^n_s\right], \quad \xi \geq 0, \end{aligned} \tag{3.3}$$

and we have $v_t^n(0) = V_t^n$ and $v_t^n \in \mathcal{G}_{K^n}$, for all $t \geq 0$ and $n \geq 1$.

We start with a lemma regarding the regularity for the approximated adjusted forward processes v^n, $n \geq 1$.

Lemma 3.1. *Let $T, M \geq 0$ and $p > \max\{2, 4/\gamma\}$. Suppose that Assumption 2.5 holds and for $n \geq 1$ define the processes*

$$\tilde{v}_t^n(\xi) = v_t^n(\xi) - v_0^n(t+\xi)$$

with v^n as in (3.3). Then

$$\mathbb{E}^n[|\tilde{v}_t^n(\xi') - \tilde{v}_s^n(\xi)|^p] \leq C(\max(|t-s|, |\xi-\xi'|))^{p\gamma/2}, \quad (s,\xi), (t,\xi') \in [0,T]\times[0,M],$$

where C is a constant that only depends on $p, T, M, \lambda, \eta, \gamma, c_K$ and $\sup_{n\geq 1}\|v_0^n\|_{\mathcal{C}[0,T]}$. As a consequence $(\tilde{v}_t^n(\xi))_{(t,\xi)\in[0,T]\times[0,M]}$ admits an α-Hölder continuous version for any $\alpha < \frac{\gamma}{2}$. Moreover, for this version and for $\alpha < \frac{\gamma}{2} - \frac{2}{p}$ we have

$$\mathbb{E}^n\left[\left(\sup_{(t,\xi')\neq(s,\xi)\in[0,T]\times[0,M]} \frac{|\tilde{v}_t^n(\xi') - \tilde{v}_s^n(\xi)|}{|(t-s, \xi'-\xi)|^\alpha}\right)^p\right] < c, \tag{3.4}$$

where $c < \infty$ is a constant that only depends on $p, \alpha, T, M, \lambda, \eta, \gamma, c_K$ and $\sup_{n\geq 1}\|v_0^n\|_{\mathcal{C}[0,T]}$.

[4] We called (v_t) the *adjusted* forward process to distinguish it from the classical Musiela parametrization of the forward process $(\mathbb{E}[V_{t+\cdot}|\mathcal{F}_t])$.

Proof. Thanks to (3.3), we have for $s \le t$ and $\xi, \xi' \le M$

$$\begin{aligned}\tilde{v}_t^n(\xi') - \tilde{v}_s^n(\xi) &= \tilde{v}_t^n(\xi') - \tilde{v}_s^n(\xi') + \tilde{v}_s^n(\xi') - \tilde{v}_s^n(\xi) \\ &= \int_0^s (K^n(t+\xi'-u) - K^n(s+\xi'-u))\,\mathrm{d}Z_u^n + \int_s^t K^n(t+\xi'-u)\,\mathrm{d}Z_u^n \\ &\quad + \int_0^s (K^n(s+\xi'-u) - K^n(s+\xi-u))\,\mathrm{d}Z_u^n\end{aligned}$$

where $Z_t^n = -\lambda \int_0^t V_s^n\,\mathrm{d}s + \eta \int_0^t \sqrt{V_s^n}\,\mathrm{d}W_s^n$. From this point onwards, using Assumption 2.5 and the bound (2.9), the argument is analogous to the proof of [5, Lemma 2.4] and it is based on successive applications of Jensen and Burkhölder-Davis-Gundy inequalities, and Kolmogorov's continuity theorem; see [40, Theorem I.2.1]. □

Remark 3.2. *As an immediate consequence of Lemma 3.1, if Assumption 2.5 holds then*

$$\sup_{n\ge 1} \mathbb{E}^n\Big[\sup_{t\in[0,T]} V_t^n\Big] \le c, \tag{3.5}$$

where $c < \infty$ is a constant that only depends on $T, \lambda, \eta, \gamma, c_K$ and $\sup_{n\ge1} \|v_0^n\|_{\mathcal{C}[0,T]}$.

We are now able to establish the convergence of the approximated adjusted forward process in the next proposition.

Proposition 3.3. *Suppose that Assumptions 2.2 and 2.5 hold. Let X (resp. X^n) be as in (2.1) (resp. (2.8)) and let v (resp. v^n) be as in (3.1) (resp. (3.3)). Then, as n goes to infinity, $(X_t^n, v_t^n(\xi))_{(t,\xi)\in\mathbb{R}_+^2}$ converges in law to $(X_t, v_t(\xi))_{(t,\xi)\in\mathbb{R}_+^2}$ in $\mathcal{C}(\mathbb{R}_+^2, \mathbb{R}^2)$.*

Proof. This proof is similar to the proof of [4, Theorem 3.6 and Proposition 4.2]. We include a short explanation for completeness. Lemma 3.1 and Assumption 2.5(iv) imply tightness for the uniform topology of the triple (X^n, v^n, Z^n), where $Z_t^n = -\lambda \int_0^t V_s^n\,\mathrm{d}s + \eta \int_0^t \sqrt{V_s^n}\,\mathrm{d}W_s^n$. Suppose that (X, v, Z) is a limit point. Thanks to (3.3) and [2, Lemma 3.2], we have

$$\begin{aligned}1 * v^n(\xi) &= 1 * v_0^n(\xi + \cdot) + 1 * (\Delta_\xi K^n * dZ^n) \\ &= 1 * v_0^n(\xi + \cdot) + \Delta_\xi K^n * Z^n \\ &= 1 * v_0^n(\xi + \cdot) + \Delta_\xi K * Z^n + (\Delta_\xi K - \Delta_\xi K^n) * Z^n, \quad \xi \ge 0.\end{aligned} \tag{3.6}$$

In the previous identities, we have used the notation $\Delta_\xi K * dZ$ for the stochastic integral $(\Delta_\xi K * dZ)_t = \int_0^t K(t-s+\xi)dZ_s$. Assumption 2.5 and the convergence in law of (v^n, Z^n) towards (v, Z) yield

$$1 * v(\xi) = 1 * v_0(\xi + \cdot) + \Delta_\xi K * Z \quad \xi \ge 0.$$

One can show, as in [4, Theorem 3.6], that Z is of the form $Z_t = -\lambda \int_0^t V_s \, \mathrm{d}s + \eta \int_0^t \sqrt{V_s} \, \mathrm{d}W_s$ for some Brownian motion W, where $V = v(0)$. Once again, [2, Lemma 3.2] implies that

$$v_t(\xi) = v_0(\xi + t) + (\Delta_\xi K * dZ)_t, \quad t, \xi \geq 0.$$

Hence, $V = v(0)$ is the (unique) weak solution to the stochastic Volterra equation in (2.8) and v is the associated adjusted forward process. Furthermore, one can prove that (X, V) is the unique weak solution to (2.1). ☐

3.2 Conditional Fourier-Laplace transforms

This section studies the conditional Fourier-Laplace transform of the log price and the adjusted forward variance in the Volterra Heston model based on previous considerations in [36, 3, 18]. The results of this section will be useful to establish the convergence of Bermudan option prices in the approximated models to the Bermudan option prices in the original model, i.e. (2.20) in Theorem 2.7, using a dynamic programming approach.

We start by introducing some notation. For a kernel $K \in \mathcal{K}$ define

$$\mathcal{G}_K^* = \left\{ h \in \mathcal{B}_c(\mathbb{R}_+, \mathbb{C}) : t \mapsto -\mathrm{Re}\left(\int_0^\infty h(\xi) K(t + \xi) \, \mathrm{d}\xi \right) \in \mathcal{G}_K \right\} \tag{3.7}$$

with $\mathcal{G}_K$ as in (2.3). This space is a *dual space* that we will consider in the computation of the Fourier-Laplace transform of the adjusted forward process.

The next proposition characterizes the conditional Fourier Laplace transform of the log price X and the adjusted forward variance v through solutions of some Riccati equations.

