Multilevel Monte Carlo Simulation for the Heston Stochastic Volatility Model

Chao Zheng *

Abstract

We combine the multilevel Monte Carlo (MLMC) method with the numerical scheme for the Heston model that simulates the variance process exactly or almost exactly and applies the stochastic trapezoidal rule to approximate the time-integrated variance process within the SDE of the logarithmic asset process. We conduct separate simulation for path-independent options and path-dependent options. In both situations, novel MLMC estimators are established, and the theoretical convergence rates are derived for the full parameter regime. The analysis is supported by the numerical results.

Keywords: Heston model, multilevel Monte Carlo, convergence rate
AMS subject classifications (2000): 60H35, 65C30, 91G60

1 Introduction

The Heston stochastic volatility model, introduced by Heston (1993), is one of the most fundamental models in mathematical finance, of theoretical and practical interest. It is a two-dimensional SDE with two components: the variance process and the asset process. This model is widely applied in various financial markets, such as the markets of equity, fixed income and foreign exchange, due to its tractability in modelling the term structure of implied volatility. Although the price of a standard European option can be expressed as a complex integral under the Heston model, which is easy to calculate, the majority of options cannot be priced in closed form. Thus Monte Carlo simulation becomes essential. It is well known that the Heston model admits a unique strong solution, although the closed-form expression is not available in the literature. The traditional time-discrete schemes, such as those in Kloeden and Platen (1999), often produce erratic results when applied to the Heston model, because its diffusion term does not satisfy the Lipschitz condition.

To address this problem, many Monte Carlo methods have been proposed for the Heston model, and the results can typically be placed into three categories. First, one technology is to sample the end point of the variance process, and then simulate the time-integrated variance process conditional on the initial and end points. An early contribution was due to Broadie and Kaya (2006), and the method was further developed by Glasserman and Kim (2011). They are useful for options

*School of Data Sciences, Zhejiang University of Finance and Economics, Hangzhou, China (chao.zheng12@gmail.com). This research is supported partially by National Science Foundation of China grant 11801504.

with path-independent payoffs, but might be computationally inconvenient for options with path-dependent payoffs. Second, an alternative approach is to develop numerical schemes that discretize the SDE of the variance process and that of the asset process. These works include Alfonsi (2010), Higham and Mao (2005), Kahl and Jäckel (2006) and a number of Euler schemes discussed in Lord, Koekkoek and Van Dijk (2010).

Third, since the transition density of the variance process is known explicitly, there is another category, which consists of schemes that use small-biased or almost exact approximations of the variance process and apply the trapezoidal rule to the time-integrated variance process in the SDE of the logarithmic asset process. Andersen (2008) made a breakthrough in approximating the variance process as a Gaussian-type random variable with the coefficients calculated through a moment matching technique, which is known as the quadratic exponential (QE) scheme. The QE scheme is highly efficient in Monte Carlo simulation of the Heston model. Van Haastrecht and Pelsser (2010) implemented an efficient caching technique to approximate the variance process. Malham and Wiese (2013) proved that the variance process can be represented by the sum of several generalized Gaussian random variables, leading to a direct inversion scheme. The last two methods are almost exact, and their efficiency is demonstrated to be comparable to Andersen (2008) for the Heston model.

In a parallel development, the multilevel Monte Carlo (MLMC) method was proposed by Giles (2008b), which can be regarded as a variance reduction technique. It has been shown to provide a significant improvement in computational efficiency over the traditional Monte Carlo method on many time-discrete schemes for SDEs (e.g., Giles 2008b for the Euler scheme and Giles 2008a for the Milstein scheme). An overview of MLMC can be found in Giles (2015). Thus, a challenging question is how to combine MLMC with an efficient numerical scheme for the Heston model. The literature includes attempts to combine them, and these MLMC methods are typically based on time-discrete schemes in the second category (e.g., Giles 2008b, Kloeden and Neuenkirch 2012, Giles and Szpruch 2014 and Altmayer and Neuenkirch 2015). Nevertheless, these MLMC methods seem to be restricted to the parameter regime where the zero boundary of the variance process is not attainable, whereas in practice, it is often observed that the variance process is attainable and reflecting (e.g., Andersen 2008). As is well known, an efficient MLMC method requires that the MLMC variance decays with a high-order convergence when the level goes to infinity, i.e., when the step size approaches zero. However, a recent study by Hefter and Jentzen (2019) reveals that the approximation of the variance process through explicit or implicit Euler- or Milstein-type discretization methods with equidistant step sizes would typically lead to arbitrarily slow strong convergence rates, which makes them inconvenient for MLMC applications. To the best of our knowledge, such a MLMC estimator that has a stable convergence rate for the full parameter regime of the Heston model is missing in the literature.

In this article, we consider a different approach by developing MLMC based on a time-discrete scheme that simulates the variance process exactly. Specifically, we combine MLMC with the numerical scheme for the Heston model, which simulates the variance process exactly and approximates the time-integrated variance process within the SDE of the logarithmic asset price by the trapezoidal rule. The numerical scheme we consider is consistent with the methods in the third category. We separately define novel MLMC estimators for path-independent options and for path-dependent options. Moreover, we prove that, under some Lipschitz assumptions on the payoff, the convergence rate of the MLMC variance is 2 for path-independent options and 1 for path-dependent options. The results apply for all parameter regimes.

There are two more important advantages of the MLMC estimators we develop. One is that

the algorithm is very easy to implement and can be used for a relatively large class of options. The other is that the MLMC estimators do not rely on which exact method we choose for the variance process, and are typically applicable when the variance process is simulated almost exactly. This feature is attractive in practice, because there are several efficient almost exact methods for the variance process, such as the NCI method from Van Haastrecht and Pelsser (2010). To illustrate the efficiency of our MLMC estimators, we further combine them with the NCI method, and make a comparison with some leading methods in the literature. In particular, for the path-independent case, we choose the gamma expansion method of Glasserman and Kim (2011) as the benchmark, and for the path-dependent case, we compare it with the QE scheme of Andersen (2008). The improvement in both cases is confirmed by the numerical results.

The article is organized as follows: In section 2, we introduce some basic facts of the Heston model and MLMC. In section 3, novel MLMC estimators are introduced for path-independent options and for path-dependent options. Their corresponding theoretical convergence rates are derived. Section 4 reports some numerical results. Section 5 extends the MLMC estimators for more stochastic volatility models. Finally, we conclude the article in section 6.

2 Preliminaries

2.1 Heston stochastic volatility model

The dynamics of the Heston stochastic volatility model under a risk-neutral probability measure is as follows

$$\begin{aligned}\mathrm{d}S_t &= rS_t\mathrm{d}t + \sqrt{V_t}S_t(\rho\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2)\\ \mathrm{d}V_t &= k(\theta - V_t)\mathrm{d}t + \sigma\sqrt{V_t}\mathrm{d}W_t^1,\end{aligned}$$

in which $(S_t)_{t\geq 0}$ is the asset process, representing the price of an asset, and $(V_t)_{t\geq 0}$ is the variance process, standing for the evolution of the volatility, with the initial values $S_0 > 0, V_0 \geq 0$. The parameters $k, \theta, \sigma > 0$, $r \in \mathbb{R}$ and $\rho \in [-1, 1]$ are constant. Here, $(W_t^1)_{t\geq 0}$ and $(W_t^2)_{t\geq 0}$ are two independent Brownian motions.

