



Siow Woon Jeng ^{†,‡} and Adem Kiliçman ^{*,†,‡}

Institute for Mathematical Research, Faculty of Science, University Putra Malaysia, Serdang 43400, Selangor, Malaysia; gs55636@student.upm.edu.my

* Correspondence: akilic@upm.edu.my

+ Current address: Faculty of Science and Institute for Mathematical Research, University Putra Malaysia, Serdang 43400, Selangor, Malaysia.

‡ These authors contributed equally to this work.

Abstract: The rough Heston model is a form of a stochastic Volterra equation, which was proposed to model stock price volatility. It captures some important qualities that can be observed in the financial market—highly endogenous, statistical arbitrages prevention, liquidity asymmetry, and metaorders. Unlike stochastic differential equation, the stochastic Volterra equation is extremely computationally expensive to simulate. In other words, it is difficult to compute option prices under the rough Heston model by conventional Monte Carlo simulation. In this paper, we prove that Euler's discretization method for the stochastic Volterra equation with non-Lipschitz diffusion coefficient $\mathbb{E}[|V_t - V_t^n|^p]$ is finitely bounded by an exponential function of t. Furthermore, the weak error $|\mathbb{E}[V_t - V_t^n]|$ and convergence for the stochastic Volterra equation are proven at the rate of $\mathcal{O}(n^{-H})$. In addition, we propose a mixed Monte Carlo method, using the control variate and multilevel methods. The numerical experiments indicate that the proposed method is capable of achieving a substantial costadjusted variance reduction up to 17 times, and it is better than its predecessor individual methods in terms of cost-adjusted performance. Due to the cost-adjusted basis for our numerical experiment, the result also indicates a high possibility of potential use in practice.

Keywords: rough Heston model; weak convergence error rate; Monte Carlo method; control variate method; multilevel Monte Carlo method

1. Introduction

Rough volatility has recently emerged as an impressive tool to model an asset's volatility in the financial market. The empirical study [1] by Gatheral and his co-authors have made all this possible. In short, rough volatility is a type of fractional Brownian motion (fBm) with a Hurst parameter of 0 < H < 0.5, or in terms of alpha notation, $0.5 < \alpha = H + 0.5 < 1$. Specifically, fBm with 0 < H < 0.5 has rougher fluctuation than the ordinary Brownian motion, whereas fBm with 0.5 < H < 1 has a smaller/lesser fluctuation than the ordinary Brownian motion. Before continuing the discussion on rough volatility, we would like to mention some of the work that made this discovery possible. First, before the introduction of rough volatility, the fractional Brownian motion with 0.5 < H < 1 was used to model the long-term memory effect in the asset's volatility by Comte and Renault [2]. Subsequently, mixed fractional Brownian motion, was proposed to also model the volatility of the asset return in the work by Cheridito et al. [3].

One of the earliest work that uses fractional Brownian motion with 0 < H < 0.5 as a volatility component in option pricing is by Alos et al. [4]. In particular, the authors managed to generate a short-time-to-maturity at-the-money-term-structure skew of the order $\mathcal{O}(T^c)$ for c > -0.5. This is a great accomplishment, as the individual local and stochastic volatility models are unable to replicate the mentioned behavior previously.



Citation: Siow, W.J.; Kiliçman, A. On Multilevel and Control Variate Monte Carlo Methods for Option Pricing under the Rough Heston Model. *Mathematics* **2021**, *9*, 2930. https:// doi.org/10.3390/math9222930

Academic Editor: Anatoliy Swishchuk

Received: 16 September 2021 Accepted: 11 November 2021 Published: 17 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Fukasawa [5] also demonstrated that option prices with volatility driven by the stochastic volatility model with the fractional Brownian motion of the Hurst parameter 0 < H < 0.5 can generate an at-the-money volatility skew of the order $\mathcal{O}(T^{H-0.5})$.

Due to recent advancements in rough volatility, there are many studies revolving around the development of rough volatility models. In particular, the authors of [6] proposed the rough Bergomi (rBergomi) model, which is a rough extension of the Bergomi model [7]. The result of the study found that the rBergomi model with fewer parameters fits the SPX volatility relatively better than Markovian stochastic volatility models; this is consistent with the SPX variance swap during the year 2008 economic crisis and the flash crash event. The rough Heston model is the main discussion of this paper and it was proposed by El Euch et al. in the work of [8,9]. What is so special about the rough Heston model is that it can fit the implied volatility of the options very well, and it can replicate the explosive behavior of the term-structure at-the-money skew observed in the financial market. Several studies [10–15] mainly focus on computational reduction of the rough Heston model is due to the lack of an explicit solution for the fractional Riccati equation, which is a part of the characteristic function of the option prices under the rough Heston model.

This paper contributes in three ways: (1) presents and proves a time bound on Euler's discretization method for the simulation of the stochastic Volterra equation $\mathbb{E}[|V_t - V_t^n|^p]$ with a non-Lipschitz diffusion coefficient; (2) proves the weak error rate $|\mathbb{E}[V_t - V_t^n]|$ and weak convergence of Euler's discretization method on the same stochastic Volterra equation; and (3) proposes a control variate estimator and a mixed Monte Carlo method to enhance the computational efficiency in computing the option price under the rough Heston model. Our work is possible mainly due to the work of [16-21]. The following Section 1.1 provides a simple and short introduction to our work. Section 2 mainly talks about discretization methods-the Euler-Maruyama (first-order Euler) method for the stochastic differential equation and stochastic Volterra equation, as well as the second-order Euler method for the stochastic differential equation. The bounded and weak errors of the Euler discretization method are also presented; the proofs are located in Appendices A–C. Furthermore, we discuss some existing Monte Carlo methods (base, control variate, and multilevel) and propose a mixed Monte Carlo (multilevel control variate) method in Section 3. Section 4 is dedicated to the numerical experiment of some existing methods and the proposed method. Specifically, we consider a few of the tests to benchmark our methods: (1) the capability of implied volatility to change on parameters ρ , α , and T; (2) cost-adjusted variation comparisons for OTM, ATM, and ITM options as well as the comparisons for different estimators; the (3) discretization bias for different estimators on a different number of time steps *n* and level *L*.

1.1. Prerequisites

The rough Heston model can be formulated as follows:

$$\frac{dS_t}{S_t} = \sigma_t \left[\rho dW_t^1 + \sqrt{1 - \rho^2} dW_t^2 \right] \tag{1}$$

and

$$\sigma_t^2 = \sigma_0^2 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \lambda [\theta - \sigma_s^2] ds + \frac{\nu}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \sigma_s dW_s^1, \tag{2}$$

where *S* is the stock price, ρ is the correlation between dW_t^1 and dW_t^2 , $0.5 < \alpha < 1$ is the parameter to control the roughness of the volatility movements, θ is the mean reversion level for the variance of return, $\lambda > 0$ is the speed at which the variance is reverting to its mean reversion level, $\nu > 0$ is the magnitude of random movement of the variance of the return.

Proposition 1. Assume the rough Heston model in Equation (2), and then the conditional expectation of the variance is as follows:

$$\mathbb{E}_0[\sigma_t^2] = \sigma_0^2 + (\theta - \sigma_0^2) \int_0^t f^{\alpha,\lambda}(s) ds,$$
(3)

where

$$\mathcal{L}^{\alpha,\lambda}(t) = \lambda t^{\alpha-1} E_{\alpha,\alpha}(-\lambda t^{\alpha}),$$
(4)

such that $E_{\alpha,\beta}(x) = \sum_{n=0}^{\infty} \frac{x^n}{\Gamma(\alpha n+\beta)}$ denotes the generalized Mittag–Leffler function.

Proof. It was proven in several studies—Proposition 3.1 [22], Proposition 2.1 [23], and page 7–8 of [24]. \Box

2. Discretization Methods

This section mainly discusses the first-order Euler discretization method for the stochastic differential equation and stochastic Volterra equation, and together with the second-order Euler discretization methods for the stochastic differential equation.

2.1. Euler-Maruyama Discretization Method

We first consider the stock process *S* satisfying the following stochastic differential equation (SDE):

$$dS_t = a(S_t)dt + b(S_t)dW_t,$$
(5)

where $a, b : \mathbb{R} \to \mathbb{R}$ are coefficient functions, and W is an independent standard Brownian motion. Specifically, the coefficient functions a and b are assumed to satisfy the usual conditions (see Appendix B.2 in [18]) for the existence and uniqueness of a strong solution to the SDE (5). The initial condition of SDE (5) is set as S(0).

First off, we denote the time-discretized approximation of the process S as \hat{S} . Under the Euler–Maruyama method [25], the initial condition remains the same as $\hat{S}_0 = S_0$, and the SDE (5) is discretized on a time grid $0 = t_0 < t_1 < ... < t_m$ as follows:

$$\widehat{S}_{t_{i+1}} = \widehat{S}_{t_i} + a(\widehat{S}_{t_i})[t_{i+1} - t_i] + b(\widehat{S}_{t_i})\sqrt{t_{i+1} - t_i}Z_{i+1},$$
(6)

where i = 0, ..., m - 1, and Z is an independent standard Brownian motion $Z \sim N(0, 1)$. In this work, we limit the method's time grid to linear fixed spacing such that $t_i = i\delta$, where δ is the fixed space. It is widely known that the Euler–Maruyama method for SDE has a strong error of $\mathcal{O}(\sqrt{\delta})$. Furthermore, it can be easily further refined to $\mathcal{O}(\delta)$, using Itô's formula to expand b(S(t)) [26] instead, but we will avoid further discussing this method, as we will introduce the second-order Euler method or Milstein scheme later. See Chapter 6 in [18] for more details.

Suppose now we consider the stochastic Volterra equation (SVE) of the following form:

$$V_t = V_0 + \int_0^t K_1(t,s)m(s,V_s)ds + \int_0^t K_2(t,s)g(s,V_s)dW_s, \quad t \in [0,T],$$
(7)

where $m, g : [0, T] \times \mathbb{R} \to \mathbb{R}$ and K_1, K_2 are the (possibly singular) kernels. We refer to the work of [21] on the discussion of discrete-time simulation of SVE for the Lipschitz condition. Notably, we can outline the following modified assumptions (one-dimensional space) stated in page 3 of [21] in order to provide theorems on the error of Euler's discretization method for SVE related to the rough Heston model (non-Lipschitz diffusion coefficient).

Assumption 1. We first denote the fixed space as $\delta := \Delta t_{i+1} = t_{i+1} - t_i \leq 1$, then let $n \geq 1$ such that $t_n = T$. Furthermore, we denote $\eta(s) := t_i$ for $s \in [t_i, t_{i+1})$ for $i \geq 0$ and again denote the following fixed constants $\beta \in (1, 1/(1-2H)), H \in (0, 1/2), \kappa \in (0, 1), C > 0$.

(AA1) For i = 1, 2, let $K_i(t, u) = 0$ when $u \ge t$, and the following holds:

$$\int_{0}^{t} \left(K_{i}(t,u)^{2} + K_{i}(t,u)^{2\beta} + K_{i}(t,\eta(u))^{2} + K_{i}(t,\eta(u))^{2\beta} \right) du < \infty$$
(8)

for all $t \in [0, T]$. (AA2) *For* i = 1, 2, and $s \le t$, the following holds:

$$\int_{s}^{t} K_{i}(t,u)^{2} ds + \int_{s}^{t} K_{i}(t,\eta(u))^{2} ds \leq C(t-s)^{2H}.$$
(9)

(AA3) For i = 1, 2, and $s \le t$, the following holds:

$$\int_{s}^{t} K_{i}(t,u)^{2\beta} ds + \int_{s}^{t} K_{i}(t,\eta(u))^{2\beta} ds \leq C(t-s)^{(2H-1)\beta+1}.$$
 (10)

(AA4) For i = 1, 2, the following holds:

$$\int_0^t [K_i(t+\delta,u) - K_i(t,u)]^2 ds + \int_0^t [K_i(t+\delta,\eta(u)) - K_i(t,\eta(u))]^2 du \le C\delta^{2H}.$$
 (11)

(AA5) For i = 1, 2, the following holds:

$$\int_0^t [K_i(t+\delta, u) - K_i(t, u)]^{2\beta} du \le C\delta^{(2H-1)\beta+1}.$$
(12)

(AA6) For i = 1, 2, the following holds:

$$\int_{0}^{t} [K_{i}(t,u) - K_{i}(t,\eta(u))]^{2} ds \leq C\delta^{2H}.$$
(13)

(AA7) For i = 1, 2, the following holds:

$$\int_0^t [K_i(t,u) - K_i(t,n(u))]^{2\beta} du \le C\delta^{(2H-1)\beta+1}.$$
(14)

(BB1) For $x, y \in \mathbb{R}$ and $s, t \in [0, T]$, let $|(m, g)(0, 0)| \leq C$ hold. The following inequalities hold:

$$|m(t,x) - m(t,y)| \le C|x - y|,$$
(15)

$$|g(t,x) - g(t,y)| \le C|x - y|^{\kappa},$$
(16)

$$|(m,g)(t,x)| \le C(1+|x|),$$
(17)

$$|(m,g)(t,x) - (m,g)(s,x)| \le C|t-s|^H(1+|x|).$$
(18)

Lemma 1. Suppose that we let $K_i(t, u) = (t - u)^{H-1/2}$ for i = 1, 2; then, it satisfies conditions (AA1)–(AA7).