Proposition 3.4. *Suppose that Assumption 2.2 holds, and let X be the log price process given by (2.1) and v be the adjusted forward process given by (3.1). Fix $T \geq 0$, $w \in \mathbb{C}$ with $\mathrm{Re}(w) \in [0, 1]$ and $h \in \mathcal{G}_K^*$. Then the conditional Fourier-Laplace transform of (X, v)*

$$L_t(w, h; X_T, v_T) = \mathbb{E}\left[\exp\left(w X_T + \int_0^\infty h(\xi) v_T(\xi) \, \mathrm{d}\xi \right) | \mathcal{F}_t \right], \quad t \leq T \tag{3.8}$$

can be computed thanks to the following formula

$$L_t(w, h; X_T, v_T) = \exp\left(w(X_t + r(T - t)) + \int_0^\infty \Psi(T - t, \xi; w, h) v_t(\xi) \, \mathrm{d}\xi \right), \tag{3.9}$$

where Ψ satisfies

$$\xi \mapsto \Psi(t, \xi; w, h) \in \mathcal{G}_K^*, \quad t \geq 0, \tag{3.10}$$

and it is a solution to the following Riccati equation

$$\Psi(t, \xi; w, h) = h(\xi - t) \mathbf{1}_{\{\xi \geq t\}} + \mathcal{R}\left(w, \int_0^\infty \Psi(t - \xi, z; w, h) K(z) \, \mathrm{d}z \right) \mathbf{1}_{\{\xi < t\}}, \quad t, \xi \geq 0, \tag{3.11}$$

and the operator $\mathcal{R}$ is defined by

$$\mathcal{R}(w,\varphi) = \frac{1}{2}(w^2 - w) + \left(\rho\eta w - \lambda + \frac{\eta^2}{2}\varphi\right)\varphi. \tag{3.12}$$

Moreover, if

$$\mathrm{Re}(w) = 0, \quad \int_0^\infty \mathrm{Re}(h(\xi))v_T(\xi)\,\mathrm{d}\xi \le 0$$

then

$$\int_0^\infty \mathrm{Re}(\Psi(T-t,\xi;w,h))v_t(\xi)\,\mathrm{d}\xi \le 0, \quad t \le T.$$

Remark 3.5. *Existence of solutions to equations (3.11) satisfying (3.10) is shown in Appendix A (see Propostion A.1). Notice that by setting*

$$\psi(t) = \int_0^\infty \Psi(t,\xi;w,h)K(\xi)\,\mathrm{d}\xi, \tag{3.13}$$

then the Riccati equation (3.11) can be recast as the following Riccati-Volterra equation for ψ

$$\psi(t) = \int_0^\infty h(\xi)K(t+\xi)\,\mathrm{d}\xi + (K * \mathcal{R}(w,\psi(\cdot)))(t), \tag{3.14}$$

and we have the identity

$$\Psi(t,\xi;w,h) = h(\xi - t)\mathbf{1}_{\{\xi \ge t\}} + \mathcal{R}(w,\psi(t-\xi))\,\mathbf{1}_{\{\xi < t\}}. \tag{3.15}$$

Proof of Proposition 3.4. Let Ψ be a solution to (3.11), satisfying (3.10) (see Proposition A.1). To simplify notation, throughout the proof we will omit the parameters w and h. Let Z be the semimartingale $Z_t = -\lambda\int_0^t V_s\,\mathrm{d}s + \eta\int_0^t \sqrt{V_s}\,\mathrm{d}W_s$, ψ be as in (3.13), and set

$$\theta = T - t, \quad \widetilde{Y}_t = \int_0^\infty \Psi(\theta,\xi)(v_t(\xi) - v_0(\xi + t))\,\mathrm{d}\xi.$$

The identity (3.2), equation (3.11), the stochastic Fubini theorem (see [39, Theorem 65]), and a change of variables yield

$$\begin{aligned}\widetilde{Y}_t &= \int_0^\infty \int_0^t \Psi(\theta,\xi)K(t+\xi-s)\,\mathrm{d}Z_s\,\mathrm{d}\xi \\ &= \int_0^t \int_\theta^\infty h(\xi-\theta)K(t+\xi-s)\,\mathrm{d}\xi\,\mathrm{d}Z_s + \int_0^t\int_0^\theta \mathcal{R}(\psi(\theta-\xi))K(t+\xi-s)\,\mathrm{d}\xi\,\mathrm{d}Z_s \\ &= \int_0^t \int_{T-s}^\infty h(\xi-T+s)K(\xi)\,\mathrm{d}\xi\,\mathrm{d}Z_s + \int_0^t\int_{t-s}^{T-s} \mathcal{R}(\psi(T-s-\xi))K(\xi)\,\mathrm{d}\xi\,\mathrm{d}Z_s.\end{aligned} \tag{3.16}$$

Equation (3.14) implies that

$$\psi(T-s) = \int_{T-s}^{\infty} h(\xi - T + s)K(\xi)\,\mathrm{d}\xi + \int_0^{T-s} \mathcal{R}(\psi(T-s-\xi))K(\xi)d\xi. \tag{3.17}$$

We then plugg (3.17) into (3.16) and obtain

$$\widetilde{Y}_t = \int_0^t \psi(T-s)\,\mathrm{d}Z_s - \int_0^t \int_0^{t-s} \mathcal{R}(\psi(T-s-\xi))K(\xi)\,\mathrm{d}\xi\,\mathrm{d}Z_s. \tag{3.18}$$

We deduce, thanks to (3.18) and the stochastic Volterra equation for the variance process, the following semimartingale dynamics for the process $\widetilde{Y}$

$$\mathrm{d}\widetilde{Y}_t = \psi(\theta)\,\mathrm{d}Z_t - \mathcal{R}(\psi(\theta))\int_0^t K(t-s)\,\mathrm{d}Z_s\,\mathrm{d}t = \psi(\theta)\,\mathrm{d}Z_t - \mathcal{R}(\psi(\theta))(V_t - v_0(t))\,\mathrm{d}t. \tag{3.19}$$

On the other hand, similar calculations show that

$$\int_0^{\infty} \Psi(\theta,\xi)v_0(\xi+t)\,\mathrm{d}\xi = \int_0^{\infty} h(\xi)v_0(\xi+T)\,\mathrm{d}\xi + \int_0^{\theta} \mathcal{R}(\psi(\xi))v_0(T-\xi)\,\mathrm{d}\xi. \tag{3.20}$$

Define the process Y as

$$Y_t = \widetilde{Y}_t + \int_0^{\infty} \Psi(\theta,\xi)v_0(\xi+t)\,\mathrm{d}\xi = \int_0^{\infty} \Psi(\theta,\xi)v_t(\xi)\,\mathrm{d}\xi.$$

From equation (3.19) and (3.20) we obtain the following semimartingale dynamics for Y

$$dY_t = \psi(\theta)\,\mathrm{d}Z_t - \mathcal{R}(\psi(\theta))V_t\,\mathrm{d}t. \tag{3.21}$$

Consider now the semimartingale

$$M_t = \exp(w(X_t - rt) + Y_t).$$

From equation (3.21) and Itô's formula, we obtain

$$\begin{aligned}\frac{\mathrm{d}M_t}{M_t} &= w\,\mathrm{d}X_t - wr\,\mathrm{d}t + \mathrm{d}Y_t + \frac{1}{2}w^2\,\mathrm{d}\langle X\rangle_t + \frac{1}{2}\,\mathrm{d}\langle Y\rangle_t + w\,\mathrm{d}\langle X,Y\rangle_t \\ &= -\frac{w}{2}V_t\,\mathrm{d}t + w\sqrt{V_t}\,\mathrm{d}B_t + \psi(\theta)\,\mathrm{d}Z_t - \mathcal{R}(\psi(\theta))V_t\,\mathrm{d}t \\ &\quad + \frac{1}{2}w^2V_t\,\mathrm{d}t + \frac{1}{2}\psi^2(\theta)\eta^2V_t\,\mathrm{d}t + \rho\eta w\psi(\theta)V_t\,\mathrm{d}t\end{aligned} \tag{3.22}$$

where $B = \rho W + \sqrt{1-\rho^2}W^{\perp}$. From the definition of $\mathcal{R}$ in (3.12), we finally get

$$\frac{dM_t}{M_t} = w\sqrt{V_t}\,\mathrm{d}B_t + \psi(T-t)\eta\sqrt{V_t}\,\mathrm{d}W_t. \tag{3.23}$$

M is then a local martingale and

$$M_T = \exp\left(w(X_T - rT) + \int_0^\infty h(\xi) v_T(\xi)\, \mathrm{d}\xi \right)$$

since $\Psi(0,\xi) = h(\xi)$. As pointed out in the proof of Propostion A.1 in Appendix A, thanks to the continuity of $\int_0^\infty h(\xi) K(\cdot + \xi)\, \mathrm{d}\xi$, the function ψ is a continuous, and hence bounded, function on $[0,T]$. Using a similar argument to the one used in [5, Lemma 7.3], we can show that M is a true martingale. This implies the formula for the Fourier-Laplace transform (3.9). The last implication in the statement of the proposition is a direct consequence of (3.9).

□

To establish the convergence of approximated Bermudan option prices, we will use convergence results of the conditional Fourier-Laplace transform, which we present in the following section.