The asset price S_t at time t, given the values S_u and V_u at time $u < t$, can be written as

$$S_t = S_u \exp\left(r(t-u) - \frac{1}{2}\int_u^t V_s\mathrm{d}s + \rho\int_u^t \sqrt{V_s}\mathrm{d}W_s^1 + \sqrt{1-\rho^2}\int_u^t \sqrt{V_s}\mathrm{d}W_s^2\right),$$

and V_t is given by

$$V_t = V_u + k\theta(t-u) - k\int_u^t V_s\mathrm{d}s + \sigma\int_u^t \sqrt{V_s}\mathrm{d}W_s^1.$$

Conditioned on V_u, the variance process V_t at time t, $t \geq u$, follows a scaled non-central chi-squared distribution. More precisely, we have

$$V_t \overset{d}{=} \frac{\sigma^2(1-\mathrm{e}^{-k(t-u)})}{4k}\chi_d^2\left(\frac{4k\mathrm{e}^{-k(t-u)}}{\sigma^2(1-\mathrm{e}^{-k(t-u)})}V_u\right),$$

where $\chi_d^2(\lambda)$ denotes a non-central chi-squared random variable with $d := \frac{4\theta k}{\sigma^2}$ degrees of freedom and non-centrality parameter λ. From the equation for V_t, we find

$$\int_u^t \sqrt{V_s}\mathrm{d}W_s^1 = \frac{1}{\sigma}\left(V_t - V_u - k\theta(t-u) + k\int_u^t V_s\mathrm{d}s\right).$$

Substituting it into the equation for S_t, we get

$$\begin{aligned} S_t &= S_u \exp\left[r(t-u) - \frac{1}{2}\int_u^t V_s \mathrm{d}s + \frac{\rho}{\sigma}\left(V_t - V_u - k\theta(t-u) + k\int_u^t V_s \mathrm{d}s\right)\right. \\ &\quad \left. + \sqrt{1-\rho^2}\int_u^t \sqrt{V_s}\mathrm{d}W_s^2\right] \\ &\stackrel{d}{=} S_u \exp\left[r(t-u) - \frac{1}{2}\int_u^t V_s \mathrm{d}s + \frac{\rho}{\sigma}\left(V_t - V_u - k\theta(t-u) + k\int_u^t V_s \mathrm{d}s\right)\right. \\ &\quad \left. + \sqrt{1-\rho^2}\sqrt{\int_u^t V_s \mathrm{d}s}N\right], \end{aligned}$$

where N represents a standard normal random variable. The last step is valid because V is independent of W^2. Then given $\int_u^t V_s \mathrm{d}s$, the integral $\int_u^t \sqrt{V_s}\mathrm{d}W_s^2$ is normal distributed with mean 0 and variance $\int_u^t V_s \mathrm{d}s$. The formula for S_t above is due to Broadie and Kaya (2006). The integral $\int_u^t V_s \mathrm{d}s$ can be approximated by the trapezoidal rule.

2.2 Simulations for path-independent and path-dependent options

Under the Heston model, the time-discrete scheme for pricing path-independent options and that for pricing path-dependent options are slightly different. This is because the prices of path-independent options only rely on the asset price S_T at maturity T. Therefore, it is convenient to call the simulation for S_T only at maturity 'path-independent simulation', and that for the path of S 'path-dependent simulation'.

To be more specific, for path-independent options, denote by $\tilde{S}_{T,h}$ the approximation of the asset price S_T at time T with step size h, and we have

$$\begin{aligned} \ln \tilde{S}_{T,h} = \ln S_0 + \left[rT - \frac{1}{2}\fint_0^T V_s^h \mathrm{d}s + \frac{\rho}{\sigma}\left(V_T - V_0 - k\theta T + k\fint_0^T V_s^h \mathrm{d}s\right)\right. \\ \left. + \sqrt{1-\rho^2}\sqrt{\fint_0^T V_s^h \mathrm{d}s}N\right], \end{aligned}$$

where N is a standard normal random variable, and $\fint_0^T V_s^h \mathrm{d}s$ denotes the trapezoidal approximation on $\int_0^T V_s \mathrm{d}s$ with step size h, i.e.,

$$\fint_0^T V_s^h \mathrm{d}s = \frac{V_0 + V_T}{2}h + \sum_{i=1}^{T/h-1} V_{hi}h.$$

On the other hand, for path-dependent options, we have to simulate the asset process path-wisely,

and therefore we approximate S_H, $H = h, 2h, ..., T$ by $\hat{S}_H^h$ based on the discretization

$$\ln \hat{S}_H^h = \ln S_0 + \left[rH - \frac{1}{2} \fint_0^H V_s^h \mathrm{d}s + \frac{\rho}{\sigma} \left(V_H - V_0 - k\theta H + k \fint_0^H V_s^h \mathrm{d}s \right) + \sqrt{1-\rho^2} \sum_{i=1}^{H/h} \sqrt{\frac{V_{h(i-1)} + V_{hi}}{2}} h N_i \right],$$

where N_i, $i = 1, 2, ...H/h$ are independent standard normal random variables, and $\fint_0^H V_s^h \mathrm{d}s$ is the trapezoidal approximation on $\int_0^H V_s \mathrm{d}s$ with step size h.

2.3 Multilevel Monte Carlo

The multilevel Monte Carlo (MLMC) method can be regarded as a variance reduction technique of the standard Monte Carlo, introduced by Giles (2008b). The idea of MLMC is as follows: It uses several approximations through the same discretization scheme but with different step sizes and different numbers of samples. Specifically, more samples are taken for the discretization of large step sizes (coarse levels) with low accuracy but less computational cost, while less samples are taken for that of small step sizes (fine levels) with high accuracy but more cost. In this way, it minimizes the overall computational cost with a given accuracy.

Denote by P the payoff of an option, and $\hat{P}_l$ its approximation on level l, $l = 0, 1, 2, ...$, based on a time-discrete scheme with step size $h_l = M^{-l}T$, where $M \geq 2$ is an integer, then the expected value $\mathbb{E}(\hat{P}_L)$ can be written as

$$\mathbb{E}(\hat{P}_L) = \mathbb{E}(\hat{P}_0) + \sum_{l=1}^{L} \mathbb{E}(\hat{P}_l - \hat{P}_{l-1}).$$