Proof. See Appendix A. \Box

The SVE in the first-order Euler discretized form is as follows:

$$V_{t_{i+1}}^{n} = V_{t_{i}}^{n} + \sum_{k=0}^{i-1} (K_{1}(t_{i+1}, t_{k}) - K_{1}(t_{i}, t_{k}))m(t_{k}, V_{t_{k}}^{n})\Delta t_{k+1} + K_{1}(t_{i+1}, t_{i})m(t_{i}, V_{t_{i}}^{n})\Delta t_{i+1} + \sum_{k=0}^{i-1} (K_{2}(t_{i+1}, t_{k}) - K_{2}(t_{i}, t_{k}))g(t_{k}, X_{t_{k}}^{n})\Delta W_{k+1} + K_{2}(t_{i+1}, t_{i})g(t_{i}, V_{t_{i}}^{n})\Delta W_{i+1},$$
(19)

where $\Delta t_{i+1} = t_{i+1} - t_i = \delta$ and $\Delta W_{i+1} = W_{t_{i+1}} - W_{t_i}$ for i = 0, 1, ..., n - 1. Note that we use the following solution V_t^n of the Euler scheme (integral form) in the bounded proof of the Euler discretization method.

$$V_t^n = V_0 + \int_0^t K_1(t,\eta(u)) m(\eta(u), V_{\eta(u)}^n) du + \int_0^t K_2(t,\eta(u)) g(\eta(u), V_{\eta(u)}^n) dW_u.$$
(20)

Theorem 1. Suppose that Assumption 1, conditions (AA1)–(AA7) and condition (BB1) hold. Let $p \ge \frac{2\beta}{\kappa(\beta-1)}$ for $\beta \in (1, 1/(1-2H))$ and $\kappa \in (0, 1)$; then, there exists a finite constant C > 0 that depends on T, p, V_0 , and β such that for $t \in [0, T]$ and $n \to \infty$, the following is true:

$$\lim_{n \to \infty} \mathbb{E}[|V_t - V_t^n|^p] \le [C(\exp(Ct) - 1)]^{1/(1-\kappa)}.$$
(21)

Proof. Take the limit $n \to \infty$ in Theorem A1, and note that $\epsilon > 0$; then, the result follows directly. See Appendix B

Remark 1. We note that the Assumptions 1, conditions (AA1)–(AA7) and condition (BB1) satisfy the case of the rough Heston model. In addition, in the case of the rough Heston model, we have the non-Lipschitz diffusion coefficient factor of $\kappa = 1/2$. Finally, it can be seen that for small time t, the bounded errors are small too. On the other note, we do have to mention that in the case of the Lipschitz condition for the diffusion coefficient in the stochastic Volterra equation, *i.e.*, $|g(t, x) - g(t, y)| \leq C|x - y|$, $\lim_{n\to\infty} \mathbb{E}[|V_t - V_t^n|^p] = 0$ as proven in [21].

Theorem 2. Suppose that Assumption 1, conditions (AA1)–(AA7) and condition (BB1) hold. *Then, the following is true:*

$$\lim_{n \to \infty} |\mathbb{E}[V_t] - \mathbb{E}[V_t^n]| = 0.$$
(22)

Proof. Take the limit $n \to \infty$ in Theorem A2, then the result follows immediately. See Appendix C. \Box

Remark 2. Theorem 2 shows the weak convergence of the variance process in the rough Heston model. This was actually proven in [27], but here, we provide the weak error order for the variance process of the rough Heston model. From Theorem 2, we know that the variance process converges at the rate of $\delta^{(1/2\beta)[(2H-1)\beta+1]}$, where $\delta = T/n$ and $\beta \in (1,1/(1-2H))$. Since β can be specifically picked, our goal is to maximize the convergence order $(1/2\beta)[(2H-1)\beta+1]$ by changing β . We differentiate the order with respect to β as follows:

$$\frac{\partial}{\partial\beta} \left[\frac{1}{2\beta} ((2H-1)\beta+1) \right] = -\frac{1}{2\beta^2}.$$
(23)

From Equation (23), we know that $(1/2\beta)[(2H-1)\beta+1]$ is strictly decreasing on the range of $\beta \in (1, 1/(1-2H))$. Then, the weak error order can be obtained as follows:

$$\lim_{\beta \to 1} \frac{1}{2\beta} ((2H - 1)\beta + 1) = H.$$
(24)

This indicates that the weak error of the rough Heston model is $O(n^{-H})$.

2.2. Second-Order Euler Discretization Method

In this subsection, we discuss the second-order Euler discretization method that was introduced by Milstein [28] (we refer to the text [18]). Let us consider the stock process (5) in another manner, i.e., the following:

$$S_{t+\delta} = S_t + \int_t^{t+\delta} a(u, S_u) du + \int_t^{t+\delta} b(u, S_u) dW_u.$$
⁽²⁵⁾

We denote the following operators:

$$\mathcal{L}^{0} = \frac{\partial}{\partial t} + a \frac{\partial}{\partial s} + \frac{1}{2} b^{2} \frac{\partial^{2}}{\partial s^{2}}$$
(26)

and

$$\mathcal{L}^1 = b \frac{\partial}{\partial s}.$$
(27)

Then, the terms $\int_t^{t+\delta} a(S(u)) du$ and $\int_t^{t+\delta} b(S(u)) dW_u$ can be approximated as follows:

$$\int_{t}^{t+\delta} a(u,S(u))du \approx a(t,S_t)\delta + \mathcal{L}^0 a(t,S_t)I_1 + \mathcal{L}^1 a(t,S_t)I_2$$
(28)

and

$$\int_{t}^{t+\delta} b(u, S(u)) dW_{u} \approx b(t, S_{t}) [W(t+\delta) - W(t)] + \mathcal{L}^{0} b(t, S_{t}) I_{3} + \mathcal{L}^{1} b(t, S_{t}) I_{4}, \quad (29)$$

where I_i for i = 1, ..., 4 denotes the double integral, which can be computed as follows:

$$I_1 = \frac{1}{2}\delta^2,\tag{30}$$

$$I_3 = \delta \Delta W - I_2, \tag{31}$$

$$I_4 = \frac{1}{2} [(\Delta W)^2 - \delta], \tag{32}$$

where $\Delta W = W_{t+\delta} - W_t$. The computation of I_2 requires some derivation. Based on Chapter 6.2 in [18], I_2 can be simulated as follows:

$$\binom{\Delta W}{I_2} \sim N\left(0, \begin{pmatrix}\delta & \frac{1}{2}\delta^2\\ \frac{1}{2}\delta^2 & \frac{1}{3}\delta^3\end{pmatrix}\right),$$
(33)

The correlation between ΔW and I_2 can be computed as follows:

$$\rho_{\Delta W, I_2} = \frac{Cov(\Delta W, I_2)}{\sqrt{Var(\Delta W)Var(I_2)}}$$
(34)

$$=\frac{\frac{1}{2}\delta^2}{\sqrt{\delta}\sqrt{\frac{1}{3}\delta^3}}=\frac{\sqrt{3}}{2}.$$
(35)

Then, we let *B* denote an independent Brownian motion and $\Delta B \sim N(0, 1)$. Finally, we can obtain the following:

$$I_{2} = \sqrt{\frac{1}{3}\delta^{3}} \left[\rho_{\Delta W, I_{2}} \sqrt{\frac{1}{\delta}} \Delta W + \sqrt{1 - \rho_{\Delta W, I_{2}}^{2}} \Delta B \right]$$
(36)

$$=\sqrt{\frac{1}{3}\delta^{3}}\left[\frac{\sqrt{3}}{2}\sqrt{\frac{1}{\delta}}\Delta W + \sqrt{1-\frac{3}{4}}\Delta B\right]$$
(37)

$$=\frac{1}{2}\delta\Delta W + \frac{1}{2\sqrt{3}}\delta^{3/2}\Delta B.$$
(38)

See [18] for specific details of the derivation of Equations (30)–(33). It was shown by [29] that the second-order Euler discretization method has a weak order of error of $\mathcal{O}(\delta^2)$; it has a better convergence order than the first-order Euler discretization method, which possesses a weak error of $\mathcal{O}(\delta)$. However, the requirement that comes with the weak order of convergence on the second-order Euler discretization method requires the evaluated functions *a* and *b* to have uniformly bounded derivatives, and the functions need to be six times continuously differentiable. The purpose of giving a short review on this method is because while we will not be using this method for the discretization of the stochastic Volterra equation, we can actually use this on the discretization of the stock process in the control variate method and multilevel control variate method (control variate part) instead of the first-order Euler discretization method. We can reduce the overall error of the simulation in numerical experiment as per the weak error that the second-order Euler discretization method has suggested.

3. Monte Carlo Methods

This section outlines the conventional Monte Carlo method, control variate method, multilevel Monte Carlo methiod, and the novel multilevel control variate Monte Carlo method.

Suppose that the interest of our study is to estimate the $\mathbb{E}[P]$, the conventional Monte Carlo method estimates $\mathbb{E}[P]$, using *N* independent samples of ω with a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, that is, the following:

$$\widehat{P} = \frac{1}{N} \sum_{n=1}^{N} P(\omega^{(n)}).$$
(39)

The variance of this estimator is $\frac{1}{N} Var[P]$, and the root mean square error (RMS) of the estimator has an order of $\mathcal{O}(N^{-1/2})$. Particularly, suppose that we require an accuracy of ϵ for the RMS for the estimator; then, a total of $N = \mathcal{O}(\epsilon^{-2})$ samples are needed for the conventional Monte Carlo.

3.1. Control Variate Method

The control variate method is one of the variance reduction methods along with antithetic variates, quasi Monte Carlo, and importance sampling methods. The idea of the control variate method is somewhat simple. Suppose we consider another random variable *X* that is paired with *P* such that the pairs (X_n, P_n) for n = 1, ..., N are simulated with the sample path ω . Furthermore, it is required that $\mathbb{E}[X]$ is known, or it can be computed relatively cheaply and accurately. Then, we denote a fixed constant ς such that we can compute the following:

$$P_n(\varsigma) = P_n - \varsigma(X_n - \mathbb{E}[X]).$$
(40)

Equation (40) is a control variate estimator, and it is an unbiased estimator to *P* as follows:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} P_n(\varsigma) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} [P_n - \varsigma(X_n - \mathbb{E}[X])]$$
$$= \mathbb{E}[P].$$

To optimally reduce the variance or mean square error of the estimator (40), we can pick the constant ς to be the following:

$$\varsigma * = \frac{\sigma_P}{\sigma_X} \rho_{X,P} = \frac{Cov(X,P)}{Var(X)}.$$
(41)

The ratio of the control variate estimator to the base estimator is as follows:

$$\frac{Var[P - \varsigma(X - \mathbb{E}[X])]}{Var[P]} = 1 - \rho_{X,P}^2.$$

$$\tag{42}$$

Equation (42) tells us that as long as the correlation between the random variable X and P is negatively correlated or positively correlated, variance reduction is guaranteed.

Since the quantity of interest *P* is usually lacking $\mathbb{E}[P]$ in practice, we ought to use the optimal sample estimate of ς from Equation (41) as follows:

$$\widehat{\varsigma}^{*} = \frac{\sum_{n=1}^{N} (X_n - \overline{X}) (P_n - \overline{P})}{\sum_{n=1}^{N} (X_n - \overline{X})^2},$$
(43)

where \overline{X} and \overline{P} are the sample mean of X_n and P_n , respectively. Accordingly, the expectation $\mathbb{E}[P]$ can be estimated using the sample mean estimator $\overline{P}(\widehat{\varsigma}^*)$ such that the optimally variance reduced control variate estimator is as follows:

$$P^{CV} = \frac{1}{N} \sum_{n=1}^{N} \left[P_n - \hat{\varsigma}^* (X_n - \overline{X}) \right].$$
(44)

In practice, $\hat{\zeta}^*$ and \overline{X} are most likely unknown variables, so it should be estimated beforehand. Pre-running a small number of samples is usually the only way to estimate $\hat{\zeta}^*$ and \overline{X} since this study focuses on call option pricing under the rough Heston model. Obviously, $P = (S_T - K)^+$ is the payoff of the call option prices where S_T is the stock price at maturity time and K is the strike price. Now comes the question of what control variate we should employ in order to get $\rho_{X,P}$ to be closest to 1 or -1. We propose a control variate based on the idea of [19].

Proposition 2. Let S_0^{CV} be the initial stock price and W be an independent Brownian motion. Consider the stock process as follows:

$$dS_t^{CV} = \sigma(t)S_t^{CV}dW_t,\tag{45}$$

where $\sigma(t)$ is a nonrandom function of t. Then, there is an analytic solution for the European call option as follows:

$$C = \mathbb{E}[(S_T^{CV} - K)^+]$$

= $S_0^{CV}N(d_+) - KN(d_-)$

where $N(\cdot)$ denotes the cumulative distribution function of the standard normal distribution and

$$d_{\pm} = \frac{\log \frac{S_0^{CV}}{K} \pm \frac{1}{2}\bar{\sigma}_T^2 T}{\bar{\sigma}_T \sqrt{T}}$$

$$\tag{46}$$

such that $\bar{\sigma}_T^2 = \frac{1}{T} \int_0^T \sigma^2(t) dt$.