3.3 Convergence of the Fourier-Laplace transform

Suppose that the kernels $(K^n)_{n\geq 1}$ and the functions $(v_0^n)_{n\geq 1}$ satisfy Assumption 2.5. Let $(X^n, V^n)_{n\geq 1}$ be the solutions to (2.8) and let $(v^n)_{n\geq 1}$ be the corresponding adjusted forward processes as in (3.3). We define, analogously to (3.8), the associated conditional Fourier-Laplace transform

$$L^n(w, h^n; X_T^n, v_T^n) = \mathbb{E}^n \left[\exp\left(w X_T^n + \int_0^\infty h^n(\xi) v_T^n(\xi)\, \mathrm{d}\xi \right) |\mathcal{F}_t^n \right] \tag{3.24}$$

with $h^n \in \mathcal{G}_{K^n}$ and $\mathrm{Re}(w) \in [0,1]$. Proposition (3.4) implies that

$$L_t^n(w, h^n; X_T^n, v_T^n) = \exp\left(w(X_t^n + r(T-t)) + \int_0^\infty \Psi^n(T-t, \xi; w, h^n) v_t^n(\xi)\, \mathrm{d}\xi \right), \tag{3.25}$$

where Ψ^n solves (3.11) with h replaced by h^n and K replaced by K^n. We have the following convergence result for the conditional Fourier-Laplace transforms.

Proposition 3.6. *Suppose that Assumptions 2.2 and 2.5 hold. Let X (resp. X^n) be as in (2.1) (resp. (2.8)) and let v (resp. v^n) be as in (3.1) (resp. (3.3)). Fix $T \geq 0$, $w \in \mathbb{C}$ with $\mathrm{Re}(w) \in [0,1]$, and $(h^n)_{n\geq 1}$ with $h^n \in \mathcal{G}_{K^n}^*$, $n \geq 1$. Assume that there is $M \geq 0$ such that*

$$\mathrm{supp}(h^n) \subseteq [0, M],\ n \geq 1; \quad \text{and} \quad h^n \to h \in \mathcal{G}_K^* \text{ in } \mathcal{B}([0,M], \mathbb{C}), \quad \text{as } n \to \infty.$$

Then

$$L^n(w, h^n; X_T^n, v_T^n) \text{ converges in law to } L(w, h; X_T, v_T) \text{ in } \mathcal{C}[0,T], \quad \text{as } n \to \infty,$$

where $L(w, h; X_T, v_T)$ and $L^n(w, h^n; X_T^n, v_T^n)$ are the conditional Fourier-Laplace transforms defined in (3.8) and (3.24), respectively.

The proof of Proposition 3.6 is based on Proposition 3.3 and the following lemma, whose proof can be found in Appendix A.

Lemma 3.7. *Assume that the hypotheses of Proposition 3.6 hold. Let Ψ (resp. Ψ^n) be solutions to the Riccati equation (3.11) with kernel K (resp. K^n) and initial condition h (resp. h^n). Define*

$$\psi(t) = \int_0^\infty \Psi(t,\xi;w,h)K(\xi)\,\mathrm{d}\xi, \quad \psi^n(t) = \int_0^\infty \Psi^n(t,\xi;w,h^n)K^n(\xi)\,\mathrm{d}\xi.$$

Then, as n goes to infinity, ψ^n converges to ψ in $\mathcal{C}[0,T]$. Moreover, letting $\widetilde{M} = \max\{M,T\}$, the support of $\Psi^n(t,\cdot;w,h^n)$ is contained in $[0,\widetilde{M}]$ for all $n \geq 1$ and $t \leq T$, and $\Psi^n(t,\cdot;w,h^n)$ converges to $\Psi(t,\cdot;w,h)$ in $\mathcal{B}([0,\widetilde{M}],\mathbb{C})$ uniformly in $t \in [0,T]$.

Proof of Proposition 3.6. By Proposition 3.3 and Skorohod's representation theorem we can construct (X^n,v^n) and (X,v) on the same probability space such that, as n goes to infinity, (X^n,v^n) converges almost surely to (X,v) in $\mathcal{C}(\mathbb{R}_+^2,\mathbb{R}^2)$. This observation and Lemma 3.7 imply that

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \left| wX_t^n + \int_0^\infty \Psi^n(t,\xi;w,h^n)v_t^n(\xi)\,\mathrm{d}\xi - wX_t - \int_0^\infty \Psi(t,\xi;w,h)v_t(\xi)\,\mathrm{d}\xi \right| = 0, \quad \text{a.s.}$$

Hence, $wX^n + \int_0^\infty \Psi^n(\cdot,\xi;w,h^n)v_t^n(\xi)\,\mathrm{d}\xi$ converges in law to $wX + \int_0^\infty \Psi(\cdot,\xi;w,h)v_t(\xi)\,\mathrm{d}\xi$ in $\mathcal{C}[0,T]$. An application of the continuous mapping theorem with the exponential function, together with Proposition 3.4, yields the conclusion. ☐

We now possess all the elements necessary for the proof of Theorem 2.7.

4 Proof of the main convergence result

We break down the argument into different parts. We start by establishing, in the next section, the convergence of the Bermudan option prices as stated in (2.20). To this end, we will consider a more general payoff structure that is better suited for an inductive argument.

4.1 Convergence of Bermudan option prices

Throughout this section we will use the notation

$$\langle h,\hat{h}\rangle = \int_0^\infty h(\xi)\hat{h}(\xi)\,\mathrm{d}\xi$$

for $h \in \mathcal{B}_c(\mathbb{R}_+,\mathbb{C})$ and $\hat{h} \in \mathcal{C}$. In addition, for a given finite set of indices J, we define

$$\mathcal{D}_J = \{(x,(\eta_j)_{j\in J}) \in (\mathbb{R},\mathbb{C}^{\#J}) : \ \mathrm{Re}(\eta_j) \leq 0 \text{ for all } j \in J\}. \tag{4.1}$$

We will consider options with intrinsic payoff processes $(Z_t)_{0\le t\le T}$ defined as

$$Z_t = \begin{cases} f(X_t), & \text{for } 0 \le t < T, \\ g(X_T, (\langle h_j, v_T\rangle)_{j\in J}), & \text{for } t = T, \end{cases} \tag{4.2}$$

where v denotes the adjusted forward process (3.1), J is a finite set of indexes, $f \in \mathcal{C}_b(\mathbb{R})$, $g \in \mathcal{C}_b(\mathcal{D}_J)$, $h_j \in \mathcal{G}_K^*$ for all $j \in J$, and $(X_T, (\langle h_j, v_T\rangle)_{j\in J}) \in \mathcal{D}_J$. In this setting, the Bermudan option discrete value process over the grid $(t_i)_{i=0}^N$ takes the form

$$U_i^N = \operatorname{ess\,sup}_{\tau\in\mathcal{T}_{t_i,T}^N} \mathbb{E}\Big[\mathrm{e}^{-r(\tau-t_i)} Z_\tau | \mathcal{F}_{t_i}\Big], \quad 0 \le i \le N. \tag{4.3}$$

For the approximating models, and in an analogous manner, we will consider options with payoff processes $(Z_t^n)_{0\le t\le T}$ defined as

$$Z_t^n = \begin{cases} f(X_t^n), & \text{for } 0 \le t < T, \\ g(X_T^n, (\langle h_j^n, v_T^n\rangle)_{j\in J}), & \text{for } t = T, \end{cases} \tag{4.4}$$

where $h_j \in \mathcal{G}_{K^n}^*$, for all $j \in J$, and $(X_T^n, (\langle h_j^n, v_T^n\rangle)_{j\in J}) \in \mathcal{D}_J$. The Bermudan option discrete value process, in the approximated model and over the grid $(t_i)_{i=0}^N$, takes the form

$$U_i^{N,n} = \operatorname{ess\,sup}_{\tau\in\mathcal{T}_{t_i,T}^N} \mathbb{E}^n\Big[\mathrm{e}^{-r(\tau-t_i)} Z_\tau^n | \mathcal{F}_{t_i}^n\Big], \quad 0 \le i \le N. \tag{4.5}$$

The following is the main result of this section.

Theorem 4.1. *Suppose that Assumptions 2.2 and 2.5 hold. Let X (resp. X^n) be as in (2.1) (resp. (2.8)) and let v (resp. v^n) be as in (3.1) (resp. (3.3)). Fix $T \ge 0$, J a finite set of indexes, $f \in \mathcal{C}_b(\mathbb{R})$, $g \in \mathcal{C}_b(\mathcal{D}_J)$, and $(h^n)_{n\ge1}$ with $h^n \in \mathcal{G}_{K^n}^*$, $n \ge 1$. Assume that there is $M \ge 0$ such that*

$$\operatorname{supp}(h^n) \subseteq [0, M], \ n \ge 1; \quad \textit{and} \quad h^n \to h \in \mathcal{G}_K^* \textit{ in } \mathcal{B}([0, M], \mathbb{C}), \quad \textit{as } n \to \infty.$$

Then

$$U_i^{N,n} \textit{ converges in law to } U_i^N, \quad i = 0, \dots, N, \quad \textit{as } n \to \infty,$$

where U_i^N and $U_i^{N,n}$ are given by (4.3) and (4.5), respectively.