The MLMC estimates each of the expectations independently. Let Y_0 be an estimator for $\mathbb{E}(\hat{P}_0)$ using N_0 samples, and Y_l be an estimator for $\mathbb{E}(\hat{P}_l - \hat{P}_{l-1})$ using N_l paths. One can define

$$Y_0 = N_0^{-1} \sum_{i=1}^{N_0} \hat{P}_0^{(i)}, \tag{1}$$

and

$$Y_l = N_l^{-1} \sum_{i=1}^{N_l} (\hat{P}_l^{(i)} - \hat{P}_{l-1}^{(i)}), \tag{2}$$

for $l > 0$, where $\hat{P}_l^{(i)}$ is the i-th sample of $\hat{P}_l$. Then, the combined MLMC estimator Y for $\mathbb{E}(\hat{P}_L)$ is

$$Y = \sum_{l=0}^{L} Y_l. \tag{3}$$

The variance of Y_l, $l \geq 0$, is $\mathrm{Var}(Y_l) = N_l^{-1} \hat{V}_l$, where $\hat{V}_0 := \mathrm{Var}(\hat{P}_0)$ and $\hat{V}_l := \mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$, $l > 0$. Thus, the variance of Y is $\mathrm{Var}(Y) = \sum_{l=0}^{L} N_l^{-1} \hat{V}_l$. The computational complexity of MLMC is provided by the following theorem from Giles (2008b):

Theorem 2.1. (Giles). *Let P denote a functional of the solution of stochastic differential equation for a given Brownian path, and $\hat{P}_l$ denote the corresponding approximation using a numerical discretisation with time step $h_l = M^{-l}T$.*

If there exist independent estimators Y_l based on N_l Monte Carlo samples, and positive constant $\beta \geq \frac{1}{2}$, α, c_1, c_2, c_3 such that

(i) $\mathbb{E}(\hat{P}_l - P) \leq c_1 h_l^{\beta}$

(ii) $\mathbb{E}(Y_l) = \begin{cases} \mathbb{E}(\hat{P}_0), & l = 0 \\ \mathbb{E}(\hat{P}_l - \hat{P}_{l-1}), & l > 0 \end{cases}$

(iii) $\mathrm{Var}(Y_l) \leq c_2 N_l^{-1} h_l^{\alpha}$

(iv) C_l, the computational complexity of Y_l, is bounded by

$$C_l \leq c_3 N_l h_l^{-1},$$

then there exists a positive constant c_4 such that for any $\xi < e^{-1}$ there are values L and N_l for which the multilevel estimator (3) has a mean-square-error with the bound

$$\mathrm{MSE} = \mathbb{E}[(Y - \mathbb{E}(P))^2] < \xi^2$$

with a computational complexity C with bound

$$C \leq \begin{cases} c_4 \xi^{-2}, & \alpha > 1 \\ c_4 \xi^{-2} (\ln \xi)^2, & \alpha = 1 \\ c_4 \xi^{-2-(1-\alpha)/\beta}, & 0 < \alpha < 1. \end{cases}$$

Note that the values L and N_l in Theorem 2.1 need to be specified. Giles (2008b) shows that the optimal N_l in (1) and (2) is

$$N_l = \left\lceil 2\xi^{-2} \sqrt{\hat{V}_l h_l} \left(\sum_{l=0}^{L} \sqrt{\hat{V}_l / h_l} \right) \right\rceil, \tag{4}$$

so that $V(Y) \leq \frac{1}{2}\xi^2$, where $\lceil x \rceil$ represents the smallest integer that is no smaller than x. The value L is determined in the algorithm for MLMC. The algorithm has been provided in Giles (2008b) for a numerical scheme with $\beta = 1$, and can be extended for that with a higher β. Asymptotically, as $l \to \infty$, we have

$$\mathbb{E}(P - \hat{P}_l) \approx c_1 h_l^{\beta},$$

for some constant c_1, and hence

$$\begin{aligned} \mathbb{E}(\hat{P}_l - \hat{P}_{l-1}) &= \mathbb{E}(P - \hat{P}_{l-1}) - \mathbb{E}(P - \hat{P}_l) \\ &\approx (M^{\beta} - 1) c_1 h_l^{\beta} \\ &\approx (M^{\beta} - 1) \mathbb{E}(P - \hat{P}_l). \end{aligned} \tag{5}$$

One increases the value for L until

$$|Y_L| < \frac{1}{\sqrt{2}} (M^{\beta} - 1) \xi, \tag{6}$$

where Y_L is defined in (2) for $l = L$. From (5) and (6), the magnitude of the bias is expected to be less than $\xi/\sqrt{2}$. The choice of optimal N_l in (4) guarantees $\mathrm{Var}(Y) \leq \frac{1}{2}\xi^2$. Therefore, it is expected that the mean squared error is less than ξ^2. However, since (5) is only an approximate equality, there is no guarantee to achieve a MSE error less than ξ^2. The numerical algorithm is provided below:

1. Start with $L = 0$.
2. Estimate V_L using an initial set of $N_L = 10000$ samples.
3. Define optimal N_l, $l = 0, ..., L$, using equation (4).
4. Evaluate extra samples at each level as needed for new N_l.
5. $L \geq 2$, test for convergence using the formula (6).
6. If $L < 2$, or it is not converged, set $L = L + 1$ and go to Step 2.

There are several important features of MLMC. First, for the standard Monte Carlo associated with a time-discrete scheme, the simulation accuracy relies on the step size and the number of samples, whereas for MLMC, the step size and number of samples are optimized in the MLMC algorithm, and the accuracy is controlled by the user-specified input parameter ξ. Second, as we can see from Theorem 2.1, the computational complexity C relies on the convergence rate of the MLMC variance α and the weak convergence rate β. The rate β is a minor issue, as the theorem only requires $\beta \geq \frac{1}{2}$, which is typically satisfied for time-discrete schemes. The rate α has more influence on the computational complexity and is thus more important for MLMC. Therefore, the essential question is how to couple $\hat{P}_{l-1}$ at the coarse level and $\hat{P}_l$ at the fine level such that there is a large α. The standard approach is to keep $\hat{P}_{l-1}$ and $\hat{P}_l$ from two discrete approximations with different time steps but with the same Brownian motion path. However, for some time-discrete schemes, such as the stochastic trapezoidal rule we consider with one component simulated exactly, there is more flexibility to connect $\hat{P}_{l-1}$ and $\hat{P}_l$, which will be discussed in the next section. Third, to avoid the introduction of additional bias, we shall make sure that $\hat{P}_l$ in $\mathbb{E}(\hat{P}_l - \hat{P}_{l-1})$ and in $\mathbb{E}(\hat{P}_{l+1} - \hat{P}_l)$ have the same expectation $\mathbb{E}(\hat{P}_l)$, as discussed in Giles (2008a).

In summary, MLMC is a variance reduction technique for Monte Carlo simulation of stochastic differential equations. To combine a time-discrete scheme for the Heston model with MLMC, the challenge is to properly couple the fine path $\hat{P}_l$ and the coarse path $\hat{P}_{l-1}$, such that there is a large convergence rate α of $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$. That is our main target.