Proof. See page 253 of [30]. □

We propose a control variate X by using Proposition 1. We start by denoting

$$X_T = (S_T^{CV} - K)^+,$$
 (47)

where S_t^{CV} is the stock process governed by the following:

$$dS_t^{CV} = \sigma(t)S_t^{CV}dW_t.$$
(48)

Specifically, $\sigma(t)$ is a nonrandom function and the squared-function is denoted as follows:

$$\sigma^2(t) = \mathbb{E}_0(\sigma_{t+\varepsilon}^2),\tag{49}$$

where $\mathbb{E}_0(\sigma_t^2)$ is the expectation from Proposition 1, and $\varepsilon > 0$ is a small positive constant. The use of $t + \varepsilon$ in Equation (49) is to ensure that the conditions and requirements of the second-order Euler discretization method are being met. We set $\varepsilon = \delta/10$ for our study to avoid the singularity at $\sqrt{\mathbb{E}'_0(\sigma_t^2)}|_{t=0}$, where δ is the step-size of the discretization method. Hence, the expectation of the call option price can be determined using Proposition 2. In a numerical experiment, the forward variance of the rough Heston model (Equation (3)) in Proposition 1 is computed through the truncated Mittag–Leffler function and numerical integration.

3.2. Multilevel Monte Carlo

Multilevel Monte Carlo is a form of the control variate method which uses multiple controls. Specifically, it uses itself with different level of coarseness as control variates. Instead of simulating the process at the finest level, the method approximates P_L , using the sequence of $P_0, ..., P_{L-1}$. In particular, the sequence $P_0, ..., P_{L-1}$, P_L is a sequence with increasing cost to simulate. The expectation of $\mathbb{E}[P_L]$ can be decomposed as follows:

$$\mathbb{E}[P_L] = \mathbb{E}[P_0] + \sum_{i=1}^{L} \mathbb{E}[P_i - P_{i-1}].$$
(50)

Consequently, we can rely on the following unbiased estimator of $\mathbb{E}[P_L]$ in practice:

$$P^{ML} = \frac{1}{N_0} \sum_{k=1}^{N_0} P_0^{(0,k)} + \sum_{i=1}^L \frac{1}{N_i} \sum_{k=1}^{N_i} \left(P_i^{(i,k)} - P_{i-1}^{(i,k)} \right).$$
(51)

For our study, we let $M \ge 2$ and D > 1 such that the total discretization steps for $P_i^{(\cdot,\cdot)}$ is $n_i = M/D^{L-i} \in \mathbb{N}$ for i = 0, ..., L, then the step size is defined as $\delta_i = T/n_i$. One thing to take note of is that the difference $P_i^{(i,k)} - P_{i-1}^{(i,k)}$ in Equation (51) actually relies on the same sample path ω but at a different coarseness level. For example, if k is the step number of the simulation (same as in Equation (51)) and we have the following parameters M = 256, D = 2, L = 4, and i = 4, we would simulate one set of sample stochastic path $\omega^{(n_4)}$ with $n_4 = 256$ discretization steps (in other words, the simulated Brownian motions are W_a , a = 0, 1, ..., 256) to compute $P_4^{(4,k)}$, and using the same simulated set of sample path $\omega^{(n_4)}$, we would only use W_{2a} , a = 0, 1, ..., 128 to compute $P_3^{(4,k)}$. To achieve minimization of the variance of the multilevel estimator using optimal N_i with a fixed total cost, we need to solve an optimization problem. We first define V_0 and C_0 to be the variance and cost of one sample P_0 respectively, whereas we can define V_i and C_i to be the variance and cost of one sample $P_i - P_{i-1}$ respectively. The total variance of the multilevel estimator is as follows:

$$V(N_0, N_1, ..., N_L) = \sum_{i=0}^{L} \frac{1}{N_i} V_i,$$
(52)

and total cost of the multilevel estimator is as follows:

$$C(N_0, N_1, ..., N_L) = \sum_{i=0}^L N_i C_i.$$
(53)

Note that each level of correction in Equation (51) uses independent samples as indicated from the presence of the level *i* in superscript (*i*, *k*) of Equation (51), which is why Equation (52) is of this form (absence of correlation on other random variables). Our optimization problem is to do the following:

Minimize
$$V(N_0, N_1, ..., N_L)$$
, (54)

subject to
$$\Psi(N_0, N_1, ..., N_L) = C(N_0, N_1, ..., N_L) - C = 0,$$
 (55)

where C > 0 is a fixed total cost. The Lagrangian function can be constructed as follows:

$$L(N_0, ..., N_L, \lambda) = V(N_0, N_1, ..., N_L) + \lambda_L \Psi(N_0, N_1, ..., N_L),$$
(56)

where λ_L is the Langrange multiplier. Solving it, we can obtain the optimal N_i as follows:

$$N_i^* = \sqrt{\frac{V_i}{C_i}} \left(\frac{C}{\sum_{a=0}^L \sqrt{V_a C_a}}\right).$$
(57)

Using Equation (57), we are able to achieve the optimal least total variance with a fixed cost by distributing N_i^* to a different coarse level of $P_i^{(i,\cdot)}$. In practice, we will never know the exact V_i and C_i ; therefore, we will need to replace them with the estimates \hat{V}_i and \hat{C}_i . What we can do in practice is that we can start by pre-running a small number of the simulation to compute for \hat{V}_i and \hat{C}_i ; then, we can use \hat{V}_i and \hat{C}_i to obtain an estimate of N_i^* in Equation (57), i.e., \hat{N}_i^* .

3.3. Multilevel Control Variate Method

We propose a mixed estimator that uses the idea of the control variate method and the multilevel Monte Carlo method. Readers may also see similar work in [31,32]. We refer to this method as the multilevel control variate method from now onward. From Equation (50), we estimate $\mathbb{E}[P_0]$ and $\mathbb{E}[P_i - P_{i-1}]$ using the control variate method. Let $X_0, X_1, ..., X_L$ be the control variate estimator proposed in Equation (47) such that $\mathbb{E}[X_i]$ for i = 0, ..., L. Then, the estimator for the decomposition $\mathbb{E}[P_L]$ is the following:

$$P_{L}(\varsigma) = [P_{0} - \varsigma_{0}(X_{0} - \mathbb{E}[X_{0}])] + \sum_{i=1}^{L} [P_{i} - P_{i-1} - \varsigma_{i}(X_{i} - X_{i-1} - \mathbb{E}[X_{i} - X_{i-1}])]$$

= $[P_{0} - \varsigma_{0}(X_{0} - \mathbb{E}[X_{0}])] + \sum_{i=1}^{L} [P_{i} - P_{i-1} - \varsigma_{i}(X_{i} - X_{i-1})].$ (58)

It can be easily shown that Equation (58) is an unbiased estimator for $\mathbb{E}[P_L]$:

$$\mathbb{E}[P_{L}(\varsigma)] = \mathbb{E}[P_{0} - \varsigma_{0}(X_{0} - \mathbb{E}[X_{0}])] + \sum_{i=1}^{L} \mathbb{E}[P_{i} - P_{i-1} - \varsigma_{i}(X_{i} - X_{i-1})]$$

$$= \mathbb{E}[P_{0}] - \varsigma_{0} \mathbb{E}[X_{0} - \mathbb{E}[X_{0}]] + \sum_{i=1}^{L} \mathbb{E}[P_{i} - P_{i-1}] - \varsigma_{i} \mathbb{E}[X_{i} - X_{i-1}]$$

$$= \mathbb{E}[P_{0}] + \sum_{i=1}^{L} \mathbb{E}[P_{i} - P_{i-1}] = \mathbb{E}[P_{L}].$$
(59)

To obtain the ratio of variance of the proposed estimator to the Multilevel estimator, we first define the total variance of the multilevel estimator as follows:

$$V^{ML} = \sum_{i=0}^{L} \frac{1}{N_i^{(1)}} V_i^{(1)},$$
(60)

where $V_0^{(1)} = Var(P_0)$ and $V_i^{(1)} = Var(P_i - P_{i-1})$ for i = 1, ..., L. Similarly, we can define the total variance for the multilevel control variate estimator as follows:

$$V^{MLCV} = \sum_{i=0}^{L} \frac{1}{N_i^{(2)}} V_i^{(2)},$$
(61)

where $V_0^{(2)} = Var(P_0 - \varsigma_0(X_0 - \mathbb{E}[X_0]))$ and $V_i^{(2)} = Var(P_i - P_{i-1} - \varsigma_i(X_i - X_{i-1}))$ for i = 1, ..., L. We can reduce the total variance for the multilevel control variate estimator by minimizing $V_i^{(2)}$ individually with optimal ς_i for i = 0, 1, ..., L, that is, the following:

$$\frac{\partial V_0^{(2)}}{\partial \zeta_0} = \frac{\partial}{\partial \zeta_0} \Big[\sigma_{P_0}^2 - 2\zeta_0 \sigma_{P_0} \sigma_{X_0} \rho_{P_0 X_0} + \zeta_0^2 \sigma_{X_0}^2 \Big] = 0 \implies \zeta_0^* = \frac{\sigma_{P_0}}{\sigma_{X_0}} \rho_{P_0 X_0}, \tag{62}$$

where ζ_0^* is the optimal value that minimizes $V_0^{(2)}$. Similarly, if we let $Y_i = P_i - P_{i-1}$ and $R_i = X_i - X_{i-1}$ for i = 1, ..., L, then the optimal ζ_i^* that reduces $V_i^{(2)}$ for i = 1, 2, ..., L is as follows:

$$\varsigma_i^* = \frac{\sigma_{Y_i}}{\sigma_{R_i}} \rho_{Y_i R_i}.$$
(63)

Equations (62) and (63) thus imply the least individual variance as follows:

$$V_0^{(2)} = \sigma_{P_0}^2 (1 - \rho_{P_0 X_0}^2), \qquad V_i^{(2)} = \sigma_{Y_i}^2 (1 - \rho_{Y_i R_i}^2) \quad \forall i = 1, ..., L$$
(64)

or in another form

$$V_0^{(2)} = V_0^{(1)} (1 - \rho_{P_0 X_0}^2), \qquad V_i^{(2)} = V_i^{(1)} (1 - \rho_{Y_i R_i}^2) \quad \forall i = 1, ..., L.$$
(65)

Using Equations (60), (61) and (65), we can compute the ratio of the total variance of the optimally controlled multilevel control variate estimator to the total variance of the multilevel estimator. Suppose we let $N_i = N_i^{(2)} = N_i^{(1)}$. To simplify the calculation, we can obtain the following ratio:

$$\frac{V^{MLCV}}{V^{ML}} = \frac{\sum_{i=0}^{L} \frac{1}{N_i} V_i^{(2)}}{\sum_{i=0}^{L} \frac{1}{N_i} V_i^{(1)}} \\
= \frac{\frac{1}{N_0} V_0^{(1)} (1 - \rho_{P_0 X_0}^2) + \frac{1}{N_1} V_1^{(1)} (1 - \rho_{Y_1 R_1}^2) + \dots + \frac{1}{N_L} V_1^{(1)} (1 - \rho_{Y_L R_L}^2)}{\frac{1}{N_0} V_0^{(1)} + \frac{1}{N_1} V_1^{(1)} + \dots + \frac{1}{N_L} V_1^{(1)}} \\
= 1 - \frac{\frac{1}{N_0} V_0^{(1)} \rho_{P_0 X_0}^2 + \frac{1}{N_1} V_1^{(1)} \rho_{Y_1 R_1}^2 + \dots + \frac{1}{N_L} V_1^{(1)} \rho_{Y_L R_L}^2}{\frac{1}{N_0} V_0^{(1)} + \frac{1}{N_1} V_1^{(1)} + \dots + \frac{1}{N_L} V_1^{(1)}}.$$
(66)