Proof. We prove the result by induction on the number of exercise dates $N + 1$.
Initialization: Assume that $N = 0$. We just have to prove that

$$\lim_{n\to+\infty} g(X_0, (\langle h_j^n, v_0^n\rangle)_{j\in J}) = g(X_0, (\langle h_j, v_0\rangle)_{j\in J}).$$

This follows from continuity of g on $\mathcal{D}_J$, because our hypotheses readily imply

$$\lim_{n\to+\infty} \langle h_j^n, v_0^n\rangle = \langle h_j, v_0\rangle.$$

Induction: Assume that the claim holds for Bermudan options with N exercise dates. We have to consider three different cases.

1) Suppose that g on $\mathcal{D}_J$ has the form

$$g(x,(\eta_j)_{j\in J}) = \mathrm{Re}\left(\sum_{k\in I} c_k \exp\left(i\left(\nu_k x + \sum_{j\in J}\beta_{j,k}\mathrm{Im}(\eta_j)\right) + \sum_{j\in J}\alpha_{j,k}\mathrm{Re}(\eta_j)\right)\right), \tag{4.6}$$

with I a finite set of indices, $c_k\in\mathbb{C}$, $\nu_k\in\mathbb{R}$, $\alpha_{j,k}\geq 0$, $\beta_{j,k}\in\mathbb{R}$. In this case the value of the option at maturity (in the original Volterra model) is

$$Z_T = \mathrm{Re}\left(\sum_{k\in I} c_k \exp(i\nu_k X_T + \langle y_k(0), v_T\rangle)\right),$$

with

$$y_k(0) = \sum_{j\in J}\alpha_{j,k}\mathrm{Re}(h_j) + i\sum_{j\in J}\beta_{j,k}\mathrm{Im}(h_j), \quad k\in I.$$

One can verify that for each $k\in I$, $y_k(0)\in\mathcal{G}_K^*$ thanks to the fact that $\alpha_{j,k}\geq 0$, $j\in J$, and the definition of $\mathcal{G}_K^*$ in (3.7) and $\mathcal{G}_K$ in (2.3). Since the process U^N discounted coincides with the Snell envelope of the discounted payoff process, we have

$$\begin{aligned} U_{N-1}^N &= \max\left(Z_{t_{N-1}},\ e^{-r(\Delta t_{N-1})}\mathbb{E}\left[U_T^N|\mathcal{F}_{t_{N-1}}\right]\right) \\ &= \max\left(f(X_{t_{N-1}}),\ e^{-r\Delta t_{N-1}}\mathbb{E}\left[g(X_T,(\langle h_j, v_T\rangle)_{j\in J}))|\mathcal{F}_{t_{N-1}}\right]\right), \end{aligned}$$

where $\Delta t_{N-1} = t_N - t_{N-1} = T - t_{N-1}$. According to the affine transform formula in Proposition 3.4, with w being purely imaginary, the value of the option at time $N-1$ is then

$$U_{N-1}^N = \max\left\{f(X_{t_{N-1}}), \mathrm{e}^{-r\Delta t_{N-1}}\mathrm{Re}\left(\sum_{k\in I} c_k \mathrm{e}^{i\nu_k(X_{t_{N-1}}+r\Delta t_{N-1})+\langle y_k(\Delta t_{N-1}), v_{t_{N-1}}\rangle}\right)\right\},$$

where $y_k(\Delta t_{N-1})\in\mathcal{G}_K^*$ is a solution at time Δt_{N-1} of the associated Riccati equation (with initial condition $y_k(0)$), $k\in I$. Similarly, in the approximated model, we have

$$U_{N-1}^{N,n} = \max\left\{f(X_{t_{N-1}}^n), \mathrm{e}^{-r\Delta t_{N-1}}\mathrm{Re}\left(\sum_{k\in I} c_k \mathrm{e}^{i\nu_k(X_{t_{N-1}}^n+r\Delta t_{N-1})+\langle y_k^n(\Delta t_{N-1}), v_{t_{N-1}}^n\rangle}\right)\right\},$$

where $y_k^n(\Delta t_{N-1})\in\mathcal{G}_{K^n}^*$ is a solution at time Δt_{N-1} of the associated Riccati equation with initial condition

$$y_k^n(0) = \sum_{j\in J}\alpha_{j,k}\mathrm{Re}(h_j^n) + i\sum_{j\in J}\beta_{j,k}\mathrm{Im}(h_j^n)\in\mathcal{G}_{K^n}^*.$$

Propositions 3.3 and 3.6 imply that $U_{N-1}^{N,n}$ converges in law to U_{N-1}^{N}. To prove that $U_i^{N,n}$ converges in law to U_i^N for $i = 0, \dots, N-2$, we apply Lemma 3.7 together with the induction hypothesis in the case of a Bermudan option with maturity t_{N-1}, N exercise dates and final payoff $\hat{g}(X_{t_{N-1}}, (\langle \hat{h}_k, v_{t_{N-1}} \rangle)_{k\in I})$ where, for $k \in I$, $\hat{h}_k = y_k(\Delta t)$ and

$$\hat{g}(x,(\eta_k)_{k\in I}) = \max\left\{ f(x),, \mathrm{e}^{-r\Delta t_{N-1}} \mathrm{Re}\left(\sum_{k\in I} c_k \exp\left(i\nu_k(x + r\Delta t_{N-1}) + \eta_k\right)\right)\right\}.$$

Notice that $(X_{t_{N-1}}, (\langle \hat{h}_k, v_{t_{N-1}} \rangle)_{k\in I}) \in \mathcal{D}_I$ thanks to the last implication in Proposition 3.4.

2) Assume now that g vanishes outside a compact set $\Gamma \subset \mathcal{D}_J$.
Let $\varepsilon > 0$. By tightness of the sequence (X_T^n, v_T^n), its convergence to (X_T, v_T), and the convergence of h_j^n to h_j for all $j \in J$, there exists a compact set $\Gamma' \subset \mathcal{D}_J$ such that $\Gamma \subset \Gamma'$ and

$$\mathbb{P}\left(\left((X_T, (\langle h_j, v_T\rangle)_{j\in J})\right) \notin \Gamma'\right) < \varepsilon, \quad \mathbb{P}^n\left(\left((X_T^n, (\langle h_j^n, v_T^n\rangle)_{j\in J})\right) \notin \Gamma'\right) < \varepsilon, \quad n \geq 1. \tag{4.7}$$

Furthermore, we can assume that there exists a constant $A > 0$ such that

$$\Gamma' = \left\{ (x, (\eta_j)_{j\in J}) \in \mathcal{D}_J : \ |x| + \max_{j\in J}(|\eta_j|) \leq A \right\}.$$

Let $\mathcal{A}$ be an algebra of functions defined as follows. We say that a function $\hat{g}$ on $\mathcal{D}_J$ belongs to $\mathcal{A}$ if it is of the form

$$\hat{g}(x,(\eta_j)_{j\in J}) = \mathrm{Re}\left(\sum_{k\in I} c_k \exp\left(2\pi i\left(\frac{n_k}{2A}x + \sum_{j\in J}\frac{m_{k,j}}{2A}\mathrm{Im}(\eta_j)\right) + \sum_{j\in J}\alpha_{j,k}\mathrm{Re}(\eta_j)\right)\right),$$

with I a finite set of indices, $c_k \in \mathbb{C}$, $\alpha_{j,k} \geq 0$, and n_k and $m_{k,j}$ integers. We also define the following compact subset of $\mathcal{D}_J$

$$\widetilde{\Gamma} = \left\{ (x, (\eta_j)_{j\in J}) \in \mathcal{D}_J : \ |x| + \max_{j\in J}|\mathrm{Im}(\eta_j)|) \leq A \right\}.$$

Notice that we have $\Gamma' \subset \widetilde{\Gamma}$ and, if we denote by $\mathcal{A}|_{\widetilde{\Gamma}}$ the restriction of all the functions in $\mathcal{A}$ to $\widetilde{\Gamma}$, $\mathcal{A}|_{\widetilde{\Gamma}}$ is a subset of $\mathcal{C}_0(\widetilde{\Gamma}, \mathbb{R})$ – the space of continuous functions that vanish at infinity – that satisfies the hypothesis of Stone-Weierstrass Theorem. Therefore, there exists $\hat{g} \in \mathcal{A}$ such that

$$\sup_{(x,(\eta_j)_{j\in J})\in\widetilde{\Gamma}} |g(x,(\eta_k)_{k\in I}) - \hat{g}(x,(\eta_k)_{k\in I})| \leq \varepsilon. \tag{4.8}$$

Now observe that for all $(x, (\eta_j)_{j\in J}) \in \mathcal{D}_J$, there exists $(x', (\eta'_j)_{j\in J}) \in \widetilde{\Gamma}$ such that $\hat{g}(x, (\eta_j)_{j\in J}) = \hat{g}(x', (\eta'_j)_{j\in J})$. Hence

$$\|\hat{g}\|_\infty \le \varepsilon + \|g\|_\infty, \tag{4.9}$$

where $\|\cdot\|_\infty$ denotes the sup norm on $\mathcal{D}_J$.