3 Novel MLMC estimators for the Heston model

3.1 Simulation for path-independent options

To price a path-independent option with MLMC, we have to estimate $\mathbb{E}(P(\tilde{S}_{T,Mh}) - P(\tilde{S}_{T,h}))$, where $P : [0, +\infty) \to [0, +\infty)$ is the option payoff. We simply let the asset price $\tilde{S}_{T,Mh}$ at the coarse level keep the same standard normal random variable N as required to simulate $\tilde{S}_{T,h}$ at the

fine level. Specifically, at the fine level, we have

$$\ln \tilde{S}_{T,h} = \ln S_0 + \left[rT - \frac{1}{2} \oint_0^T V_s^h \mathrm{d}s + \frac{\rho}{\sigma} \left(V_T - V_0 - k\theta T + k \oint_0^T V_s^h \mathrm{d}s \right) + \sqrt{1-\rho^2} \sqrt{\oint_0^T V_s^h \mathrm{d}s} N \right],$$

and at the coarse level, we take

$$\ln \tilde{S}_{T,Mh} = \ln S_0 + \left[rT - \frac{1}{2} \oint_0^T V_s^{Mh} \mathrm{d}s + \frac{\rho}{\sigma} \left(V_T - V_0 - k\theta T + k \oint_0^T V_s^{Mh} \mathrm{d}s \right) + \sqrt{1-\rho^2} \sqrt{\oint_0^T V_s^{Mh} \mathrm{d}s} N \right],$$

where N in both equations are the same standard normal random variable. This estimator applies to any path-independent option, although in the theoretical analysis of $\mathrm{Var}(P(\tilde{S}_{T,Mh}) - P(\tilde{S}_{T,h}))$, we have to impose the Lipschitz assumption on the payoff function.

3.2 Convergence analysis for path-independent simulation

Throughout the analysis, we should denote by c a constant regardless its value. Let $P : [0, +\infty) \to [0, +\infty)$ be a Lipschitz payoff function, such that

$$|P(U) - P(V)| \leq c|U - V|$$

for all $U, V \geq 0$. Then, we have $\mathbb{E}[(P(U) - P(V))^2] \leq c\mathbb{E}[(U - V)^2]$, where $\mathbb{E}[(U - V)^2]$ is usually more convenient for the error analysis. For notational simplicity, we may write $\mathbb{E}[(U - V)^2]$ as $\mathbb{E}(U - V)^2$. Typically, for a put option with a Lipschitz payoff and any $U, V > 0$, we further have

$$\mathbb{E}(P(U) - P(V))^2 \leq c\mathbb{E}(\ln U - \ln V)^2, \tag{7}$$

which is justified by Theorem 3.1 below.

Theorem 3.1. *Let $P : [0, +\infty) \to [0, +\infty)$ be a Lipschitz continuous function. Suppose there is a constant $U_{max} > 0$, such for all $U \geq U_{max}$, we have $P(U) = P(U_{max})$. Then*

$$|P(U) - P(V)| \leq c|\ln U - \ln V|$$

for any $U, V > 0$.

Proof. We shall firstly consider the case where both $U, V \in (0, U_{max}]$, and it follows by the Lipschitz continuity that

$$|P(U) - P(V)| \leq c|U - V| \leq c|\ln U - \ln V|, \tag{8}$$

Next, we consider the case where $V \in (0, U_{max}]$ and $U \in (U_{max}, +\infty)$, and by (8), we have

$$|P(U) - P(V)| = |P(U_{max}) - P(V)| \leq c|\ln U_{max} - \ln V| < c|\ln U - \ln V|. \tag{9}$$

The analysis of the situation where $U \in (0, U_{max}]$ and $V \in (U_{max}, +\infty)$ is analogous. When both $U, V \in (U_{max}, +\infty)$, we obtain

$$|P(U) - P(V)| = 0.$$

Then the proof is complete. □

Remark 3.1. *The condition $P(U) = P(U_{max})$, when $U \geq U_{max}$, implies that the payoff must be bounded, which is particularly relevant to the put option. The put option becomes worthless when the price of the underlying asset is sufficiently large.*

In our case, we let $U = \tilde{S}_{T,Mh}$, and $V = \tilde{S}_{T,h}$. It follows by (7) that

$$\mathbb{E}(P(\tilde{S}_{T,Mh}) - P(\tilde{S}_{T,h}))^2 \leq c\mathbb{E}(\ln \tilde{S}_{T,Mh} - \ln \tilde{S}_{T,h})^2, \tag{10}$$

which is essential to estimate the computational complexity of MLMC. For call options, the inequality (10) is usually not satisfied. However, for many types of options in finance (e.g., the standard European option, the digital option), we have the put-call parity, which says the difference of the prices between a call option and its corresponding put option with the same strike and the same maturity is known at the present time. This means the computational complexity to calculate a call option price, through the use of the put-call parity, is just the same as that to compute its corresponding put option price.

Note that V_t at time t follows a scaled non-central Chi-squared distribution, and we have (e.g., Dereich, Neuenkirch and Szpruch 2012)

$$\sup_{t\in[0,T]} \mathbb{E}\left(V_t^p\right) < \infty$$

for all $p > -2k\theta/\sigma^2$ and $T > 0$. To derive the convergence rate for (10), we further need the lemma below from Theorem 4.1(a) in Dufresne (2001).

Lemma 3.1. (Dufresne). *For any $\tau \in \mathbb{R}$, and $T > 0$, we have $\mathbb{E}\left(\int_0^T V_s \mathrm{d}s\right)^\tau < \infty$.*

Theorem 3.2. *Let $T > 0$ and $M \in \mathbb{N}^+$. We have*

$$\mathbb{E}\left(\ln \tilde{S}_{T,Mh} - \ln \tilde{S}_{T,h}\right)^2 = O(h^2).$$

Proof. It holds that

$$\mathbb{E}\left(\ln \tilde{S}_{T,Mh} - \ln \tilde{S}_{T,h}\right)^2 \leq 2\mathbb{E}\left(\ln \tilde{S}_{T,Mh} - \ln S_T\right)^2 + 2\mathbb{E}\left(\ln \tilde{S}_{T,h} - \ln S_T\right)^2, \tag{11}$$

where the exact solution $\ln S_T$ can be written as

$$\ln S_T = \ln S_0 + \left[rT - \frac{1}{2}\int_0^T V_s \mathrm{d}s + \frac{\rho}{\sigma}\left(V_T - V_0 - k\theta T + k\int_0^T V_s \mathrm{d}s\right) + \sqrt{1-\rho^2}\sqrt{\int_0^T V_s \mathrm{d}s}N\right],$$

in which N is defined to be the same standard normal random variable as that in the equation for $\tilde{S}_{T,h}$ or $\tilde{S}_{T,Mh}$. Due to the independence of N and V, it follows by the straight-forward calculation

and the Cauchy-Schwarz inequality that

$$\begin{aligned}
&\mathbb{E}\left(\ln \tilde{S}_{T,h} - \ln S_T\right)^2 \\
&= c\mathbb{E}\left(\fint_0^T V_s^h \mathrm{d}s - \int_0^T V_s \mathrm{d}s\right)^2 + c\mathbb{E}\left(\sqrt{\fint_0^T V_s^h \mathrm{d}s} - \sqrt{\int_0^T V_s \mathrm{d}s}\right)^2 \\
&= c\mathbb{E}\left(\fint_0^T V_s^h \mathrm{d}s - \int_0^T V_s \mathrm{d}s\right)^2 + c\mathbb{E}\left(\frac{\fint_0^T V_s^h \mathrm{d}s - \int_0^T V_s \mathrm{d}s}{\sqrt{\fint_0^T V_s^h \mathrm{d}s} + \sqrt{\int_0^T V_s \mathrm{d}s}}\right)^2 \\
&\leq c\mathbb{E}\left(\fint_0^T V_s^h \mathrm{d}s - \int_0^T V_s \mathrm{d}s\right)^2 + c\mathbb{E}^{1/2}\left(\fint_0^T V_s^h \mathrm{d}s - \int_0^T V_s \mathrm{d}s\right)^4 \mathbb{E}^{1/2}\left(\frac{1}{\int_0^T V_s \mathrm{d}s}\right)^2. \qquad (12)
\end{aligned}$$