Equation (66) indicates that the variance reduction factor is weighted by $\rho_{P_0X_0}^2, \rho_{Y_1R_1}^2, ..., \rho_{Y_LR_L}^2$. Since $-1 \le \rho \le 1$, the variance reduction on the multilevel estimator is basically guaranteed, unless $\rho_{P_0X_0} = \rho_{Y_1R_1} = ... = \rho_{Y_LR_L} = 0$. Furthermore, if $\rho_{P_0X_0} = \rho_{Y_1R_1} = ... = \rho_{Y_LR_L} = 1$, then the variance is minimized to zero. Unfortunately, we are faced with the same issue as any Monte Carlo estimation methods—we are unable simulate $X_0, X_1, ..., X_L$ directly and exactly. Therefore, we do have to rely on discretization methods and therefore, discretization bias is unavoidable. The estimator that we will be using to conduct numerical experiment is as follows:

$$\widehat{P}_{L}(\varsigma) = \frac{1}{N_{0}^{(2)}} \sum_{k=1}^{N_{0}^{(2)}} \left(P_{0}^{(0,k)} - \varsigma_{0}[X_{0}^{(0,k)} - \mathbb{E}[X]] \right) \\
+ \sum_{i=1}^{L} \frac{1}{N_{i}^{(2)}} \sum_{k=1}^{N_{i}^{(2)}} \left(\left[P_{i}^{(i,k)} - P_{i-1}^{(i,k)} \right] - \varsigma_{i} \left[X_{i}^{(i,k)} - X_{i-1}^{(i,k)} \right] \right),$$
(67)

where the optimal ς_0 and ς_i from Equations (62) and (63) are estimated as follows:

$$\widehat{\varsigma}_{0}^{*} = \frac{\sum_{k=1}^{N_{0}^{(2)}} \left(P_{0}^{(0,k)} - \overline{P}_{0}^{(0,\cdot)} \right) \left(X_{0}^{(0,k)} - \overline{X}_{0}^{(0,\cdot)} \right)}{\sum_{k=1}^{N_{0}^{(2)}} \left(X_{0}^{(0,k)} - \overline{X}_{0}^{(0,\cdot)} \right)^{2}}$$
(68)

and

$$\widehat{\varsigma}_{i}^{*} = \frac{\sum_{k=1}^{N_{i}^{(2)}} \left(\left[P_{i}^{(i,k)} - P_{i-1}^{(i,k)} \right] - \left[\overline{P}_{i}^{(i,\cdot)} - \overline{P}_{i-1}^{(i,\cdot)} \right] \right) \left(\left[X_{i}^{(i,k)} - X_{i-1}^{(i,k)} \right] - \left[\overline{X}_{i}^{(i,\cdot)} - \overline{X}_{i-1}^{(i,\cdot)} \right] \right)}{\sum_{k=1}^{N_{i}^{(2)}} \left(\left[X_{i}^{(i,k)} - X_{i-1}^{(i,k)} \right] - \left[\overline{X}_{i}^{(i,\cdot)} - \overline{X}_{i-1}^{(i,\cdot)} \right] \right)^{2}$$

$$(69)$$

such that $(\overline{P}_i^{(i,\cdot)}, \overline{P}_{i-1}^{(i,\cdot)})$ and $(\overline{X}_i^{(i,\cdot)}, \overline{X}_{i-1}^{(i,\cdot)})$ are the sample mean for $(P_i^{(i,\cdot)}, P_{i-1}^{(i,\cdot)})$ and $(X_i^{(i,\cdot)}, X_{i-1}^{(i,\cdot)})$. In practice, we will not have the optimal $\widehat{\varsigma}_i^*$ for i = 0, ..., L immediately; pre-running the simulation with a small sample is needed. We need to clarify that $X_i^{(\cdot,\cdot)}$ has the same coarseness level as $P_i^{(\cdot,\cdot)}$. In other words, $X_i^{(\cdot,\cdot)}$ has the same number of discretization steps and the same step size as $P_i^{(\cdot,\cdot)}$. Additionally, in order to generate the largest correlation for the pairs $(P_0, X_0), (Y_1, R_1), ..., (Y_L, R_L)$, we should employ the same sample path to each of the pairs $(P_0^{(0,k)}, X_0^{(0,k)}), (P_0^{(1,k)}, X_1^{(1,k)}), (P_L^{(1,k)}, X_L^{(1,k)})$. Furthermore, to avoid large discretization error on the simulation of X_i for i = 0, 1, ..., L, we will simulate X_i for i = 0, 1, ..., L, using the second-order Euler method (introduced in Section 2.2), which has a weak order of convergence $O(\delta^2)$. We will see that in the numerical experiment Section 4, the discretization biases are ignorable once a certain level of the multilevel control variate method is employed.

Remark 3. The total discretization steps n_i and step size δ_i for the multilevel control variate method are defined the same as the multilevel method. In addition, the optimal $N_i^{(2)}$ for Equation (61) is computed similarly as in Section 3.2. Note that a fair comparison between the multilevel and multilevel control variate should be done using cost-adjusted performance. In other words, the benefit of using the extra control variates to reduce variance must outweigh the cost of generating the extra paths for the control variate.

4. Numerical Experiments

This section discusses the numerical experiments for our evaluated Monte Carlo methods in four different ways. First, we demonstrate that by using the multilevel control variate Monte Carlo, the implied volatility under the rough Heston model has the ability to change according to different ρ and α with different maturity time *T*. Then, we will evaluate the cost-adjusted distribution for the OTM, ATM, and ITM volatility under the rough Heston model on the following estimators: (1) base estimator (BE), (2) control variate estimator (MLCV). Subsequently, we will compare the bias of the four above-mentioned estimators. Finally, we compare our methods to the implied volatility under the rough Heston model computed, using the Fourier inversion method.

Accordingly, we define the following parameters for four estimators that will be used in the following numerical experiment unless stated otherwise.

$$S = 100, \quad K = 100, \quad \alpha = 0.6, \quad \sigma_0^2 = 0.16, \quad T = 1$$

$$\theta = 0.16, \quad \lambda = 0.5, \quad \nu = 0.5, \quad \rho = -0.6 \quad n = 256$$

$$M = 256, \quad D = 2, \quad L = 4.$$
(70)

The parameter n is used by BE and CV, whereas the parameters M, D and L are used by ML and MLCV. Note that we did not specify the number of simulations N for all the estimators, and instead, we will be using the same run-time or cost for all the methods, as

this will keep the benchmark of different methods fair. The general steps of our Monte Carlo simulation are as follows:

- 1. Pre-run the simulation of the methods with relatively small N^1 and store the intended result.
- 2. Compute time-taken to run a single simulation.
- 3. Compute N^2 as the number of simulations needed to achieve the specific run-time or cost *C*.
- 4. With the previous N^1 simulations stored, conduct the rest of $N^2 N^1$ simulations.

The numerical experiments are coded in MATLAB R2019b software with i7-8750H CPU @ 2.20 GHz (Max Turbo Frequency: 4.10 GHz) and 16 GB memory.

4.1. The ρ and α Change

Here, we show that the Monte Carlo method on implied volatility under the rough Heston model is capable of demonstrating various smile shapes, according to α and ρ with different maturity time *T*. Figure 1 is produced using MLCV with the parameters stated in Equation (70), except for the α and ρ . In addition, in order to make the graph looks smoother, we used 250 simulations of cost *C* = 1 and then averaged the computed implied volatility values.

We can see that the behaviors from Figure 1 are quite reasonable for short- and longterm implied volatility. The plots somewhat fit the empirical observation from the work of [6], i.e. their Figures 4 and 8 indicate that as the maturity *T* of the implied volatility plots gets smaller, the center of the smile approaches to log(K/S) = 0 becomes much more obvious. In addition, we can observe the following effects from the implied volatility plots Figure 1: (1) larger α causes the implied volatilities for different log-strike to have larger discrepancies at in-the-money options, whereas out-of-the-money options are priced below the Black–Scholes option prices on different maturity lengths *T*; (2) higher correlation ρ flattens the implied volatility such that it behaves more like a smile shape.

Remark 4. Note that during the computational work in the numerical experiment, the implied volatility with a higher correlation between the stock return and volatility movement ρ requires more computational effort to compute to an accurate number as compared to the lower ρ . Furthermore, while we have only used MLCV to estimate the implied volatility plots on Figure 1, similar plots can be constructed using the BE, CV, and ML methods, but it would take a larger number of simulations to create a smooth line across different log strikes in the figures. The reason that it requires a larger number of simulations is discussed in the subsequent subsection.



Figure 1. Implied volatility against log-strike log(K/S) plots on using $\alpha = \{0.6, 0.8\}$ and $\rho = \{-0.6, -0.2\}$ with different maturity time *T*.

4.2. Cost-Adjusted Variation and Bias

This section focuses on the following four matters: (1) evaluate the distribution and standard deviation of the out-of-the-money (OTM), at-the-money (ATM), and in-the-money (ITM) implied volatilities under the rough Heston model using MLCV; (2) compare the standard deviation of the MLCV on the changes of the parameters *T*, ρ , and α ; (3) evaluate the distribution and standard deviations between the BE, CV, ML, and MLCV; and (4) compute discretization biases for all the previously mentioned methods on a different number of time steps *n* and level *L*. We set the cost of numerical experiments as *C* = 1 (second) to avoid improper comparison between estimators.

Figure 2 shows the histogram plots of $1000(\sigma_{imp}^i - \bar{\sigma}_{imp})$ for i = 1, 2, ..., 3000 simulations (3000 runs of 1 second performance) of MLCV with a normal distribution fit; $\bar{\sigma}_{imp}$ is the sample mean of the computed implied volatility σ_{imp} . Furthermore, we have fitted normal distribution plot $N(0, \sigma_{norm})$, where σ_{norm} is the standard deviation of $1000(\sigma_{imp}^i - \bar{\sigma}_{imp})$. Based on Figure 2, we can observe that ITM options have the largest standard deviation, and therefore, it is harder to pin down an exact option price as compared to the ATM and OTM option. On a bright note, the concept of normal distribution ensures that there is a 99.7% (6 standard deviations gap) chance that the ITM implied volatility would have an absolute error of at most 0.459% (implied volatility error) or roughly 1% percentage error for the implied volatility.

Similarly, we can compute the standard deviations for ATM implied volatility on different values of ρ , α , and T. Table 1 displays the standard deviation of the implied volatility estimator computed by using the MLCV. From Table 1, we can infer that by using the MLCV, option prices under the rough Heston model that have longer maturity time T

15 of 32

incur more error than the shorter maturity time *T*. In addition, based on the observation of the changes in ρ and α , it is evident that a higher α value will lead to a smaller error for the implied volatility, whereas a higher ρ will do the opposite. It is important to realize that among the evaluated ATM implied volatility standard deviations, even the worst case scenario ($\rho = -0.2$, $\alpha = 0.6$, T = 1) would have 99.7% chance of having a percentage error of roughly 1.2% only.

In addition, we have computed and plotted the histogram shown in Figure 3 for the ATM implied volatility of the four estimators. Unsurprisingly, the MLCV performed the best, with the smallest standard deviation among all the estimators. To put it in a perspective, by using the multilevel control variate estimator, we managed to achieve a roughly 17 times variance reduction as compared to BE. CV and ML are also able to demonstrate a substantial variance reduction of roughly 11 times and 12 times.



Figure 2. Histograms and normal distribution fits of OTM, ATM and ITM ($k = \log(K/S) = \{-0.1053, 0, 0.0953\}$) implied volatility. The x-axis is $1000(\sigma_{imp}^k - \bar{\sigma}_{imp})$, whereas the y-axis is simply the probability that the implied volatility falls into the interval of the histogram. The σ_{norm} is the standard deviation of the normal distribution fit.

	σ_{norm}			
	T = 1/52	T = 3/12	T = 6/12	T = 1
$\rho = -0.6, \ \alpha = 0.6$	0.00084	0.00084	0.00100	0.00123
$ ho = -0.6, \ lpha = 0.8$	0.00025	0.00036	0.00045	0.00062
$ ho = -0.2, \ lpha = 0.6$	0.00089	0.00107	0.00123	0.00156
$\rho = -0.2, \ \alpha = 0.8$	0.00033	0.00044	0.00062	0.00085

Table 1. Computation of implied volatility's standard deviation.



Figure 3. Histograms and normal distribution fits of ATM implied volatility errors for base, control variate, multilevel and multilevel control variate estimator. The x-axis is transformed to $1000(\sigma_{imp}^k - \bar{\sigma}_{imp})$, whereas the y-axis is simply the probability that the implied volatility falls into the interval of the histogram. The σ_{norm} is the standard deviation of the normal distribution fit.

Finally, Figure 4 shows the discretization biases on n = 64, 128, 256, 512, 1024 for BE and CV, as well as the biases on L = 1, ..., 4 for ML and MLCV. Note that for ML and MLCV, we are using $(L, M) = \{\{1, 128\}, \{2, 256\}, \{3, 512\}, \{4, 1024\}\}$. Figure 4 shows the number of time steps n or M against the ATM implied volatility. Evidently, the observed implied volatilities of the estimators do differ by a slight bit. The most notable difference in implied volatility, but ultimately, when n becomes larger or level L becomes larger, Figure 4 shows that the estimators do appear to converge and have little variance between one another. A 2 level MLCV with M = 256 has 0.15% percentage error as compared to BE with number of time steps n = 1024, which is acceptable in the financial industry.



Figure 4. Number of time steps or levels against the implied volatility plot to assess the discretization error.

4.3. Calibration to SPX Options

In this section, we will calibrate the implied volatility under the rough Heston model computed through MLCV to the options of an index consisting of the top 500 companies in the U.S. market, which is the S&P 500 (SPX) option. The SPX options prices and stock are taken from the ending on 22 October 2021 with maturity dates ranging from T = 0.0992 up to T = 0.4812. The interest rate is assumed to be zero in the inversion of option price to implied volatility and computation of implied volatility through the rough Heston model. The main criterion of optimizing the parameters is the minimization of the root mean squared error between the bid-ask implied volatility. The calibrated parameters are as follows:

$$\alpha = 0.7600, \quad \theta = 0.0450, \quad V_0 = 0.0095,$$
(71)

$$\lambda = 2.0558, \quad \nu = 3.0480, \quad \rho = -0.9507.$$
 (72)

Figure 5 shows the calibration of implied volatility under the rough Heston model to the SPX options against log-moneyness log(K/S) (on the *x*-axis), using the MLCV Monte Carlo method. It shows that the calibrated implied volatility performs reasonably well for the different option maturities considered. It can be seen from Figure 5 that the implied volatility on T = 0.2579 and T = 0.4812 have some deviations at its end (in-the-money and out-of-the-money). One thing to take note of is that we have actually attempted to calibrate the rough Heston's implied volatility, using MLCV up to maturity of T = 1.66, but were unable to do so because of two main reasons: (1) with cost C = 1 s, the implied volatility at larger maturity actually has large deviation, so the optimization algorithm simply would not work well with such a large deviation at each iteration; (2) there are six parameters to consider, and therefore, the optimization cost is extremely high. We are actually unsure whether the parameters in Equations (71) and (72) are even the global minimum of the optimization problem or not. We have tried three different optimization algorithms, including the gradient-based method (interior-point method), evolutionary algorithm (genetic algorithm), and pattern search algorithm. The gradient-based method obviously does not work well with stochastic problems, whereas the evolutionary algorithm simply requires extremely large computational cost to give a moderate optimal parameter. The parameters in Equations (71) and (72) are obtained through a pattern search algorithm with initial parameters given by the genetic algorithm.