Denote by $\widehat{U}^N$ (resp. $\widehat{U}^{N,n}$) the value processes for the Bermudan options corresponding to the payoff process $\widehat{Z}$ (resp. $\widehat{Z}^n$) obtained by replacing g by $\widehat{g}$ in (4.2) (resp. (4.4)). As shown in the previous case, we already know that

$$\widehat{U}_i^{N,n} \text{ converges in law to } \widehat{U}_i^N \text{ for } i = 0, \dots, N-1. \tag{4.10}$$

Moreover, since the process U^N discounted coincides with the Snell envelope of the discounted payoff process, we have

$$|U_i^N - \widehat{U}_i^N| \le \mathbb{E}\left[|U_{i+1}^N - \widehat{U}_{i+1}^N| \middle| \mathcal{F}_{t_i}\right], \quad i = 0, \dots, N-1.$$

By iterating this inequality, we deduce

$$|U_i^N - \widehat{U}_i^N| \le \mathbb{E}\left[|g(X_T, \langle h_j, v_T\rangle)_{j\in J})) - \hat{g}(X_T, \langle h_j, v_T\rangle)_{j\in J}))||\mathcal{F}_{t_i}\right], \quad i = 0, \dots, N.$$

Therefore, thanks to the inequalities (4.7), (4.8) and (4.9),

$$\mathbb{E}\Big[|U_i^N - \widehat{U}_i^N|\Big] \le \varepsilon(1 + \|\hat{g}\|_\infty) \le \varepsilon(1 + \varepsilon + \|g\|_\infty), \quad i = 0, \dots, N. \tag{4.11}$$

Similarly we can prove that

$$\mathbb{E}^n\Big[|U_i^{N,n} - \widehat{U}_i^{N,n}|\Big] \le \varepsilon(1 + \varepsilon + \|g\|_\infty), \quad i = 0, \dots, N,\ n \ge 0. \tag{4.12}$$

Since ε is abitrary we conclude, using (4.10), (4.11) and (4.12), that $U_i^{N,n}$ converges in law to U_i^N, for $i = 0, \dots, N$.

3) Suppose now that g belongs to $\mathcal{C}_b(\mathcal{D}_J)$.
Let $\varepsilon > 0$ be arbitrary. As before, tightness of the sequence (X_T^n, v_T^n), its convergence to (X_T, v_T), and the convergence of h_j^n to h_j, $j \in J$, imply that there is a compact set $\Gamma \subset \mathcal{D}_J$ such that

$$\mathbb{P}\left(((X_T, (\langle h_j, v_T\rangle)_{j\in J})) \notin \Gamma)\right) < \varepsilon, \quad \mathbb{P}^n\left(((X_T^n, (\langle h_j^n, v_T^n\rangle)_{j\in J})) \notin \Gamma)\right) < \varepsilon, \quad n \ge 1. \tag{4.13}$$

Let $\varphi : \mathcal{D}_J \to [0,1]$ be a function of compact support such that $\varphi \equiv 1$ on Γ.

Denote $\overline{U}^N$ (resp. $\overline{U}^{N,n}$) the value processes for the Bermudan options corresponding to the payoff process $\overline{Z}$ (resp. $\overline{Z}^n$) obtained by replacing g by $\overline{g} = \varphi g$ in (4.2) (resp. (4.4)). As shown in the previous case, we already know that

$$\overline{U}_i^{N,n} \text{ converges in law to } \overline{U}_i^N \text{ for } i = 1, \dots, N-1. \tag{4.14}$$

Additionally, we have

$$\begin{aligned} \mathbb{E}\Big[|U_i^N - \overline{U}_i^N|\Big] &\leq \mathbb{E}\left[|g(X_T, \langle h_j, v_T \rangle)_{j \in J})) - \overline{g}(X_T, \langle h_j, v_T \rangle)_{j \in J}))|\right] \\ &\leq \varepsilon \|g\|_\infty, \end{aligned} \tag{4.15}$$

and

$$\mathbb{E}^n\Big[|U_i^{N,n} - \overline{U}_i^{N,n}|\Big] \leq \varepsilon \|g\|_\infty. \tag{4.16}$$

Since ε is abitrary we conclude, from (4.14), (4.15) and (4.16), that $U_i^{N,n}$ converges in law to U_i^N, for $i = 0, \dots, N$.

□

4.2 Approximation of American options with Bermudan options

The following theorem establishes the convergence of Bermudan option prices towards American option prices and it is crucial in order to prove (2.21) in Theorem 2.7.

Theorem 4.2. *Suppose that Assumption 2.2 holds. Let (X, V) be the unique weak solution to (2.1). For a function $f \in \mathcal{C}_b^2(\mathbb{R})$ consider the American and Bermudan option prices given by (2.7) and (2.17), respectively. Then*

$$0 \leq P_0 - P_0^N \leq c \left(1 + \mathbb{E}\Big[\sup_{t \in [0,T]} V_t\Big]\right) \pi_N, \tag{4.17}$$

where π_N is the mesh of the partition $(t_i)_{i=0}^N$ and c is a constant that only depends on r, T and $\|f^{(m)}\|_{\mathcal{C}[0,T]}$, $m = 0, 1, 2$.

Proof. We obviously have $0 \leq P_0 - P_0^N$. Let $\varepsilon > 0$. There exists $\tau_\varepsilon^* \in \mathcal{T}_{0,T}$, ε-optimal in the sense that

$$P_0 \leq \mathbb{E}\Big[\mathrm{e}^{-r\tau_\varepsilon^*} f(X_{\tau_\varepsilon^*})\Big] + \varepsilon.$$

Now, we introduce the lowest stopping time taking values in $\{t_0, ..., t_N\}$, greater than τ_ε^*, this is

$$\tau_\varepsilon^{N,*} = \inf\{t_k : \; t_k \geq \tau_\varepsilon^*\}.$$

We have that $\tau_\varepsilon^{N,*}$ belongs to $\mathcal{T}_{0,T}^N$. Since the drift and the quadratic variation of X are affine in V, applying Itô's formula to the process $\left(\mathrm{e}^{-rt}f(X_t)\right)_{0\le t\le T}$ between τ_ε^* and $\tau_\varepsilon^{N,*}$ yields

$$
\begin{aligned}
P_0 - P_0^N &\le c\mathbb{E}\Big[\int_{\tau_\varepsilon^*}^{\tau_\varepsilon^{N,*}} (1+V_s)\,ds\Big] + \varepsilon && (4.18)\\
&\le c\mathbb{E}\Big[(\tau_\varepsilon^{N,*} - \tau_\varepsilon^*) \sup_{t\in[0,T]} (1+V_t)\Big] + \varepsilon && (4.19)\\
&\le c\left(1 + \mathbb{E}\Big[\sup_{t\in[0,T]} V_t\Big]\right)\pi_N + \varepsilon, && (4.20)
\end{aligned}
$$

where c is a constant that only depends on r, T and $\|f^{(m)}\|_{\mathcal{C}[0,T]}$, $m = 0,1,2$. Since $\varepsilon > 0$ was arbitrary, we deduce (4.17). □

We are now ready to prove our main theorem.

Proof of Theorem 2.7. The convergence in (2.20) is a direct consequence of Theorem 4.1. On the other hand, (3.5) and Theorem 4.2 yield (2.21). The limit (2.22) follows from (2.20) and (2.21). □

5 Numerical illustrations

In this section we illustrate with numerical examples the convergence and behavior of Bermudan put option prices in the approximated sequence of models. To this end, we consider the framework of the rough Heston model in Example 2.4 and the approximation scheme of Example 2.6.

We choose the same model parameters as in [1], namely

$$
V_0 = 0.02, \quad \bar{\nu} = 0.02, \quad \lambda = 0.3, \quad \eta = 0.3, \quad \rho = -0.7. \tag{5.1}
$$

We fix a maturity $T = 0.5$ and a spot interest rate $r = 0.06$.