By Lemma 3.1, we have

$$\mathbb{E}\left(\frac{1}{\int_0^T V_s \mathrm{d}s}\right)^2 < \infty.$$

Thus, it remains to analyse $\mathbb{E}\left(\fint_0^T V_s^h \mathrm{d}s - \int_0^T V_s \mathrm{d}s\right)^{2n}$ for any $n \in \mathbb{N}^+$, and we have

$$\mathbb{E}\left(\fint_0^T V_s^h \mathrm{d}s - \int_0^T V_s \mathrm{d}s\right)^{2n} \leq c\mathbb{E}\left(\int_0^T V_{\eta(s)} \mathrm{d}s - \int_0^T V_s \mathrm{d}s\right)^{2n} + c\mathbb{E}(V_T - V_0)^{2n} h^{2n}, \qquad (13)$$

where $\eta(s) = \max\{lh : lh \leq s, l = 0, 1, 2, ...\}$. The first term on the right side of (13) can be regarded as the Euler scheme applied on the integrated variance process. From the SDE of the variance process, we obtain

$$\begin{aligned}
&\mathbb{E}\left(\int_0^T V_{\eta(s)} \mathrm{d}s - \int_0^T V_s \mathrm{d}s\right)^{2n} \\
&\leq c\mathbb{E}\left(\int_0^T \left(\int_{\eta(s)}^s k(\theta - V_u)\mathrm{d}u\right) \mathrm{d}s\right)^{2n} + c\mathbb{E}\left(\int_0^T \left(\int_{\eta(s)}^s \sigma\sqrt{V_u}\mathrm{d}W_u^1\right) \mathrm{d}s\right)^{2n}. \qquad (14)
\end{aligned}$$

For the first term of (14), the Fubini theorem implies that

$$\int_0^T \left(\int_{\eta(s)}^s k(\theta - V_u)\mathrm{d}u\right) \mathrm{d}s = \int_0^T \left(\int_u^{\eta(u)+h} k(\theta - V_u)\mathrm{d}s\right) \mathrm{d}u = h\int_0^T \left(1 + \frac{\eta(u) - u}{h}\right) k(\theta - V_u)\mathrm{d}u.$$

Thus, the first term of (14) can be written as

$$\begin{aligned}
&\mathbb{E}\left(\int_0^T \left(\int_{\eta(s)}^s k(\theta - V_u)\mathrm{d}u\right) \mathrm{d}s\right)^{2n} \\
&= h^{2n}\mathbb{E}\left(\int_0^T \left(1 + \frac{\eta(u) - u}{h}\right) k(\theta - V_u)\mathrm{d}u\right)^{2n} \leq ch^{2n}\int_0^T \mathbb{E}\left(\theta - V_u\right)^{2n} \mathrm{d}u = O(h^{2n}),
\end{aligned}$$

where the last step is valid due to that $\sup_{t\in[0,T]}\mathbb{E}\left(V_t^p\right)<\infty$ for any $p>0$. For the second term of (14), it follows by the Fubini theorem and the Burkholder-Davies-Gundy inequality that

$$\begin{aligned}&\mathbb{E}\left(\int_0^T\left(\int_{\eta(s)}^s\sigma\sqrt{V_u}\mathrm{d}W_u^1\right)\mathrm{d}s\right)^{2n}\\&=h^{2n}\mathbb{E}\left(\int_0^T\left(1+\frac{\eta(u)-u}{h}\right)\sigma\sqrt{V_u}\mathrm{d}W_u^1\right)^{2n}\leq ch^{2n}\mathbb{E}\left(\int_0^T\sigma^2V_u\mathrm{d}u\right)^n=O(h^{2n}).\end{aligned}$$

Therefore, they, together with the inequalities (13) and (14), indicate that

$$\mathbb{E}\left(\oint_0^T V_s^h\mathrm{d}s-\int_0^T V_s\mathrm{d}s\right)^{2n}=O(h^{2n}).$$

With the aid of the inequality (12), it follows that

$$\mathbb{E}(\ln\tilde{S}_{T,h}-\ln S_T)^2=O(h^2).$$

The analysis of $\mathbb{E}(\ln\tilde{S}_{T,Mh}-\ln S_T)^2$ is analogous by replacing h with Mh. The application of the inequality (11) finishes the proof. □

Theorem 3.3. *Let $T>0$ and $M\in\mathbb{N}^+$. For any payoff function P that satisfies the assumptions in Theorem 3.1, We have*

$$\mathrm{Var}(P(\tilde{S}_{T,h})-P(\tilde{S}_{T,Mh}))=O(h^2),$$

for the full parameter regime of the Heston model.

Proof. It follows from Theorem 3.1 and Theorem 3.2 that

$$\mathbb{E}\left(P(\tilde{S}_{T,h})-P(\tilde{S}_{T,Mh})\right)^2\leq\mathbb{E}\left(\ln\tilde{S}_{T,Mh}-\ln\tilde{S}_{T,h}\right)^2=O(h^2).$$

The application of the inequality

$$\mathrm{Var}\left(P(\tilde{S}_{T,h})-P(\tilde{S}_{T,Mh})\right)\leq\mathbb{E}\left(P(\tilde{S}_{T,h})-P(\tilde{S}_{T,Mh})\right)^2$$

completes the proof. □

Remark 3.2. *If we consider the Euler scheme with the time integrated variance process approximated as*

$$\int_0^T V_s\mathrm{d}s\approx\sum_{i=1}^{T/h}hV_{(i-1)h},$$

then we can easily adapt the MLMC estimator we have analysed to the Euler case and prove $\mathrm{Var}(P(\tilde{S}_{T,h})-P(\tilde{S}_{T,Mh}))=O(h^2)$. Although there is no improvement on the convergence rate when we use the trapezoidal rule for the Heston model, the trapezoidal rule generally has a higher weak order than the Euler scheme, which is preferred in MLMC applications.

3.3 Simulation for path-dependent options

For path-dependent options, there is more flexibility to define MLMC estimators, and we provide two MLMC estimators.