Figure 5. Implied volatility surface of SPX options (22 October 2021) and its calibration. Red and blue solid lines are bid and ask implied volatilities respectively, whereas green-dashed line is implied volatility under the rough Heston model computed through the MLCV Monte Carlo method.

5. Conclusions and Future Research

This paper mainly contributes to the Monte Carlo computation for option prices under the rough Heston model. In particular, we have proved finite bounds on *t* for the stochastic Volterra equation with a non-Lipschitz diffusion coefficient as well as the weak error and weak convergence of the discretization method. The proofs are located in the Appendices A and B. In addition, we have proposed a Control Variate estimator for variance reduction purpose and a mixed Monte Carlo method, which consists of the control variate method and multilevel method.

In Section 4, we conducted extensive numerical experiments that consider many scenarios, which we believe many readers or practitioners might find interesting. The most notable part is in Section 4.2, where all the computations and simulations are performed, using the cost-adjusted method rather than the number of simulations *N*. Time to time comparisons make it easier to identify which estimator is performing the best. In this case, the most favorable estimator goes to the proposed MLCV method. A substantial 17 times variance reduction can be achieved by using MLCV as compared to the BE. The numerical experiment for the discretization bias of the estimators was also conducted. The results show that there is little difference between n = 256 and n = 1024 in terms of discretization bias. A calibration of the SPX option was also performed on the MLCV Monte Carlo

method at the end of Section 4. The overall performance of the calibration is overall satisfied, but there were some struggles in calibrating the optimal parameters.

We are aware that there was a recent study on an efficient simulation method of the stochastic Volterra equation [33] that used a hybrid exponential scheme. In particular, the author proved the convergence of the method with Lipschitz coefficient functions to the stochastic Volterra equation. The method can be potentially used to replace the discrete simulation of the stochastic Volterra equation for rough Heston volatility in practice once the convergence of the non-Lipschitz coefficient functions is proven. For future research, it would be interesting to see the hybrid exponential scheme [20,34] being used on a rough Heston model simulation and then consider the cost-adjusted performance of different estimators, including the proposed multilevel control variate estimator. In general, we also do hope to see improvements on the finite bound of $\mathbb{E}[|V_t - V_t^n|^p]$ or perhaps even strong convergence of the stochastic Volterra equation with a non-Lipschitz diffusion coefficient, i.e., $\lim_{n\to\infty} \mathbb{E}[|V_t - V_t^n|^p] = 0$. The main hurdle of proving the strong convergence in Lemma A1 is because we cannot use the Grönwall inequality to Equation (A37) the same way as we did to Equation (A49). Instead, we have to use Perov's inequality to prove the convergence that is bounded on time *t*. In addition, Remark 2 indicates that the weak convergence rate of stochastic Volterra equation $|\mathbb{E}[V_t] - \mathbb{E}[V_t^n]|$ is of the order $O(n^{-H})$, but in the study [35], the author proved that the discretization scheme of the Heston model (with H = 1/2 in the stochastic Volterra process) under a certain hypothesis has a weak convergence rate of $O(n^{-1})$. This indicates that the weak convergence rate of stochastic Volterra equation $|\mathbb{E}[V_t] - \mathbb{E}[V_t^n]|$ can potentially be improved to the order of $O(n^{-H-1/2})$, which we leave for future work.

Author Contributions: Conceptualization, S.W.J. and A.K.; methodology, S.W.J. and A.K.; software; S.W.J.; validation, S.W.J. and A.K.; formal analysis, S.W.J. and A.K.; investigation, S.W.J. and A.K.; resources, S.W.J. and A.K.; data curation, S.W.J.; writing—original draft preparation, S.W.J. and A.K.; writing—review and editing, A.K.; visualization, S.W.J.; supervision, A.K.; project administration, A.K.; funding acquisition, A.K. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to acknowledge that this research is partially supported by Special Graduate Research Allowance (SGRA) with Vote number of 5540153.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: SPX data is available on: https://github.com/WoonJeng/On-Multileveland-Control-Variate-Monte-Carlo-Methods-for-Option-Pricing-under-the-RoughHestonModel.git, accessed date 16 September 2021.

Acknowledgments: The article was partially funded by the Geran Putra Berimpak (GPB) with project number GPB/2017/9543000.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

- OTM Out-of-the-money
- ATM At-the-money
- ITM In-the-money
- BE Base estimator
- CV Control variate estimator
- ML Multilevel estimator
- MLCV Multilevel control variate estimator

Appendix A. Proof of Lemma 1

We prove that the kernel $K(t, u) := K_i(t, u) = (t - u)^{H-1/2}$ for i = 1, 2 satisfies the conditions (AA1)–(AA7). Let C > 0 be a fixed constant. We omit the proofs for conditions (AA1)–(AA3), as they are easy to prove. As for condition (AA4), the proof is the same as that for condition (AA5) when $\beta = 1$. Hence, we prove that it satisfies condition (AA5) instead.

$$\int_0^t [K(t+\delta,u) - K(t,u)]^{2\beta} du = \int_0^t [(t+\delta-u)^{H-1/2} - (t-u)^{H-1/2}]^{2\beta} du.$$
(A1)

Let v = u - t, then it follows that

$$\int_0^t [K(t+\delta,u) - K(t,u)]^{2\beta} du = \int_{-t}^0 [(\delta-v)^{H-1/2} - (-v)^{H-1/2}]^{2\beta} du.$$
(A2)

Furthermore, let $w = \frac{v}{\delta}$. Then we have the following:

$$\int_{0}^{t} [K(t+\delta,u) - K(t,u)]^{2\beta} du = \int_{-t/\delta}^{0} [(\delta - w\delta)^{H-1/2} - (-w\delta)^{H-1/2}]^{\beta} \delta dw$$
$$= \delta^{(2H-1)\beta+1} \int_{-t/\delta}^{0} \left[(1-w)^{H-1/2} - (-w)^{H-1/2} \right]^{2\beta} dw.$$
(A3)

Finally, we let w = -a and the fact that for $x, y \in \mathbb{R}^+$, x > y, and $q \in (0, 1)$. Then, we notice that $(x - y)^q \ge x^q - y^q$ to obtain the following:

$$\begin{split} \int_{0}^{t} [K(t+\delta,u) - K(t,u)]^{2\beta} du &= \delta^{(2H-1)\beta+1} \int_{0}^{t/\delta} \left[(1+a)^{H-1/2} - a^{H-1/2} \right]^{2\beta} da \\ &= \delta^{(2H-1)\beta+1} \int_{0}^{t/\delta} \left[\frac{a^{1/2-H} - (1+a)^{1/2-H}}{[(1+a)a]^{1/2-H}} \right]^{2\beta} da \\ &\leq \delta^{(2H-1)\beta+1} \int_{0}^{t/\delta} \left[\frac{(1+a)^{1/2-H} - a^{1/2-H}}{[(1+a)a]^{1/2-H}} \right]^{2\beta} da \\ &\leq \delta^{(2H-1)\beta+1} \int_{0}^{t/\delta} \left[\frac{(1+a)^{1/2-H} - a^{1/2-H}}{[(1+a)a]^{1/2-H}} \right]^{2\beta} da \\ &\leq \delta^{(2H-1)\beta+1} \int_{0}^{t/\delta} \left[\frac{1}{[(1+a)a]^{(1-2H)\beta}} \right] da \\ &\leq C \delta^{(2H-1)\beta+1}. \end{split}$$
(A4)

The last inequality is because the integral is finite only if (1 - 2H)B < 1, which is fulfilled for the conditions $B \in (1, 1/(1 - 2H))$ and $H \in (0, 1/2)$. The second sum of condition (AA4) can be easily shown by taking note of the following:

$$\int_{0}^{t} [K_{i}(t+\delta,u) - K(t,u)]^{2} ds \ge \int_{0}^{t} [K_{i}(t+\delta,\eta(u)) - K_{i}(t,\eta(u))]^{2} du.$$
(A5)

Similarly, we skip condition (AA6), as it can be shown easily when B = 1 in condition (AA7). For $K(t, u) = (t - u)^{H-1/2}$, condition (AA7) can be shown as follows:

$$\int_{0}^{t} [K(t,u) - K(t,\eta(u))]^{2\beta} du = \int_{0}^{t} [(t-u)^{H-1/2} - (t-\eta(u))^{H-1/2}]^{2\beta} du$$
$$= \int_{0}^{t} [(t-u)^{H-1/2} - (t-u+u-\eta(u))^{H-1/2}]^{2\beta} du.$$
(A6)

Note that $0 \le u - \eta(u) < \delta$.

$$\int_{0}^{t} \left[K(t,u) - K(t,\eta(u)) \right]^{2\beta} du \le \int_{0}^{t} \left[(t-u)^{H-1/2} - (t-u+\delta)^{H-1/2} \right]^{2\beta} du.$$
(A7)

Then, it follows the same from the proof of condition 1.

Appendix B. Proof of Theorem 1

Before proving Theorem 1, we do need to prove two lemmas first. The following proofs are somewhat similar to the work of [21] in proving the convergence of the stochastic Volterra equation on the Lipschitz conditions, but with different rate of convergence in time. Note that Fubini's theorem is repeatedly used in the the following proofs to interchange the sign of expectation and integral without being mentioned. We assume that Assumption 1 and conditions (AA1)–(AA7) hold in all the proofs.

Lemma A1. Let $p \ge \frac{2\beta}{\kappa(\beta-1)}$ for $\kappa \in (0,1)$ and $\beta \in (1,1/(1-2H))$; then, there exists a finite constant C > 0 that depends on T, p, V_0, β such that for $t \in [0,T]$ and $n \ge 1$, we have the following:

 $\mathbb{E}[|V_t|^p] + \mathbb{E}[|V_t^n|^p] \le C,\tag{A8}$

where V_t and V_t^n are defined in Equations (7) and (20).

Proof. We start from the stochastic Volterra Equation (7). Then, by the Cauchy–Schwarz inequality and the Buckholder–Davis–Gundy (BDG) inequality, we have the following:

$$\mathbb{E}[|V_{t}|^{p}] \leq C \mathbb{E}[|V_{0}|^{p}] + C \mathbb{E}\left[\left|\int_{0}^{t} K_{1}(t, u)g(u, V_{u})du\right|^{p}\right] \\ + C \mathbb{E}\left[\left|\int_{0}^{t} K_{2}(t, u)m(u, V_{u})dW_{u}\right|^{p}\right] \\ \leq C + C \mathbb{E}\left[\left|\int_{0}^{t} K_{1}(t, u)^{2}du\int_{0}^{t} g(u, V_{u})^{2}du\right|^{p/2}\right] \\ + C \mathbb{E}\left[\left|\int_{0}^{t} K_{2}(t, u)^{2}m(u, V_{u})^{2}du\right|^{p/2}\right].$$
(A9)

By using Condition (AA1), Assumption (BB1), Jensen's inequality and Hölder's inequality, we can obtain the following:

$$\mathbb{E}[|V_t|^p] \leq C + C \int_0^t \mathbb{E}\left[|1 + V_u^p|\right] du + C \mathbb{E}\left[\left|\left(\int_0^t K_2(t, u)^{2\beta} du\right)^{1/\beta} \left(\int_0^t m(u, V_u)^{2\beta^*} du\right)^{1/\beta^*}\right|^{p/2}\right]$$
(A10)
$$\leq C + C \int_0^t \mathbb{E}\left[|1 + V_u^p|\right] du + C \mathbb{E}\left[\left|\left(\int_0^t m(u, V_u)^{2\beta^*} du\right)^{(p(\beta-1))/2\beta}\right|\right].$$
(A11)

Notice that from the Hölder inequality, $\beta > 1$, $0 < \kappa < 1$, and hence p > 2. Then it follows from condition (BB1) and Jensen's inequality that the following holds:

$$\mathbb{E}[|V_t|^p] \leq C + C \int_0^t \mathbb{E}[1 + |V_u|^p] du + C \mathbb{E}\left[\int_0^t m(u, V_u)^p du\right] \leq C + C \int_0^t \mathbb{E}[1 + |V_u|^p] du + C \mathbb{E}\left[\int_0^t (1 + |V_u|)^p du\right] \leq C + C \int_0^t \mathbb{E}[|V_u|^p] du.$$
(A12)

Lastly, we use Grönwall's inequality to obtain the following:

$$\mathbb{E}[|V_t|^p] \le C \tag{A13}$$

For the proof of $\mathbb{E}[|V_t^n|^p]$, the exact same can be done, except for the last step, where we obtain the following:

$$\mathbb{E}[|V_t^n|^p] \le C + C \int_0^t \mathbb{E}[|V_{\eta(u)}^n|^p] du.$$
(A14)

We then define the following function as follows:

$$f_n(t) := \sup_{s \in [0,t]} \mathbb{E}[|V_s^n|^p].$$
(A15)

We continue with

$$\mathbb{E}[|V_t^n|^p] \le f_n(t)$$

$$\le C + C \int_0^t \sup_{s \in [0,u]} \mathbb{E}[|V_{\eta(s)}^n|^p] du$$

$$\le C + C \int_0^t \sup_{s \in [0,u]} \mathbb{E}[|V_s^n|^p] du$$

$$\le C.$$
(A16)

The last inequality is due to Grönwall's inequality. \Box

Lemma A2. Let $p \ge \frac{2\beta}{\kappa(\beta-1)}$ for $\kappa \in (0,1)$ and $\beta \in (1,1/(1-2H))$, then for $t-s \le \delta$ there exists a finite constant C > 0 that depends on T, p, V_0 , β such that for $t \in [0,T]$ and $n \ge 1$, we have the following:

$$\mathbb{E}[|V_t - V_s|^p] + \mathbb{E}[|V_t^n - V_s^n|^p] \le C\delta^{(p/2\beta)[(2H-1)\beta+1]}.$$
(A17)

where V_t and V_t^n are defined in Equations (7) and (20).