In order to compute Bermudan option prices in the approximated model (X^n, V^n) in (2.8), we apply the Longstaff Schwartz algorithm [38] using 10^5 path simulations. Following the suggestion in [1], and based on the factor-representation (2.14), we simulate the trajectories of the variance with a *truncated* explicit-implicit Euler-scheme and the trajectories of the log price with an explicit Euler-scheme. More precisely, given a uniform partition $(s_k)_{k=0}^{N_{time}}$ of $[0,T]$ of norm Δt, and $\left(G_1^k\right)_{k\ge1}$ and $\left(G_2^k\right)_{k\ge1}$ independent sequences of independent centered and reduced gaussian variables, we simulate the log price with the scheme

$$
\widehat{X}^n_{s_{k+1}} = \widehat{X}^n_{s_k} + \left(r - \frac{\widehat{V}^n_{s_k}}{2}\right)\Delta t + \sqrt{\widehat{V}^{n\;+}_{s_k}}\sqrt{\Delta t}\left(\rho G_1^{k+1} + \sqrt{1-\rho^2}G_2^{k+1}\right), \quad \widehat{X}^n_{s_0} = X_0,
$$

and the variance with the scheme

$$\widehat{V}^n_{s_k} = v^n_0(s_k) + \sum_{i=1}^{n} c^n_i \widehat{Y}^{n,i}_{s_k}, \quad \widehat{Y}^{n,i}_0 = 0, \quad i = 1, \dots, n,$$

$$\widehat{Y}^{n,i}_{s_{k+1}} = \frac{1}{1 + x^n_i \Delta t}\left(\widehat{Y}^{n,i}_{s_k} - \lambda \widehat{V}^n_{s_k} \Delta t + \eta \sqrt{\widehat{V}^{n\,+}_{s_k}} \sqrt{\Delta t} G^{k+1}_1\right), \quad i = 1, \dots, n.$$

In this framework the initial curve v_0 in (2.6) takes the form

$$v^n_0(s_k) = V_0 + \lambda \bar{\nu} \sum_{i=1}^{n} c^n_i \left(\frac{1 - e^{-x^n_i s_k}}{x^n_i}\right).$$

We take $N_{time} = 500$ and select equidistant exercise times $(t_k)_{i=0}^N$, with $N = 50$, within the partition $(s_k)_{k=0}^{N_{time}}$. Given a strike price K, for the regressions of the Longstaff Schwartz algorithm we use the linear space of functions generated by functions with argument S, corresponding to the log price, and V corresponding to the volatility, of the form

$$f_1\left(\frac{S}{K}\right) f_2\left(\frac{V}{\bar{\nu}}\right), \quad f_1, f_2 \in \mathcal{A}$$

where $\mathcal{A}$ is given by

$$\mathcal{A} = \{1\} \cup \{\mathrm{e}^{-z} L_i(z) : i = 0, 1, 2\},$$

and L_i denotes the Laguerre polynomial of order i.[5]

To illustrate the convergence of options prices, we fix the parameter $\alpha = 0.6$ and choose parameters $r_n > 1$ in the kernel approximation such that

$$\begin{aligned} r_n &= \arg\min_r \|K - K^r\|^2_{\mathcal{L}^2(0,T)} \\ &= \arg\min_r \left(\sum_{i,j \le n} c^r_i c^r_j \frac{1 - \mathrm{e}^{-(x^r_i + x^r_j)T}}{x^r_i + x^r_j} - 2 \sum_{i \le n} c^r_i (x^r_i)^{-\alpha} \gamma(\alpha, T x^r_i)\right), \end{aligned} \tag{5.2}$$

where c^r_i, x^r_i, $i = 1, \dots, n$, are as in (2.16) with r_n replaced by r, K^r is the corresponding kernel obtained as a sum of exponentials, and $\gamma(\alpha, x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} \mathrm{e}^{-t}\, \mathrm{d}t$ is the lower incomplete gamma function. Table 1 contains the values of the parameter r_n along with the corresponding values of $\|K - K^n\|^2_{\mathcal{L}^2(0,T)}$ for $n = 4, 10, 20, 40, 200$. Figure 1 shows Bermudan put option prices for a strike $K = 100$, initial prices $S_0 = \exp(X_0)$ in $[93, 96]$,

[5]In the framework of our factor-approximation scheme, the prices of the Bermudan options at intermediate times are functions of the price S and the factors $(Y^{n,i})_{i=1}^n$ defined in (2.14). This functions could be approximated using neural network-based techniques similar to those in [37]. Our initial experiments, however, indicate that there is no significant gain in using this more complex approach. This is consistent with similar findings in [11] and [27] for American options prices in the rough Bergomi model.

n	r_n	norm_n^2
4	50.5458	0.3699
10	18.0548	0.1125
20	8.8750	0.0325
40	4.4737	0.0076
200	1.6946	1.1166e-04

Table 1: Values of r_n and $\text{norm}_n^2 = \|K - K^n\|^2_{\mathcal{L}^2(0,T)}$ obtained using (5.2) with $\alpha = 0.6$ and $T = 0.5$.

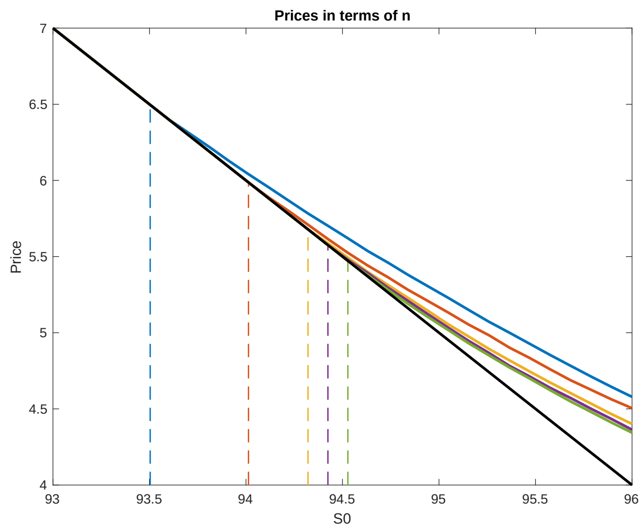

Figure 1: Bermudan put option prices in terms of n. Payoff (black), Heston model (blue), $n = 4$ (red), $n = 10$ (yellow), $n = 20$ (purple), $n = 40$ (green).

and $n = 4, 10, 20, 40$ number of factors. We also plot the prices obtained for the classical Heston model. For each set of prices we indicate the corresponding so-called critical price, this is the greatest value of the initial price for which the Bermudan option price is equal to the payoff. We observe that as n increases the option prices on this interval decrease and as a result the critical price increases. In Figure 2, we plot the critical-price as a function of the norm $\|K - K^n\|_{\mathcal{L}^2(0,T)}$ for $n = 1, 4, 10, 20, 40$, where $n = 1$ corresponds to the classical Heston model. Computing prices with $n = 200$ factors we observe the same critical price as with $n = 40$ which illustrates the convergence of the approximated models.

To study the behavior of Bermudan put option prices with respect to the parameter α, and taking into account our previous findings, we proxy the prices in the rough Heston model using the approximated model with $n = 40$ factors. We consider the same parameters

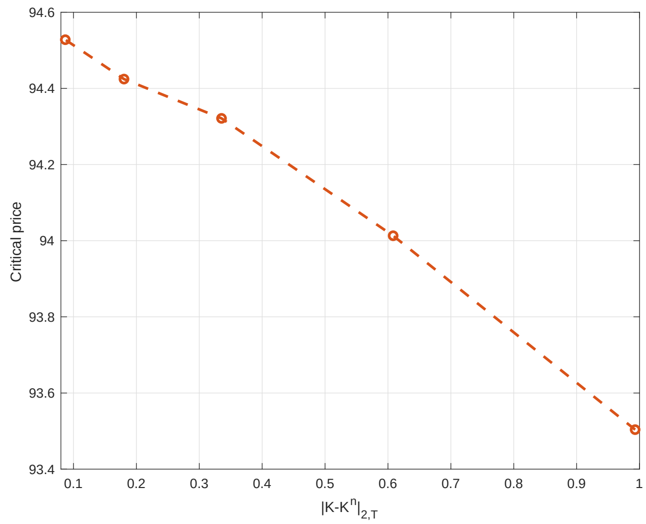

Figure 2: Critical prices as a function of $\|K - K^n\|_{\mathcal{L}^2(0,T)}$.

as in the previous example with the exception of α. The parameter r_{40} is chosen as in (5.2) depending on the parameter α of the fractional kernel K. We compute prices and critical prices for $\alpha = 0.6, 0.7, 0.8, 0.9, 1$. Figure 3 shows the Bermudan option prices obtained for these values of α and Figure 4 displays the critical price as a function of α. As α increases, we observe a similar behavior as the one obtained by increasing $\|K - K^n\|_{\mathcal{L}^2(0,T)}$ in our previous example. More precisely, as the regularity of the paths in the model increases, i.e. α increases, the prices of the option increase and the critical price decreases. This is consistent with similar findings reported in [34] within the context of the rough Bergomi model and it could be a consequence of the fact that for smaller values of α the variance has rougher paths and spends more time in a neighborhood of zero.

To illustrate the impact of the initial spot variance, we compare in Figure 5 the levels of the critical price for different values of V_0 in the rough Heston model with $\alpha = 0.6$ and the classical Heston model. The critical price seems to depend almost linearly on the initial spot variance V_0 in both the classical and the rough Heston model. In the rough Heston model the critical price, and hence the Bermudan option prices, appear to be slightly less sensitive to the initial level of the variance. This could be a result of the difference in sensitivity, with respect to V_0, of the time spent around zero by the trajectories in the classical and rough Heston models.