Let $\hat{S}_t^h$, $t = 0, h, 2h, ..., T$, be the approximation of the asset price $(S_t)_{t\geq 0}$ at the fine level with step size h, and $\hat{S}_t^{Mh}$, $t = 0, Mh, 2Mh, ..., T$, be the approximation at the coarse level with step size Mh, where the initials $\hat{S}_0^h = \hat{S}_0^{Mh} = S_0$. At the fine level with time step h, we let

$$\ln \hat{S}_{H+ih}^h = \ln \hat{S}_{H+(i-1)h}^h + \left(r - \frac{\rho k\theta}{\sigma}\right) h + \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right) \frac{V_{H+(i-1)h} + V_{H+ih}}{2} h$$
$$+ \frac{\rho}{\sigma}(V_{H+ih} - V_{H+(i-1)h}) + \sqrt{1-\rho^2}\sqrt{\frac{V_{H+(i-1)h} + V_{H+ih}}{2} h} N_i,$$

for any $i = 1, ..., M$, where $H = jMh$, $j = 0, 1, ..., T/(Mh) - 1$ and N_i are independent standard normal random variables. On the other hand, at the coarse level with time step Mh, we take

$$\ln \hat{S}_{H+Mh}^{Mh} = \ln \hat{S}_H^{Mh} + \left(r - \frac{\rho k\theta}{\sigma}\right) Mh + \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right) \frac{V_H + V_{H+Mh}}{2} Mh$$
$$+ \frac{\rho}{\sigma}(V_{H+Mh} - V_H) + \sqrt{1-\rho^2}\sqrt{\frac{V_H + V_{H+Mh}}{2} Mh} N,$$

where N is normal distributed with mean 0 and standard deviation 1.

The key step in MLMC is to establish a link between $\ln \hat{S}_{H+Mh}^{Mh}$ at the coarse level and $\ln \hat{S}_{H+h}^h$,...,$\ln \hat{S}_{H+Mh}^h$ at the fine level. This requires us to properly define N, based on N_1, N_2,..., N_M, such that N is a standard normal random variable that is independent of the variance process. There is a simple estimator

$$N := \frac{1}{\sqrt{M}}(N_1 + N_2 + ... + N_M), \tag{15}$$

which can be found at Giles (2008b). However, for the Heston model, where the variance process can be simulated in isolation, we have more freedom to construct an estimator, and we define

$$N := \frac{\sum_{i=1}^M \sqrt{\int_{H+(i-1)h}^{H+ih} V_s^h \mathrm{d}s} N_i}{\sqrt{\int_H^{H+Mh} V_s^h \mathrm{d}s}}. \tag{16}$$

The estimator (16) above does not only relies on $N_1, ..., N_M$ but also depends on the trapezoidal rule with step size h. We shall call the former (15) the standard estimator and the latter (16) the weighted average estimator.

For any path of the variance process V given, N calculated by (16) is a standard normal random variable, due to that N_1,..., N_M are independent of V. Since for any path of V, N has the same distribution, it is clear that N is independent of V. Compared with the standard estimator, the weighted average estimator has a smaller $\mathrm{Var}(\ln \hat{S}_H^{Mh} - \ln \hat{S}_H^h)$, for the same $H = Mh, 2Mh, ..., T$, which is supported by our numerical test in Section 4. We emphasize that both estimators here can be used in principle for any path-dependent option, although an option with a different payoff function may lead to different computational complexity.

3.4 Convergence analysis for path-dependent simulation

Let $\eta(t) = \max\{lh : lh \le t, l = 0, 1, 2, ..\}$ and $\eta_M(t) = \max\{lMh : lMh \le t, l = 0, 1, 2, ...\}$, where $t \in [0, T]$. For MLMC analysis of a path-dependent option, we shall always assume the payoff function P satisfies Assumption 3.1 below:

Assumption 3.1. *It holds that*

$$|P(\hat{S}^h) - P(\hat{S}^{Mh})| \le c \sup_{t\in[0,T]} |\ln \hat{S}^h_{\eta(t)} - \ln \hat{S}^{Mh}_{\eta_M(t)}|,$$

for all paths of the weak approximations $\hat{S}^h$ and $\hat{S}^{Mh}$.

This assumption is roughly for options with bounded payoffs that are Lipschitz continuous. Below are some popular options in real practice that fall into the scope of our analysis.

Example 3.1: the Asian option

A typical path-dependent option that satisfies the Lipschitz assumption is the Asian put option with the payoff

$$P(S) := \max\left(K - \frac{1}{T}\int_0^T S_t \mathrm{d}t, 0\right)$$

where $K > 0$ is the strike price. It is Lipschitz continuous when we take $\int_0^T S_t \mathrm{d}t$ as a random variable. We can use the approximation $P(\hat{S}^h) := \max\left(K - \frac{1}{T}\sum_{i=1}^{T/h} \hat{S}^h_{(i-1)h} h, 0\right)$. Since $\sum_{j=1}^{T/(Mh)} \hat{S}^{Mh}_{(j-1)Mh} Mh = \sum_{i=1}^{T/h} \hat{S}^{Mh}_{\eta_M((i-1)h)} h$, it follows from Theorem 3.1 that

$$\begin{aligned}
|P(\hat{S}^h) - P(\hat{S}^{Mh})| &\le c \left| \ln\left(\frac{1}{T}\sum_{i=1}^{T/h} \hat{S}^h_{(i-1)h} h\right) - \ln\left(\frac{1}{T}\sum_{j=1}^{T/(Mh)} \hat{S}^{Mh}_{(j-1)Mh} Mh\right)\right| \\
&= c \left| \ln\left(\frac{1}{T}\sum_{i=1}^{T/h} \hat{S}^h_{\eta((i-1)h)} h\right) - \ln\left(\frac{1}{T}\sum_{i=1}^{T/h} \hat{S}^{Mh}_{\eta_M((i-1)h)} h\right)\right| \\
&\le c \sup_{t\in[0,T]} |\ln \hat{S}^h_{\eta(t)} - \ln \hat{S}^{Mh}_{\eta_M(t)}|.
\end{aligned}$$

The last inequality is the application of Proposition 3.1.

Proposition 3.1. *For any $a_1, b_1, a_2, b_2 > 0$, we have*

$$|\ln(a_1 + b_1) - \ln(a_2 + b_2)| \le \max\{|\ln a_1 - \ln a_2|, |\ln b_1 - \ln b_2|\}.$$

Proof. Without loss of generality, we assume that $a_1 + b_1 > a_2 + b_2$. Let $a_1 = c_a a_2$, $b_1 = c_b b_2$, where $c_a, c_b > 0$, and then we have $|\ln(a_1 + b_1) - \ln(a_2 + b_2)| = \ln\left(\frac{c_a a_2 + c_b b_2}{a_2 + b_2}\right)$. Suppose that $c_a \ge c_b$, then we obtain $c_a > 1$ and

$$|\ln(a_1 + b_1) - \ln(a_2 + b_2)| = \ln\left(\frac{c_a a_2 + c_b b_2}{a_2 + b_2}\right) \le \ln c_a = |\ln a_1 - \ln a_2|.$$