Proof. We first rearrange $V_t - V_s$ as follows:

$$V_{t} - V_{s} = \int_{s}^{t} K_{1}(t, u)m(u, V_{u})du + \int_{s}^{t} K_{2}(t, u)g(u, V_{u})dW_{u} + \int_{0}^{s} K_{1}(t, u)m(u, V_{u}) - K_{1}(s, u)m(u, V_{u})du + \int_{0}^{s} K_{2}(t, u)g(u, V_{u}) - K_{2}(s, u)g(u, V_{u})dW_{u} = R_{1} + R_{2} + R_{3} + R_{4}.$$
(A18)

We start with $\mathbb{E}[|R_1|^p]$ where we use the Cauchy–Schwarz inequality, condition (AA2) and Jensen's inequality.

$$\mathbb{E}[|R_1|^p] \leq \mathbb{E}\left[\left|\int_s^t K_1(t,u)^2 du \int_s^t m(u,V_u)^2 du\right|^{p/2}\right]$$
$$\leq \mathbb{E}\left[\left|C\delta^{2H} \int_s^t m(u,V_u)^2 du\right|^{p/2}\right]$$
$$\leq C\delta^{pH}\left[\int_s^t C(1+\mathbb{E}[|V_u|^p]) du\right].$$
(A19)

Using Lemma A1, we can obtain the following:

$$\mathbb{E}[|R_1|^p] \le C\delta^{pH} \left[\int_s^t C(1+C)du \right]$$

$$\le C\delta^{pH}.$$
(A20)

For $\mathbb{E}[|R_2|^p]$, we will be using BDG inequality, Hölder inequality, condition (AA3), Jensen's inequality, condition (BB1) and Lemma A1, we can obtain the following:

$$\mathbb{E}[|R_{2}|^{p}] \leq \mathbb{E}\left[\left|\int_{s}^{t} K_{2}(t,u)g(u,V_{u})dW_{u}\right|^{p}\right]$$

$$\leq C \mathbb{E}\left[\left|\int_{s}^{t} K_{2}(t,u)^{2}g(u,V_{u})^{2}du\right|^{p/2}\right]$$

$$\leq C \mathbb{E}\left[\left|\left(\int_{s}^{t} K_{2}(t,u)^{2\beta}du\right)^{1/\beta}\left(\int_{s}^{t} g(u,V_{u})^{2\beta^{*}}du\right)^{1/\beta^{*}}\right|^{p/2}\right]$$

$$\leq C\delta^{(p/2\beta)[(2H-1)\beta+1]} \mathbb{E}\left[\int_{s}^{t} g(u,V_{u})^{p}du\right]$$

$$\leq C\delta^{(p/2\beta)[(2H-1)\beta+1]}\int_{s}^{t} C(1+\mathbb{E}[|V_{0}|^{p}])du$$

$$\leq C\delta^{(p/2\beta)[(2H-1)\beta+1]}.$$
(A21)

Do note that for $\beta \in (1, 1/(1-2H))$ and $H \in (0, 1/2)$, $(2H-1)\beta + 1 > 0$. As for $\mathbb{E}[|R_3|^p]$, the Cauchy–Schwarz inequality, the Hölder inequality, condition (AA4), Jensen's inequality, condition (BB1) and Lemma A1 are used.

$$\mathbb{E}[|R_{3}|^{p}] \leq \mathbb{E}\left[\left|\int_{0}^{s} K_{1}(t,u)m(u,V_{u}) - K_{1}(s,u)m(u,V_{u})du\right|^{p}\right]$$

$$\leq \mathbb{E}\left[\left|\int_{0}^{s} [K_{1}(t,u) - K_{1}(s,u)]m(u,V_{u})du\right|^{p}\right]$$

$$\leq \mathbb{E}\left[\left|\int_{0}^{s} [K_{1}(t,u) - K_{1}(s,u)]^{2}du\int_{0}^{s} m(u,V_{u})^{2}du\right|^{p/2}\right]$$

$$\leq C\delta^{pH} \mathbb{E}\left[\int_{0}^{s} m(u,V_{u})^{p}du\right]$$

$$\leq C\delta^{pH}.$$
(A22)

Lastly, we prove the bounded $\mathbb{E}[|R_4|^p]$ by using the BDG inequality, the Hölder inequality and condition (AA5).

$$\mathbb{E}[|R_{4}|^{p}] \leq \mathbb{E}\left[\left|\int_{0}^{s} K_{2}(t,u)g(u,V_{u}) - K_{2}(s,u)g(u,V_{u})dW_{u}\right|^{p}\right] \\
\leq \mathbb{E}\left[\left|\int_{0}^{s} [K_{2}(t,u) - K_{2}(s,u)]^{2}g(u,V_{u})^{2}du\right|^{p/2}\right] \\
\leq \mathbb{E}\left[\left|\left(\int_{0}^{s} [K_{2}(t,u) - K_{2}(s,u)]^{2\beta}du\right)^{1/\beta}\left(\int_{0}^{s} g(u,V_{u})^{2\beta^{*}}du\right)^{1/\beta^{*}}\right|^{p/2}\right] \\
\leq C\delta^{(p/2\beta)[(2H-1)\beta+1]} \mathbb{E}\left[\int_{0}^{s} g(u,V_{u})^{p}du\right] \\
\leq C\delta^{(p/2\beta)[(2H-1)\beta+1]}.$$
(A23)

Henceforth, we can conclude the following:

$$\mathbb{E}[|V_t - V_s|^p] \le C\delta^{(p/2\beta)[(2H-1)\beta+1]}.$$
(A24)

This is due to $(1/2\beta)[(2H-1)\beta+1] \le H$ for $\beta \in (1, 1/(1-2H))$. The above steps can be repeated to prove the case of $E[|V_t^n - V_s^n|^p|]$ as well. \Box

We move to the proof of $\mathbb{E}[|V_t - V_t^n|^p]$,

Theorem A1. Let $p \ge \frac{2\beta}{\kappa(\beta-1)}$ for $\kappa \in (0,1)$ and $\beta \in \left(1, \frac{1}{1-2H}\right)$; then, there exists a finite constant C > 0 that depends on T, p, V_0, β and $\epsilon = (p\kappa/2\beta)[(2H-1)\beta+1]$ such that for $t \in [0,T]$ and $n \ge 1$, we have the following:

$$\mathbb{E}[|V_t - V_t^n|^p] \le \left[(C\delta^{\epsilon})^{1-\kappa} \exp(Ct) + C(\exp(Ct) - 1) \right]^{1/(1-\kappa)}.$$
(A25)

Proof. We split $V_t - V_t^n$ as follows:

$$\begin{aligned} V_t - V_t^n &= \int_0^t (K_1(t, u) - K_1(t, \eta(u)))m(u, V_u)du \\ &+ \int_0^t K_1(t, \eta(u))(m(u, V_u) - m(\eta(u), V_{\eta(u)}))du \\ &+ \int_0^t K_1(t, \eta(u))(m(\eta(u), V_{\eta(u)}) - m(\eta(u), V_{\eta(u)}^n))du \\ &+ \int_0^t (K_2(t, u) - K_2(t, \eta(u)))g(u, V_u)dW_u \\ &+ \int_0^t K_2(t, \eta(u))(g(u, V_u) - g(\eta(u), V_{\eta(u)}))dW_u \\ &+ \int_0^t K_2(t, \eta(u))(g(\eta(u), V_{\eta}(u)) - g(\eta(u), V_{\eta(u)}^n))dW_u \\ &= Q_1 + Q_2 + Q_3 + Q_4 + Q_5 + Q_6. \end{aligned}$$
(A26)

We find the individual bound for each term starting for $\mathbb{E}[|Q_1|^p]$, as follows:

$$\mathbb{E}[|Q_{1}|^{p}] \leq \mathbb{E}\left[\left|\int_{0}^{t} (K_{1}(t,s) - K_{1}(t,\eta(u)))m(u,V_{u})\right|^{p}\right]$$

$$\leq \mathbb{E}\left[\left|\int_{0}^{t} (K_{1}(t,u) - K_{1}(t,\eta(u)))^{2}ds \int_{0}^{t} m(u,V_{u})^{2}du\right|^{p/2}\right]$$

$$\leq C \mathbb{E}\left[\left|\delta^{2H} \int_{0}^{t} m(u,V_{u})^{2}du\right|^{p/2}\right]$$

$$\leq C\delta^{pH} \mathbb{E}\left[\left|\int_{0}^{t} m(u,V_{u})^{2}du\right|^{p/2}\right]$$

$$\leq C\delta^{pH}, \qquad (A27)$$

where we used the Cauchy–Schwarz inequality, condition (AA6), Jensen's inequality and Lemma A1. As for $\mathbb{E}[|Q_2|^p]$, we use the Hölder inequality, condition (AA1), Jensen's inequality and condition (BB1) as follows:

$$\begin{split} \mathbb{E}[|Q_{2}|^{p}] &\leq \mathbb{E}\left[\left|\int_{0}^{t} K_{1}(t,\eta(u))(m(u,V_{u}) - m(\eta(u),V_{u}) + m(\eta(u),V_{u}) - m(\eta(u),V_{\eta(u)})du\right|^{p}\right] \\ &\leq \mathbb{E}\left[\left|\int_{0}^{t} K_{1}(t,\eta(u))^{2} du \int_{0}^{t} (m(u,V_{u}) - m(\eta(u),V_{u}))^{2} du\right|^{p/2}\right] \\ &+ \mathbb{E}\left[\left|\int_{0}^{t} K_{1}(t,\eta(u))^{2} du \int_{0}^{t} (m(\eta(u),V_{u}) - m(\eta(u),V_{\eta(u)}))^{2} du\right|^{p/2}\right] \\ &\leq C \mathbb{E}\left[\left|\int_{0}^{t} (m(\eta(u),V_{u}) - m(\eta(u),V_{u}))^{2} du\right|^{p/2}\right] \\ &+ C \mathbb{E}\left[\left|\int_{0}^{t} (m(\eta(u),V_{u}) - m(\eta(u),V_{\eta(u)}))^{2} du\right|^{p/2}\right] \\ &\leq C \delta^{pH} \int_{0}^{t} (1 + \mathbb{E}[|V_{u}|^{p}]) du + C \mathbb{E}\left[\left|\int_{0}^{t} (m(\eta(u),V_{u}) - m(\eta(u),V_{\eta(u)}))^{2} du\right|^{p/2}\right] \\ &\leq C \delta^{pH} + C \int_{0}^{t} \mathbb{E}[|V_{u} - V_{\eta(u)}|^{p}] du. \end{split}$$

(A28)

By Lemma A2 and noticing that $H > (1/2\beta)[(2H-1)\beta + 1]$, we obtain the following:

$$\mathbb{E}[|Q_2|^p] \le C\delta^{pH} + C\delta^{(p/2\beta)[(2H-1)\beta+1]} < C\delta^{(p/2\beta)[(2H-1)\beta+1]}.$$
(A29)

For $\mathbb{E}[|Q_3|^p]$, we use the Cauchy–Schwarz inequality, condition (AA1) and Jensen's inequality such that the following holds:

$$\mathbb{E}[|Q_{3}|^{p}] \leq \mathbb{E}\left[\left|\int_{0}^{t} K_{1}(t,\eta(u)(m(\eta(u),V_{\eta(u)}) - m(\eta(u),V_{\eta(u)}^{n}))\right|^{p}\right]$$

$$\leq \mathbb{E}\left[\left|\int_{0}^{t} K_{1}(t,\eta(u)^{2}du\int_{0}^{t}(m(\eta(u),V_{\eta(u)}) - m(\eta(u),V_{\eta(u)}^{n}))^{2}du\right|^{p/2}\right]$$

$$\leq C \mathbb{E}\left[\left|\int_{0}^{t}(m(\eta(u),V_{\eta(u)}) - m(\eta(u),V_{\eta(u)}^{n}))^{p}du\right|\right]$$

$$\leq C \int_{0}^{t} \mathbb{E}[|V_{\eta(u)} - V_{\eta(u)}^{n}|^{p}]du.$$
(A30)