A theoretical explanation of our numerical findings would require more detailed results about the path-behavior of the rough Heston model, and their impact on American and Bermudan option prices. Such study falls outside of the scope of this manuscript and it could be an interesting topic of future research, along with a deeper numerical analysis of

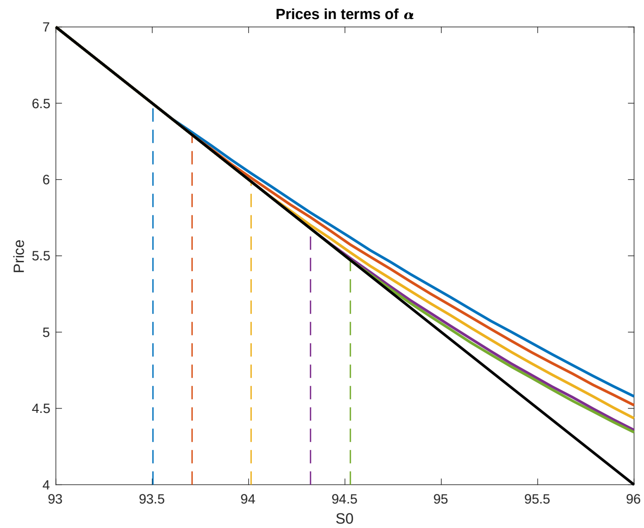

Figure 3: Bermudan put option prices and critical prices in terms of α. Payoff (black), $\alpha = 1$ (blue), $\alpha = 0.9$ (red), $\alpha = 0.8$ (yellow), $\alpha = 0.7$ (purple), $\alpha = 0.6$ (green).

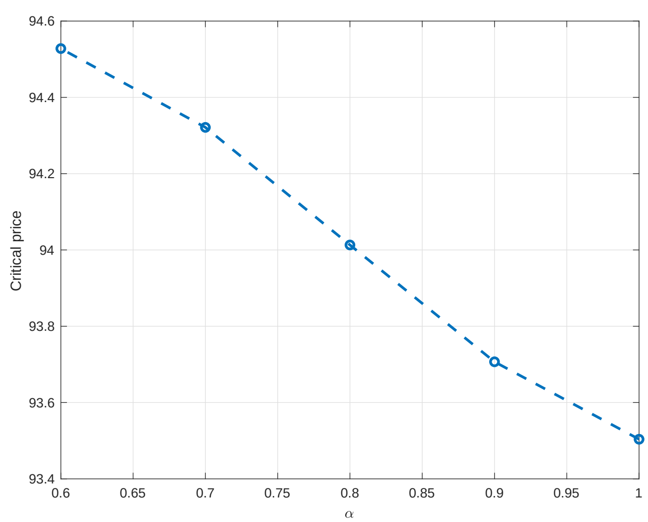

Figure 4: Critical prices as a function of α.

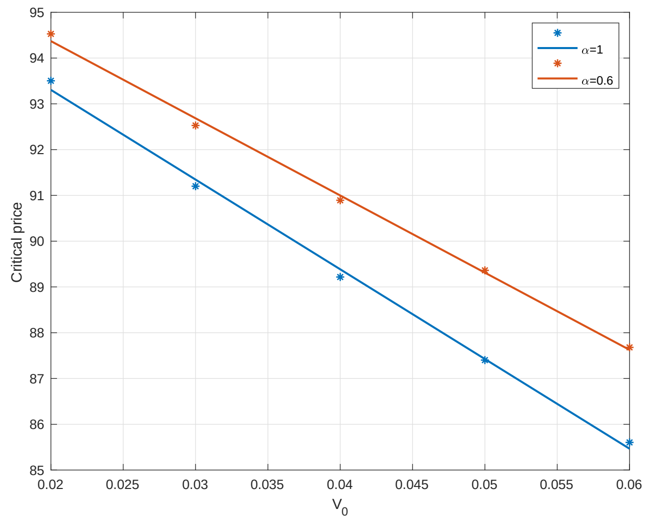

Figure 5: Critical prices for $\alpha = 0.6, 1$ and $V_0 = 0.02 + k * 0.01$, $k = 0, 1, 2, 3, 4$. The solid lines represent the linear regressions.

the behavior of American and Bermudan option prices in terms of the parameters of rough volatility models.

A Riccati-Volterra equations

Proposition A.1. *Suppose that $K \in \mathcal{L}^2_{loc}$ satisfies condition (i) in Assumption 2.2. Then, given $w \in \mathbb{C}$ with $\mathrm{Re}(w) \in [0,1]$, and $h \in \mathcal{G}^*_K$, the Riccati-Volterra equation (3.11) admits a solution Ψ such that $\Psi(t,\cdot;w,h) \in \mathcal{G}^*_K$ for all $t \geq 0$.*

Proof. As pointed out in Remark 3.5 the Riccati equation (3.11) for Ψ can be recast as the stochastic Volterra equation (3.14) for the function ψ given by (3.13). Thanks to the continuity of $\int_0^\infty h(\xi)K(\cdot+\xi)\,\mathrm{d}\xi$, [28, Theorem 12.1.1] implies the existence of a continuous solution ψ on a maximal interval $[0, T_{max})$. In order to prove that $T_{max} = \infty$, we can follow the proof of [5, Lemma 7.4]. In [5] the authors consider $\mathcal{L}^2$-solutions and a particular type of initial conditions for the Riccati-Voterra equations. In our case we consider continuous solutions and we have initial conditions of the form $\int_0^\infty h(\xi)K(t+\xi)\,\mathrm{d}\xi$ such that $-\int_0^\infty \mathrm{Re}(h(\xi))K(t+\xi)\,\mathrm{d}\xi \in \mathcal{G}_K$. The same arguments, however, can be adapted to our setting using the invariance result in [3, Theorem C.1] together with the fact that $\int_0^\infty f(\xi)K(t+\xi)\,\mathrm{d}\xi \in \mathcal{G}_K$ for all $f \in \mathcal{B}_c(\mathbb{R}_+, \mathbb{R}_+)$. Moreover, taking minus the real part in (3.14), [3, Theorem C.1] guarantees that

$$s \mapsto g_t(s) = \Delta_t g(s) - (\Delta_s K * \mathrm{Re}(\mathcal{R}(w,\psi)))(t) \in \mathcal{G}_K, \quad t \geq 0, \tag{A.1}$$

where $g(s) = -\int_0^\infty \mathrm{Re}(h(\xi))K(s+\xi)\,\mathrm{d}\xi$. We now define Ψ using (3.15), which satisfies (3.11) thanks to (3.14). The fact that $\Psi(t,\cdot;w,h) \in \mathcal{G}_K^*$, for all $t \geq 0$, is a consequence of (A.1) and the identity

$$\Delta_t g(s) - (\Delta_s K * \mathrm{Re}(\mathcal{R}(w,\psi)))(t) = -\int_0^\infty \mathrm{Re}(\Psi(t,\xi;w,h))K(s+\xi)\,\mathrm{d}\xi.$$

□

We finish this section with a sketch of the proof of Lemma 3.7.

Proof of Lemma 3.7. To prove the convergence of ψ^n towards ψ in $\mathcal{C}[0,T]$, one can use similar arguments as in the proof of [4, Theorem 4.1], replacing the zero initial condition by the initial curves $\int_0^\infty h(\xi)K(t+\xi)\,\mathrm{d}\xi$ and $\int_0^\infty h^n(\xi)K^n(t+\xi)\,\mathrm{d}\xi$, $n \geq 1$. The convergence of Ψ^n towards Ψ is a consequence of the identity (3.15), the convergence of (h^n, ψ^n) to (h,ψ), and the quadratic structure of $\mathcal{R}(w,\cdot)$. Since $\mathrm{supp}(h^n) \subseteq [0,M]$ for all $n \geq 1$, thanks to the form of the Riccati equations satisfied by Ψ^n, we conclude that the support of $\Psi^n(t,\cdot;w,h^n)$ is contained in $[0,\max\{T,M\}]$ for all $n \geq 1$ and $t \leq T$. □

B Some results on the kernel approximation

In this appendix we provide sufficient conditions on the kernel approximation which ensure condition (i) in Assumption 2.5.

Theorem B.1. *Suppose that μ is a non-negative Borel measure on $\mathbb{R}_+$ such that*

$$\int_{\mathbb{R}_+} (1 \wedge (\varepsilon x)^{-\frac{1}{2}})\mu(\mathrm{d}x) \leq c(T)\varepsilon^{\frac{\gamma-1}{2}}, \quad T > 0,\ \varepsilon \leq T,^{6} \tag{B.1}$$

with $\gamma \in (0,2]$ and $c : \mathbb{R}_+ \to \mathbb{R}_+$ a locally bounded function. If in addition

$$\sup_{n\geq 1} \sup_{i\in\{0,\ldots n-1\}} \frac{\eta_{i+1}^n}{\eta_i^n} < \infty \tag{B.2}$$

then the kernels $(K^n)_{n\geq 1}$ defined in (2.13), with $(c_i^n)_{i=1}^n$, $(x_i^n)_{i=1}^n$ given by (2.15), satisfy condition (i) in Assumption 2.5 with γ as in (B.1).