Similarly, in the case where $c_a < c_b$, we can get $c_b > 1$ and

$$|\ln(a_1 + b_1) - \ln(a_2 + b_2)| \leq |\ln b_1 - \ln b_2|.$$

The proof is complete. □

Example 3.2: the lookback option with fixed strike

Another path-dependent option that has a Lipschitz payoff is the lookback put option with fixed strike, and its payoff is

$$P(S) = \max\left(K - \inf_{t\in[0,T]} S_t, 0\right),$$

where $K > 0$. This payoff is again Lipschitz continuous with respect to the random variable $\inf_{t\in[0,T]} S_t$, which can be approximated as $\inf_{t\in[0,T]} S_t \approx \min_{j=0,1,\ldots,T/h} \hat{S}^h_{jh}$. Then by Theorem 3.1, we have

$$\begin{aligned}
|P(\hat{S}^h) - P(\hat{S}^{Mh})| &\leq c\left|\ln\left(\min_{J=0,h,2h,\ldots,T} \hat{S}^h_J\right) - \ln\left(\min_{H=0,Mh,2Mh,\ldots,T} \hat{S}^{Mh}_H\right)\right| \\
&= c\left|\min_{J=0,h,2h,\ldots,T} \ln\left(\hat{S}^h_J\right) - \min_{H=0,Mh,2Mh,\ldots,T} \ln\left(\hat{S}^{Mh}_H\right)\right| \\
&\leq c \sup_{t\in[0,T]} |\ln \hat{S}^h_{\eta(t)} - \ln \hat{S}^{Mh}_{\eta_M(t)}|.
\end{aligned}$$

There are some approaches tailored for some specific payoffs that might improve the simulation performance (e.g. for the Asian option, one can approximate $\int_0^T S_t \mathrm{d}t$ based on the trapezoidal rule; for the lookback option, one can simulate $\inf_{t\in[0,T]} S_t$ based on the method discussed in Chapt 6.4 of Glasserman (2003).). However, their analysis is usually more complicated. In our analysis, we shall focus on the weighted average estimator (16), because it is more efficient for MLMC as discussed. Our target is to derive the convergence rate of $\mathrm{Var}(P(\hat{S}^h) - P(\hat{S}^{Mh}))$ for the full parameter regime. We require Lemma 3.2 below from Corollary 2.14 in Hutzenthaler, Jentzen and Noll (2014).

Lemma 3.2. (Hutzenthaler, Jentzen and Noll). *Let $T > 0$. For any $s, t \in [0, T]$, $V_0 \geq 0$, $p > 0$, there exists a constant L (independent of s, t), such that*

$$\mathbb{E}\left|\sqrt{V_t} - \sqrt{V_s}\right|^p \leq L|t - s|^{p/2}$$

and for all $\varepsilon > 0$, it holds that

$$\mathbb{E} \sup_{s,t\in[0,T], s\neq t} \left(\frac{|\sqrt{V_t} - \sqrt{V_s}|}{|t - s|^{1/2-\varepsilon}}\right)^p < \infty.$$

Theorem 3.4. *Let $T > 0$ and $M \in \mathbb{N}^+$. For the weighted average estimator, we have*

$$\mathbb{E} \max_{H=Mh,2Mh,\ldots,T} \left(\ln \hat{S}^{Mh}_H - \ln \hat{S}^h_H\right)^2 = O(h).$$

Proof. Using the definition of the weighted average estimator (16) and then applying the Burkholder-Davis-Gundy inequality, we obtain

$$
\begin{aligned}
&\mathbb{E}\max_{H=Mh,2Mh,\ldots,T}\left[\left(\ln\hat{S}_H^{Mh}-\ln\hat{S}_H^{h}\right)^2\right]\\
&=\mathbb{E}\max_{H}\left[c\left(\fint_0^H V_s^{Mh}\mathrm{d}s-\fint_0^H V_s^{h}\mathrm{d}s\right)+c\sum_{j=1}^{H/(Mh)}\left(\sqrt{\fint_{(j-1)Mh}^{jMh}V_s^{Mh}\mathrm{d}s}-\sqrt{\fint_{(j-1)Mh}^{jMh}V_s^{h}\mathrm{d}s}\right)Z_j\right]^2\\
&\le c\mathbb{E}\max_{H}\left(\fint_0^H V_s^{Mh}\mathrm{d}s-\fint_0^H V_s^{h}\mathrm{d}s\right)^2+c\mathbb{E}\max_{H}\left[\sum_{j=1}^{H/(Mh)}\left(\sqrt{\fint_{(j-1)Mh}^{jMh}V_s^{Mh}\mathrm{d}s}-\sqrt{\fint_{(j-1)Mh}^{jMh}V_s^{h}\mathrm{d}s}\right)Z_j\right]^2\\
&\le c\mathbb{E}\max_{H}\left(\fint_0^H V_s^{Mh}\mathrm{d}s-\fint_0^H V_s^{h}\mathrm{d}s\right)^2+c\sum_{j=1}^{T/(Mh)}\mathbb{E}\left(\sqrt{\fint_{(j-1)Mh}^{jMh}V_s^{Mh}\mathrm{d}s}-\sqrt{\fint_{(j-1)Mh}^{jMh}V_s^{h}\mathrm{d}s}\right)^2
\end{aligned}
\tag{17}
$$

where Z_j is a standard normal random variable independent of the variable process. Let $\bar{V}_{s,h}=\frac{1}{2}(V_{\eta(s)}+V_{\eta(s)+h})$, if $s\in[0,T)$, and $\bar{V}_{T,h}=V_T$, where $\eta(s)=\max\{lh:lh\le s,l=0,1,2,..\}$. For the first term in (17), we obtain from Jensen's inequality that

$$
\mathbb{E}\max_{H}\left(\fint_0^H V_s^{h}\mathrm{d}s-\int_0^H V_s\mathrm{d}s\right)^2\le\int_0^T\mathbb{E}\left(\bar{V}_{s,h}-V_s\right)^2\mathrm{d}s=O(h),
$$

Hence, the first term is $O(h)$. Next we consider the second term in (17). Since

$$
\left(\sqrt{\int_0^T f(s)\mathrm{d}s}-\sqrt{\int_0^T g(s)\mathrm{d}s}\right)^2\le\int_0^T\left(\sqrt{f(s)}-\sqrt{g(s)}\right)^2\mathrm{d}s
$$

for any $f,g:[0,T]\to[0,+\infty)$, it follows that

$$
\begin{aligned}
&\sum_{j=1}^{T/(Mh)}\mathbb{E}\left(\sqrt{\fint_{(j-1)Mh}^{jMh}V_s^{Mh}\mathrm{d}s}-\sqrt{\fint_{(j-1)Mh}^{jMh}V_s^{h}\mathrm{d}s}\right)^2\\
&\le c\sum_{j=1}^{T/h}\mathbb{E}\left(\sqrt{\fint_{(j-1)h}^{jh}V_s^{h}\mathrm{d}s}-\sqrt{\int_{(j-1)h}^{jh}V_s\mathrm{d}s}\right)^2+c\sum_{j=1}^{T/(Mh)}\mathbb{E}\left(\sqrt{\fint_{(j-1)Mh}^{jMh}V_s^{Mh}\mathrm{d}s}-\sqrt{\int_{(j-1)Mh}^{jMh}V_s\mathrm{d}s}\right)^2.
\end{aligned}
$$

With the aid of Lemma 3.2, we further have

$$
\sum_{j=1}^{T/h}\mathbb{E}\left(\sqrt{\fint_{(j-1)h}^{jh}V_s^{h}\mathrm{d}s}-\sqrt{\int_{(j-1)h}^{jh}V_s\mathrm{d}s}\right)^2\le c\int_0^T\mathbb{E}\left(\sqrt{\bar{V}_{s,h}}-\sqrt{V_s}\right)^2\mathrm{d}s=O(h),
$$

which indicates that the second term in (17) is also $O(h)$. The proof is complete. □

Theorem 3.5. *Let $T > 0$ and $M \in \mathbb{N}^+$. Suppose that the payoff function P satisfies Assumption 3.1. Then for the weighted average estimator, we have*

$$\mathrm{Var}(P(\hat{S}^h) - P(\hat{S}^{Mh})) = O(h^{1-\varepsilon}),$$

for all $\varepsilon > 0$ and all parameter regimes of the Heston model.