For $\mathbb{E}[|Q_4|^p]$, $\mathbb{E}[|Q_5|^p]$ and $\mathbb{E}[|Q_6|^p]$, they are somewhat similar in the structure of their proof, but we do need to switch out the Cauchy–Schwarz inequality for the BDG inequality. As such, we prove $\mathbb{E}[|Q_4|^p]$ by using the BDG inequality, the Hölder inequality, condition (AA7), Jensen's inequality, condition (BB1) and Lemma A1.

$$\mathbb{E}[|Q_{4}|^{p}] \leq \mathbb{E}\left[\left|\int_{0}^{t} (K_{2}(t,u) - K_{2}(t,\eta(u)))g(u,V_{u})dW_{u}\right|^{p}\right] \\ \leq \mathbb{E}\left[\left|\int_{0}^{t} (K_{2}(t,u) - K_{2}(t,\eta(u)))^{2}g(u,V_{u})^{2}du\right|^{p/2}\right] \\ \leq \mathbb{E}\left[\left|\int_{0}^{t} \left(K_{2}(t,u) - K_{2}(t,\eta(u)))^{2\beta}du\right)^{1/\beta} \left(\int_{0}^{t} g(u,V_{u})^{2\beta^{*}}du\right)^{1/\beta^{*}}\right|^{p/2}\right] \\ \leq C\delta^{(p/2\beta)[(2H-1)\beta+1]} \mathbb{E}\left[\left|\left(\int_{0}^{t} g(u,V_{u})^{2\beta^{*}}du\right)^{1/\beta^{*}}\right|^{p/2}\right] \\ \leq C\delta^{(p/2\beta)[(2H-1)\beta+1]} \int_{0}^{t} \mathbb{E}[|g(u,V_{u})|^{p}]du \\ \leq C\delta^{(p/2\beta)[(2H-1)\beta+1]}.$$
(A31)

Similarly, by the BDG inequality, the Hölder inequality, condition (AA1), Jensen's inequality, and condition (BB1), we have the following:

$$\mathbb{E}[|Q_{5}|^{p}] \leq \mathbb{E}\left[\left|\int_{0}^{t} K_{2}(t,\eta(u))(g(u,V_{u}) - g(\eta(u),V_{\eta(u)}))dW_{u}\right|^{p}\right] \\
\leq \mathbb{E}\left[\left|\int_{0}^{t} K_{2}(t,\eta(u))^{2}(g(u,V_{u}) - g(\eta(u),V_{u}) + g(\eta(u),V_{u}) - g(\eta(u),V_{\eta(u)}))^{2}du\right|^{p/2}\right] \\
\leq \mathbb{E}\left[\left|\int_{0}^{t} K_{2}(t,\eta(u))^{2}(g(\eta(u),V_{u}) - g(\eta(u),V_{\eta(u)})^{2}du\right|^{p/2}\right] \\
+ \mathbb{E}\left[\left|\int_{0}^{t} K_{2}(t,\eta(u))^{2\beta}du\right|^{1/\beta} \left(\int_{0}^{t} (g(u,V_{u}) - g(\eta(u),V_{u}))^{2\beta*}du\right)^{1/\beta*}\right|^{p/2}\right] \\
\leq \mathbb{E}\left[\left|\left(\int_{0}^{t} K_{2}(t,\eta(u))^{2\beta}du\right)^{1/\beta} \left(\int_{0}^{t} (g(\eta(u),V_{u}) - g(\eta(u),V_{\eta(u)}))^{2\beta*}du\right)^{1/\beta*}\right|^{p/2}\right] \\
+ \mathbb{E}\left[\left|\left(\int_{0}^{t} K_{2}(t,\eta(u))^{2\beta}du\right)^{1/\beta} \left(\int_{0}^{t} (g(\eta(u),V_{u}) - g(\eta(u),V_{\eta(u)}))^{2\beta*}du\right)^{1/\beta*}\right|^{p/2}\right] \\
\leq C\delta^{pH} + C \mathbb{E}\left[\int_{0}^{t} (g(\eta(u),V_{u}) - g(\eta(u),V_{\eta(u)}))^{p}du\right].$$
(A32)

Again from condition (BB1) and Lemma A2, we have the following:

$$\mathbb{E}[|Q_5|^p] \le C\delta^{pH} + C \int_0^t \mathbb{E}\Big[\Big|V_u - V_{\eta(u)}\Big]\Big|^{p\kappa} du$$

$$\le C\delta^{pH} + C\delta^{(p\kappa/2\beta)[(2H-1)\beta+1]}$$

$$\le C\delta^{(p\kappa/2\beta)[(2H-1)\beta+1]}$$
(A33)

as $H > (\kappa/2\beta)[(2H-1)\beta+1]$ for $H \in (0,0.5)$, $\beta \in (1,1/(1-2H))$ and $\kappa \in (0,1)$. For the last $\mathbb{E}[|Q_6|^p]$, it can be bounded as follows:

$$\begin{split} \mathbb{E}[|Q_{6}|^{p}] &\leq \mathbb{E}\left[\left|\int_{0}^{t} K_{2}(t,\eta(u))(g(\eta(u),V_{\eta(u)}) - g(\eta(u),V_{\eta(u)}^{n}))dW_{u}\right|^{p}\right] \\ &\leq \mathbb{E}\left[\left|\int_{0}^{t} K_{2}(t,\eta(u))^{2}(g(\eta(u),V_{\eta(u)}) - g(\eta(u),V_{\eta(u)}^{n}))^{2}du\right|^{p/2}\right] \\ &\leq \mathbb{E}\left[\left|\left(\int_{0}^{t} K_{2}(t,\eta(u))^{2\beta}du\right)^{1/\beta}\left(\int_{0}^{t} (g(\eta(u),V_{\eta(u)}) - g(\eta(u),V_{\eta(u)}^{n}))^{2\beta^{*}}du\right)^{1/\beta^{*}}\right|^{p/2}\right] \\ &\leq C \mathbb{E}\left[\left(\int_{0}^{t} |g(\eta(u),V_{\eta(u)}) - g(\eta(u),V_{\eta(u)}^{n})|^{p}du\right)\right] \\ &\leq C \int_{0}^{t} \mathbb{E}\left[\left|V_{\eta(u)} - V_{\eta(u)}^{n}\right|^{p^{*}}\right]du \\ &\leq C \int_{0}^{t} \mathbb{E}\left[\left|V_{\eta(u)} - V_{\eta(u)}^{n}\right|^{p}\right]^{\kappa}du, \end{split}$$
(A34)

where we have used the BDG inequality, the Hölder inequality, Jensen's inequality, condition (AA1), condition (BB1) and Jensen's inequality again. From all the inequalities above, we can then conclude it as follows:

$$\mathbb{E}[|V_{t} - V_{t}^{n}|^{p}] \leq C\delta^{pH} + C\delta^{(p/2\beta)[(2H-1)\beta+1]} + C\delta^{(p\kappa/2\beta)[(2H-1)\beta+1]} + C\int_{0}^{t} \mathbb{E}[|V_{\eta(u)} - V_{\eta(u)}^{n}|^{p}]^{\kappa} du$$

$$\leq C\delta^{(p\kappa/2\beta)[(2H-1)\beta+1]} + C\int_{0}^{t} \mathbb{E}[|V_{\eta(u)} - V_{\eta(u)}^{n}|^{p}] du + C\int_{0}^{t} \mathbb{E}[|V_{\eta(u)} - V_{\eta(u)}^{n}|^{p}]^{\kappa} du.$$
(A35)

Let $\epsilon = (p\kappa/2\beta)[(2H-1)\beta + 1]$ and the following hold:

$$h_n(t) := \sup_{s \in [0,t]} \mathbb{E}[|V_s - V_s^n|^p].$$
(A36)

Then,

$$\mathbb{E}[|V_{t} - V_{t}^{n}|^{p}] \leq h_{n}(t)$$

$$\leq C\delta^{\epsilon} + C \int_{0}^{t} \sup_{s \in [0,u]} \mathbb{E}[|V_{\eta(u)} - V_{\eta(u)}^{n}|^{p}]du$$

$$+ C \int_{0}^{t} \sup_{s \in [0,u]} \mathbb{E}\left[\left|V_{\eta(u)} - V_{\eta(u)}^{n}\right|^{p}\right]^{\kappa}du$$

$$\leq C\delta^{\epsilon} + C \int_{0}^{t} \sup_{s \in [0,u]} \mathbb{E}[|V_{u} - V_{u}^{n}|^{p}]du$$

$$+ C \int_{0}^{t} \sup_{s \in [0,u]} \mathbb{E}\left[|V_{u} - V_{u}^{n}|^{p}\right]^{\kappa}du$$

$$\leq C\delta^{\epsilon} + C \int_{0}^{t} h_{n}(u)du + C \int_{0}^{t} h_{n}(u)^{\kappa}du.$$
(A37)

By Perov's inequality (page 360 in [36]), we obtain the following:

$$\mathbb{E}[|V_t - V_t^n|^p] \le h_n(t)$$

$$\le \left[(C\delta^{\epsilon})^{1-\kappa} \exp(Ct) + C(\exp(Ct) - 1) \right]^{1/(1-\kappa)}.$$
(A38)

That concludes the proof. \Box

Appendix C. Proof of Theorem 2

The following proof is similar to Theorem A1.

Theorem A2. Let $p \ge \frac{2\beta}{\kappa(\beta-1)}$ for $\kappa \in (0,1)$ and $\beta \in \left(1, \frac{1}{1-2H}\right)$; there exists a finite constant C > 0 that depends on T, p, V_0 , and β such that for $t \in [0,T]$ and $n \ge 1$, the following is true:

$$\mathbb{E}[V_t] - \mathbb{E}[V_t^n]|^p \le C\delta^{(1/2\beta)[(2H-1)\beta+1]}.$$
(A39)

Proof. The expectations of V_t and V_t^n take the following form:

$$\mathbb{E}[V_t] = \mathbb{E}[V_0] + \int_0^t K_1(t-u) \mathbb{E}[m(u, V_u)] du$$
(A40)

and

$$\mathbb{E}[V_t^n] = \mathbb{E}[V_0] + \int_0^t K_1(t - \eta(u)) \mathbb{E}[m(\eta(u), V_{\eta(u)}^n)] du.$$
(A41)

It follows that

$$\begin{split} |\mathbb{E}[V_{t}] - \mathbb{E}[V_{t}^{n}]|^{p} &= \left| \int_{0}^{t} K_{1}(t-u) \mathbb{E}[m(u,V_{u})] - K_{1}(t-\eta(u)) \mathbb{E}[m(\eta(u),V_{\eta(u)}^{n}]du \right|^{p} \\ &= \left| \int_{0}^{t} K_{1}(t-u) \mathbb{E}[m(u,V_{u})] + K_{1}(t-\eta(u)) \mathbb{E}[m(u,V_{u})] \\ - K_{1}(t-\eta(u)) \mathbb{E}[m(u,V_{u})] - K_{1}(t-\eta(u)) \mathbb{E}[m(\eta(u),V_{\eta(u)})]du \right|^{p} \\ &+ K_{1}(t-\eta(u)) \mathbb{E}[m(\eta(u),V_{\eta(u)})] - K_{1}(t-\eta(u)) \mathbb{E}[m(\eta(u),V_{\eta(u)})]du \right|^{p} \\ &= \left| \int_{0}^{t} (K_{1}(t-u) - K_{1}(t,\eta(u)) \mathbb{E}[m(u,V_{u})] du \\ &+ \int_{0}^{t} K_{1}(t,\eta(u)) \Big[\mathbb{E}[m(u,V_{u})] - \mathbb{E}[m(\eta(u),V_{\eta(u)}] \Big] du \\ &+ \int_{0}^{t} K_{1}(t,\eta(u)) \Big[\mathbb{E}[m(\eta(u),V_{\eta(u)})] - \mathbb{E}[m(m(\eta(u),V_{\eta(u)})] \Big] du \Big|^{p} \\ &= \left| G_{1} + G_{2} + G_{3} \right|^{p} \\ &\leq C(|G_{1}|^{p} + |G_{2}|^{p} + |G_{3}|^{p}), \end{split}$$
(A42)

where the last inequality is due to the finite form of Jensen's inequality. We start with the bounding of $|G_1|^p$:

$$\begin{aligned} |G_{1}|^{p} &= \left| \int_{0}^{t} (K_{1}(t-u) - K_{1}(t,\eta(u)) \mathbb{E}[m(u,V_{u})]du \right|^{p} \\ &\leq \left| \int_{0}^{t} (K_{1}(t-u) - K_{1}(t,\eta(u))^{2}du \int_{0}^{t} \mathbb{E}[m(u,V_{u})]^{2}du \right|^{p/2} \\ &\leq C \left| \delta^{2H} \int_{0}^{t} \mathbb{E}[m(u,V_{u})]^{2}du \right|^{p/2} \\ &\leq C \delta^{pH} \left| \int_{0}^{t} \mathbb{E}[(1+|V_{u}|)]^{2}du \right|^{p/2} \\ &\leq C \delta^{pH} \left| \int_{0}^{t} \mathbb{E}[(1+|V_{u}|)]^{p}du \right| \\ &\leq C \delta^{pH} \left| \int_{0}^{t} \mathbb{E}[1+|V_{u}|^{p}]du \right| \\ &\leq C \delta^{pH} \left| \int_{0}^{t} 1 + \mathbb{E}[|V_{u}|^{p}]du \right| \\ &\leq C \delta^{pH} \left| \int_{0}^{t} 1 + C(1+\mathbb{E}[|V_{0}|^{p}])du \right| \\ &\leq C \delta^{pH} \right| \end{aligned}$$
(A43)