To prove Theorem B.1 we use the following lemma.

Lemma B.2. *Suppose that μ is a non-negative Borel measure μ on $\mathbb{R}_+$ such that (B.1) holds. Let K be the corresponding completely monotone kernel as in (2.11). Then K satisfies condition (i) in Definition 2.1, with the locally bounded function $2c^2$ and the same constant γ as in (B.1).*

[6] This condition was considered also in [3, Section 4].

Proof. Note that

$$\|K\|_{\mathcal{L}^2(0,\varepsilon)} \leq \int_0^\infty \|e^{-\cdot x}\|_{\mathcal{L}^2(0,\varepsilon)}\mu(dx) = \int_0^\infty \sqrt{\frac{1-e^{-2x\varepsilon}}{2x}}\mu(dx) \leq \varepsilon^{\frac{1}{2}} \int_0^\infty (1 \wedge (\varepsilon x)^{-\frac{1}{2}})\mu(dx).$$

This implies, by (B.1), that $\|K\|_{\mathcal{L}^2(0,\varepsilon)} \leq c(T)\epsilon^{\frac{\gamma}{2}}$, $\varepsilon \leq T$. A similar argument shows that $\|\Delta_\varepsilon K - K\|_{\mathcal{L}^2(0,T)} \leq c(T)\epsilon^{\frac{\gamma}{2}}$. The conclusion readily follows from these observations. □

Proof of Theorem B.1. According to Lemma B.2 it is enough to show that there is a locally bounded function $\tilde{c} : \mathbb{R}_+ \to \mathbb{R}_+$ such that for all $n \geq 1$

$$\int_{\mathbb{R}_+} (1 \wedge (\varepsilon x)^{-\frac{1}{2}})\mu^n(dx) \leq \tilde{c}(T)\varepsilon^{\frac{\gamma-1}{2}}, \quad T > 0,\, \varepsilon \leq T,$$

where μ^n is a sum of Dirac measures as in (2.12). This is a routine verification, using the definition of c_i^n, x_i^n in (2.15), Jensen's inequality, and conditions (B.1) and (B.2). For the sake of brevity, we omit the details. □

Remark B.3. *Let K be the fractional kernel* (2.5) *and consider the geometric partition $\eta_i^n = r_n^{i-\frac{n}{2}}$, $i = 0, \ldots, n$. It is easy to check that the hypotheses of Theorem B.1 hold with $\gamma = 2\alpha - 1$ as long as* $\sup_{n\geq 1} r_n < \infty$.

References

[1] E. Abi Jaber. Lifting the Heston model. *Quantitative Finance*, 19(12):1995–2013, 2019.

[2] E. Abi Jaber, C. Cuchiero, M. Larsson, and S. Pulido. A weak solution theory for stochastic volterra equations of convolution type. *arXiv preprint arXiv:1909.01166. To appear in Annals of Applied Probability*, 2019.

[3] E. Abi Jaber and O. El Euch. Markovian structure of the Volterra Heston model. *Statistics & Probability Letters*, 149:63–72, 2019.

[4] E. Abi Jaber and O. El Euch. Multifactor approximation of rough volatility models. *SIAM Journal on Financial Mathematics*, 10(2):309–349, 2019.

[5] E. Abi Jaber, M. Larsson, and S. Pulido. Affine Volterra processes. *The Annals of Applied Probability*, 29(5):3155–3200, 2019.

[6] E. Abi Jaber, E. Miller, and H. Pham. Linear–quadratic control for a class of stochastic Volterra equations: solvability and approximation. *arXiv preprint arXiv:1911.01900. To appear in Annals of Applied Probability.*, 2019.

[7] E. Abi Jaber, E. Miller, and H. Pham. Markowitz portfolio selection for multivariate affine and quadratic Volterra models. *Preprint : hal-02877569. To appear in SIAM Journal on Financial Mathematics*, 2020.

[8] A. Alfonsi and A. Kebaier. Approximation of Stochastic Volterra Equations with kernels of completely monotone type. *arXiv preprint arXiv:2102.13505*, 2021.

[9] C. Bayer, P. Friz, and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

[10] C. Bayer, J. Qiu, and Y. Yao. Pricing options under rough volatility with backward SPDEs. *arXiv preprint arXiv:2008.01241*, 2020.

[11] C. Bayer, R. Tempone, and S. Wolfers. Pricing American options by exercise rate optimization. *Quantitative Finance*, 20(11):1749–1760, 2020.

[12] S. Becker, P. Cheridito, and A. Jentzen. Deep optimal stopping. *Journal of Machine Learning Research*, 20:74, 2019.

[13] M. Bennedsen, A. Lunde, and M. S. Pakkanen. Hybrid scheme for Brownian semistationary processes. *Finance and Stochastics*, 21(4):931–965, 2017.

[14] M. Bennedsen, A. Lunde, and M. S. Pakkanen. Decoupling the Short- and Long-Term Behavior of Stochastic Volatility. *Journal of Financial Econometrics*, 2021.

[15] P. Carmona, L. Coutin, and G. Montseny. Approximation of some Gaussian processes. *Statistical inference for stochastic processes*, 3(1-2):161–171, 2000.

[16] F. Comte, L. Coutin, and E. Renault. Affine fractional stochastic volatility models. *Annals of Finance*, 8(2-3):337–378, 2012.

[17] L. Coutin and P. Carmona. Fractional Brownian motion and the Markov property. *Electronic Communications in Probability*, 3:12, 1998.

[18] C. Cuchiero and J. Teichmann. Generalized Feller processes and Markovian lifts of stochastic Volterra processes: the affine case. *Journal of Evolution Equations*, pages 1–48, 2020.

[19] O. El Euch, M. Fukasawa, and M. Rosenbaum. The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics*, 22(2):241–280, 2018.

[20] O. El Euch and M. Rosenbaum. Perfect hedging in rough Heston models. *The Annals of Applied Probability*, 28(6):3813–3856, 2018.

[21] O. El Euch and M. Rosenbaum. The characteristic function of rough Heston models. *Mathematical Finance*, 29(1):3–38, 2019.

[22] J.-P. Fouque and R. Hu. Optimal portfolio under fractional stochastic environment. *Mathematical Finance*, 29(3):697–734, 2019.

[23] M. Fukasawa. Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance*, 17(2):189–198, 2017.

[24] M. Fukasawa, T. Takabatake, and R. Westphal. Is volatility rough? *arXiv preprint arXiv:1905.04852*, 2019.

[25] J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018.

[26] J. Gatheral and M. Keller-Ressel. Affine forward variance models. *Finance and Stochastics*, 23(3):501–533, 2019.

[27] L. Goudenège, A. Molent, and A. Zanette. Machine learning for pricing American options in high-dimensional markovian and non-markovian models. *Quantitative Finance*, 20(4):573–591, 2020.

[28] G. Gripenberg, S.-O. Londen, and O. Staffans. *Volterra integral and functional equations*, volume 34 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1990.

[29] H. Guennoun, A. Jacquier, P. Roome, and F. Shi. Asymptotic behavior of the fractional Heston model. *SIAM Journal on Financial Mathematics*, 9(3):1017–1045, 2018.

[30] B. Han and H. Y. Wong. Mean–variance portfolio selection under Volterra Heston model. *Applied Mathematics & Optimization*, pages 1–28, 2020.

[31] B. Han and H. Y. Wong. Merton's portfolio problem under Volterra Heston model. *Finance Research Letters*, 39:101580, 2021.

[32] P. Harms and D. Stefanovits. Affine representations of fractional processes with applications in mathematical finance. *Stochastic Processes and their Applications*, 129(4):1185–1228, 2019.

[33] 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.

[34] B. Horvath, A. J. Jacquier, and A. Muguruza. Functional central limit theorems for rough volatility. *Available at SSRN 3078743*, 2017.

[35] T. Jaisson and M. Rosenbaum. Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *The Annals of Applied Probability*, 26(5):2860–2882, 2016.

[36] M. Keller-Ressel, M. Larsson, and S. Pulido. Affine rough models. *arXiv preprint arXiv:1812.08486*, 2018.

[37] B. Lapeyre and J. Lelong. Neural network regression for Bermudan option pricing. *arXiv preprint arXiv:1907.06474*, 2019.

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

[39] P. E. Protter. Stochastic differential equations. In *Stochastic integration and differential equations*, pages 249–361. Springer, 2005.

[40] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[41] S. E. Rømer. The hybrid-exponential scheme for stochastic Volterra equations. *Available at SSRN 3706253*, 2020.