where we used the Cauchy inequality, condition (AA6), condition (BB1), Jensen's inequality, and Lemma A1. We move on to the bound of $|G_2|^p$, i.e., we used the Cauchy inequality, condition (BB1), Lemma A1, Jensen's inequality, and Lemma A2 to obtain the following:

$$\begin{split} |G_{2}|^{p} &= \left| \int_{0}^{t} K_{1}(t,\eta(u))[\mathbb{E}[m(u,V_{u})] - \mathbb{E}[m(\eta(u),V_{\eta(u)})]]du \right|^{p} \\ &= \left| \int_{0}^{t} K_{1}(t,\eta(u))[\mathbb{E}[m(u,V_{u})] - \mathbb{E}[m(\eta(u),V_{u})]] \\ &+ K_{1}(t,\eta(u))[\mathbb{E}[m(\eta(u),V_{u})] + \mathbb{E}[m(\eta(u),V_{\eta(u)})]]du \right|^{p} \\ &\leq \left| \left(\int_{0}^{t} K_{1}(t,\eta(u))^{2} du \int_{0}^{t} (\mathbb{E}[m(u,V_{u})] - \mathbb{E}[m(\eta(u),V_{u})])^{2} du \right)^{1/2} \\ &+ \left(\int_{0}^{t} K_{1}(t,\eta(u))^{2} du \int_{0}^{t} (\mathbb{E}[m(\eta(u),V_{u})] - \mathbb{E}[m(\eta(u),V_{\eta(u)})])^{2} du \right)^{1/2} \right|^{p} \\ &\leq C \left[\left| \int_{0}^{t} (\mathbb{E}[m(u,V_{u})] - \mathbb{E}[m(\eta(u),V_{u})])^{2} du \right|^{p/2} \\ &+ \left| \int_{0}^{t} (\mathbb{E}[m(\eta(u),V_{u})] - \mathbb{E}[m(\eta(u),V_{\eta(u)})])^{2} du \right|^{p/2} \right] \\ &\leq C \left[\left| \int_{0}^{t} \mathbb{E}[\delta^{H}[1 + |V_{u}|]]^{2} du \right|^{p/2} + \left| \int_{0}^{t} \mathbb{E}[V_{u} - V_{\eta(u)}]^{2} du \right|^{p/2} \right] \\ &\leq C \left[\delta^{pH} \int_{0}^{t} (1 + \mathbb{E}[|V_{u}|^{p}]) du + \int_{0}^{t} |\mathbb{E}[V_{u} - V_{\eta(u)}]|^{p} du \right] \\ &\leq C \delta^{pH} + C \int_{0}^{t} \mathbb{E}[|V_{u} - V_{\eta(u)}|^{p}] du \\ &\leq C \delta^{pH} + C \delta^{(p/2\beta)[(2H-1)\beta+1]}. \end{split}$$
(A44)

Since $H \ge (1/2\beta)[(2H-1)\beta+1]$, we can conclude the bound for $|G_2|^p$ as follows:

$$|G_2|^p \le C\delta^{(p/2\beta)[(2H-1)\beta+1]}.$$
(A45)

Then, by using the Cauchy inequality and condition (AA1), the bound $|G_3|^p$ can be formulated as follows:

$$\begin{aligned} |G_{3}|^{p} &= \left| \int_{0}^{t} K_{1}(t,\eta(u)) \left[\mathbb{E}[m(\eta(u),V_{\eta(u)})] - \mathbb{E}[m(\eta(u),V_{\eta(u)}^{n})] \right] du \right|^{p} \\ &= \left| \int_{0}^{t} K_{1}(t,\eta(u))^{2} du \int_{0}^{t} \left[\mathbb{E}[m(\eta(u),V_{\eta(u)})] - \mathbb{E}[m(\eta(u),V_{\eta(u)}^{n})] \right]^{2} du \right|^{p/2} \\ &\leq C \int_{0}^{t} \left| \mathbb{E}[m(\eta(u),V_{\eta(u)})] - \mathbb{E}[m(\eta(u),V_{\eta(u)}^{n})] \right|^{p} du \\ &\leq C \int_{0}^{t} \left| \mathbb{E}[V_{\eta(u)}] - \mathbb{E}[V_{\eta(u)}^{n}] \right|^{p} du. \end{aligned}$$
(A46)

Combining the bounds $|G_1|$, $|G_2|$, and $|G_3|$, we have the following:

$$\begin{split} |\mathbb{E}[V_{t}] - \mathbb{E}[V_{t}^{n}]|^{p} &\leq C\delta^{pH} + C\delta^{(p/2\beta)[(2H-1)\beta+1]} \\ &+ C\int_{0}^{t} |\mathbb{E}[V_{\eta}(u)] - \mathbb{E}[V_{\eta(u)}^{n}]|^{p}du \\ &\leq C\delta^{(p/2\beta)[(2H-1)\beta+1]} + C\int_{0}^{t} |\mathbb{E}[V_{\eta(u)}] - \mathbb{E}[V_{\eta(u)}^{n}]|^{p}du. \end{split}$$
(A47)

Suppose that we let the following hold:

$$g_n(t) := \sup_{s \in [0,t]} \left| \mathbb{E}[V_s] - \mathbb{E}[V_s^n] \right|^p,$$
(A48)

and then

$$\begin{split} \left| \mathbb{E}[V_{t}] - \mathbb{E}[V_{t}^{n}] \right|^{p} &\leq g_{n}(t) \\ &\leq C\delta^{(p/2\beta)[(2H-1)\beta+1]} + C \sup_{s \in [0,t]} \int_{0}^{s} \left| \mathbb{E}[V_{\eta(u)}] - \mathbb{E}[V_{\eta(u)}^{n}] \right|^{p} du \\ &\leq C\delta^{(p/2\beta)[(2H-1)\beta+1]} + C \int_{0}^{t} \sup_{s \in [0,u]} \left| \mathbb{E}[V_{\eta(s)}] - \mathbb{E}[V_{\eta(s)}^{n}] \right|^{p} du \\ &\leq C\delta^{(p/2\beta)[(2H-1)\beta+1]} + C \int_{0}^{t} g_{n}(u) du \\ &\leq C\delta^{(p/2\beta)[(2H-1)\beta+1]}, \end{split}$$
(A49)

where the last inequality is due to Grönwall's inequality. Finally, apply the power of 1/p to both sides of the inequality, and the result follows. \Box

References

- 1. Gatheral, J.; Jaisson, T.; Rosenbaum, M. Volatility is rough. Quant. Financ. 2018, 18, 933–949. [CrossRef]
- Comte, F.; Renault, E. Long memory in continuous-time stochastic volatility models. *Math. Financ.* 1998, *8*, 291–323. [CrossRef]
 Cheridito, P. Mixed fractional Brownian motion. *Bernoulli* 2001, *7*, 913–934. [CrossRef]
- 4. Alòs, E.; León, J.A.; Vives, J. On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Financ. Stoch.* 2007, 11, 571–589. [CrossRef]
- 5. Fukasawa, M. Asymptotic analysis for stochastic volatility: Martingale expansion. Financ. Stoch. 2011, 15, 635–654. [CrossRef]
- 6. Bayer, C.; Friz, P.; Gatheral, J. Pricing under rough volatility. *Quant. Financ.* 2016, 16, 887–904. [CrossRef]
- 7. Bergomi, L. Smile Dynamics II. SSRN 2005. Available online: http://dx.doi.org/10.2139/ssrn.1493302 (accessed on 3 May 2021).
- 8. El Euch, O.; Fukasawa, M.; Rosenbaum, M. The microstructural foundations of leverage effect and rough volatility. *Financ. Stoch.* **2018**, *22*, 241–280. [CrossRef]
- 9. El Euch, O.; Rosenbaum, M. The characteristic function of rough Heston models. Math. Financ. 2019, 29, 3–38. [CrossRef]
- 10. El Euch, O.; Gatheral.; Radoicic, R.; Rosenbaum, M. The zumbach effect under rough heston. *arXiv* **2018**, arXiv:1809.02098. Available online: https://arxiv.org/abs/1809.02098 (accessed on 2 April 2021).
- 11. Gatheral, J.; Radoicic, R. Rational approximation of the rough Heston solution. *Int. J. Theor. Appl. Financ.* **2019**, *22*, 1950010. [CrossRef]
- 12. Jeng, S.W.; Kilicman, A. Approximation formula for option prices under rough heston model and short-time implied volatility behaviour. *Symmetry* **2020**, *12*, 1878. [CrossRef]
- 13. Jeng, S.W.; Kilicman, A. Fractional Riccati Equation and Its Applications to Rough Heston Model Using Numerical Methods. *Symmetry* **2020**, *12*, 959. [CrossRef]
- 14. Jeng, S.W.; Kilicman, A. Series Expansion and Fourth-Order Global Padé Approximation for a Rough Heston Solution. *Mathematics* **2020**, *8*, 1968. [CrossRef]
- 15. Abi Jaber, E. Lifting the Heston model. Quant. Financ. 2019, 19, 1995–2013. [CrossRef]
- 16. Giles, M.B. Multilevel monte carlo path simulation. Oper. Res. 2008, 56, 607–617. [CrossRef]
- 17. Giles, M.B. Multilevel monte carlo methods. Acta Numer. 2015, 24, 259. [CrossRef]
- 18. Glasserman, P. *Monte Carlo Methods in Financial Engineering;* Springer Science & Business Media: Berlin/Heidelberg, Germany, 2003; Volume 53.
- 19. Liu, G.; Zhao, Q.; Gu, G. A simple control variate method for options pricing with stochastic volatility models. *Int. J. Appl. Math.* **2015**, *8*, 1968.
- 20. McCrickerd, R.; Pakkanen, M.S. Turbocharging Monte Carlo pricing for the rough Bergomi model. *Quant. Financ.* 2018, 18, 1877–1886. [CrossRef]
- Richard, A.; Tan, X.; Yang, F. Discrete-time simulation of stochastic volterra equations. *arXiv* 2020, arXiv:2004.00340. Available online: https://arxiv.org/abs/2004.00340 (accessed on 3 May 2021).
- 22. El Euch, O.; Rosenbaum, M. Perfect hedging in rough Heston models. Ann. Appl. Probab. 2018, 28, 3813–3856. [CrossRef]
- 23. Forde, M.; Gerhold, S.; Smith, B. Small-time and large-time smile behaviour for the Rough Heston model. *arXiv* 2019, arXiv:1906.09034. Available online: https://arxiv.org/abs/1906.09034 (accessed on 25 April 2021).
- 24. Gatheral, J.; Keller-Ressel, M. Affine forward variance models. Financ. Stoch. 2018, 23, 501–533. [CrossRef]

- 25. Maruyama, G. Continuous markov processes and stochastic equations. Rend. Del Circ. Mat. Di Palermo 1955, 4, 48. [CrossRef]
- 26. Milstein, G. Approximate integration of stochastic differential equations. *Theory Probab. Its Appl.* **1975**, *19*, 557–562.
- Richard, A.; Tan, X.; Yang, F. On the discrete-time simulation of the rough Heston model. *arXiv* 2021, arXiv:2107.07835. Available online: https://arxiv.org/abs/2107.07835 (accessed on 3 May 2021).
- 28. Milstein, G. A method of second-order accuracy integration of stochastic differential equations. *Theory Probab. Its Appl.* **1979**, *23*, 396–401. [CrossRef]
- 29. Talay, D. Efficient numerical schemes for the approximation of expectations of functionals of the solution of a sde, and applications. In *Filtering and Control of Random Processes*; Springer: Berlin/Heidelberg, Germany, 1984; pp. 294–313.
- Shreve, S.E. Stochastic Calculus for Finance II: Continuous Time Models; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2004; Volume 11.
- 31. Fairbanks, H.R.; Doostan, A.; Ketelsen, C.; Iaccarino, G. A low-rank control variate for multilevel Monte Carlo simulation of high-dimensional uncertain systems. *J. Comput. Phys.* 2017, 341, 121–139. [CrossRef]
- 32. Nobile, F.; Tesei, F. A Multi Level Monte Carlo method with control variate for elliptic PDEs with log-normal coefficients. *Stoch. Partial Differ. Equ. Anal. Comput.* **2015**, *3*, 398–444. [CrossRef]
- Rømer, S.E. The Hybrid-Exponential Scheme for Stochastic Volterra Equations. SSRN 2020, ssrn.3706253. Available online: http://dx.doi.org/10.2139/ssrn.3706253 (accessed on 23 April 2021).
- 34. Bennedsen, M.; Lunde, A.; Pakkanen, M.S. Hybrid scheme for Brownian semistationary processes. *Financ. Stoch.* 2017, 21, 931–965. [CrossRef]
- 35. Alfonsi, A. On the discretization schemes for the CIR (and Bessel squared) processes. De Gruyter 2005, 355–384. [CrossRef]
- 36. Mitrinovic, D.S.; Pecaric, J.; Fink, A.M. *Inequalities Involving Functions and Their Integrals and Derivatives*; Springer Science & Business Media : Berlin/Heidelberg, Germany, 2012; Volume 53.