Proof. For the payoff function that satisfies Assumption 3.1, we have

$$\mathbb{E}\left(P(\hat{S}^h) - P(\hat{S}^{Mh})\right)^2 \leq c\mathbb{E}\sup_{t\in[0,T]}\left(\ln \hat{S}^h_{\eta(t)} - \ln \hat{S}^{Mh}_{\eta_M(t)}\right)^2, \tag{18}$$

and it holds that

$$\begin{aligned}
&\mathbb{E}\sup_{t\in[0,T]}\left(\ln \hat{S}^h_{\eta(t)} - \ln \hat{S}^{Mh}_{\eta_M(t)}\right)^2 \\
&= \mathbb{E}\sup_{t\in[0,T]}\left[\left(\ln \hat{S}^h_{\eta_M(t)} - \ln \hat{S}^{Mh}_{\eta_M(t)}\right) + \left(\ln \hat{S}^h_{\eta(t)} - \ln \hat{S}^h_{\eta_M(t)}\right)\right]^2 \\
&\leq 2\mathbb{E}\max_{H=0,Mh,..,T}\left(\ln \hat{S}^{Mh}_H - \ln \hat{S}^h_H\right)^2 + 2\mathbb{E}\sup_{t\in[0,T]}\left(\ln \hat{S}^h_{\eta(t)} - \ln \hat{S}^h_{\eta_M(t)}\right)^2.
\end{aligned} \tag{19}$$

Theorem 3.4 implies that the first term $\mathbb{E}\max_{H=0,Mh,..,T}\left(\ln \hat{S}^{Mh}_H - \ln \hat{S}^h_H\right)^2 = O(h)$. Next let's pay attention to the second term from (19), and it follows that

$$\begin{aligned}
&\mathbb{E}\sup_{t\in[0,T]}\left(\ln \hat{S}^h_{\eta(t)} - \ln \hat{S}^h_{\eta_M(t)}\right)^2 \\
&\leq ch^2 + c\mathbb{E}\max_j(V_{jh} - V_{(j-1)h})^2 + c\mathbb{E}\max_j\left(\frac{V_{(j-1)h} + V_{jh}}{2}\right)^2 h^2 + c\mathbb{E}\max_j\left(\sqrt{\frac{V_{(j-1)h} + V_{jh}}{2}}hN_j\right)^2 \\
&\leq ch^2 + c\mathbb{E}\max_j(V_{jh} - V_{(j-1)h})^2 + c\mathbb{E}\max_i\left(V_{ih}^2\right)h^2 + c\mathbb{E}\max_i\left(V_{ih}\right)\cdot\mathbb{E}\max_j(N_j^2)h,
\end{aligned} \tag{20}$$

where $j = 1, ..., T/h$, $i = 0, 1, ..., T/h$, and N_j are iid standard normal random variables that are independent of the process $(V_t)_{t\in[0,T]}$. From Lemma 3.2, we obtain $\mathbb{E}\sup_{t\in[0,T]}(V_t^p) < \infty$ and $\mathbb{E}\sup_{t\in[0,T],|t-s|=h}\left|\sqrt{V_t} - \sqrt{V_s}\right|^p \leq ch^{p/2-\varepsilon}$ for any $p > 0$ and $\varepsilon > 0$. Thus,

$$\mathbb{E}\max_i\left(V_{ih}^p\right) < \infty, \tag{21}$$

for any $p > 0$, and the Cauchy-Schwartz inequality indicates

$$\mathbb{E}\max_j\left(V_{jh} - V_{(j-1)h}\right)^2 \leq \mathbb{E}^{1/2}\max_j\left(\sqrt{V_{jh}} - \sqrt{V_{(j-1)h}}\right)^4 \cdot \mathbb{E}^{1/2}\max_j\left(\sqrt{V_{jh}} + \sqrt{V_{(j-1)h}}\right)^4 \leq ch^{1-\varepsilon}, \tag{22}$$

for any $\varepsilon > 0$. Now, it remains to analyse $\mathbb{E}\max_j(N_j^2)h$. Let $(W_t)_{t\in[0,T]}$ be a Brownian motion, and it follows by the well-known fact on the Brownian motion that

$$\mathbb{E}\max_j(N_j^2)h = \mathbb{E}\max_j(W_{jh} - W_{(j-1)h})^2 \leq \mathbb{E}\sup_{0\leq s\leq t\leq T,|t-s|\leq h}(W_t - W_s)^2 = O(|\ln(h)|h),$$

where the upper bound of the estimate can be found at Fischer and Nappo (2009). Hence, we get

$$\mathbb{E}\max_j(N_j^2)h = O(h^{1-\varepsilon}) \tag{23}$$

for any $\varepsilon > 0$. Therefore, combining (21), (22) and (23) into (20), we have

$$\mathbb{E}\sup_{t\in[0,T]}\left(\ln \hat{S}^h_{\eta(t)} - \ln \hat{S}^h_{\eta_M(t)}\right)^2 = O(h^{1-\varepsilon}). \tag{24}$$

It follows from (18), (19) and (24) that

$$\mathrm{Var}(P(\hat{S}^h) - P(\hat{S}^{Mh})) \le \mathbb{E}\left(P(\hat{S}^h) - P(\hat{S}^{Mh})\right)^2 = O(h^{1-\varepsilon}).$$

□

Remark 3.3. *Analogous to the path-independent simulation, if we apply the Euler scheme instead of the trapezoidal rule on the integrated variance process, we still have* $\mathrm{Var}(P(\hat{S}^h) - P(\hat{S}^{Mh})) = O(h^{1-\varepsilon})$ *for the weighted average estimator. This result fills a gap in Altmayer and Neuenkirch (2015), when the zero boundary of the variance process is attainable.*

Let us emphasize that the options that fall into the scope of our analysis are typically those options with bounded Lipschitz payoff functions, which satisfy the conditions in Theorem 3.1. In the case of path-dependent options, for example the Asian option, it is Lipschitz continuous when we take the whole term $\int_0^T S_t \mathrm{d}t$ as a random variable. For call options with Lipschitz payoffs, the condition $P(U) = P(U_{max})$ for all $U \ge U_{max}$ is not satisfied. However, our numerical experiments in Section 4 indicate they may still have desirable convergence rates. For options with non-Lipschitz payoffs (e.g., the barrier option, the digital option), it is possible that the convergence rate could be lower, but the analysis is not conducted in this research. We refer readers to Giles, Higham and Mao (2009) for numerical analysis based on an Euler scheme applied to SDEs with Lipschitz continuous coefficients.

3.5 Weak convergent analysis

Let us have a short discussion on the weak convergence rate of the stochastic trapezoidal approximation for the Heston model. We have

$$\left|\mathbb{E}(\hat{P}_l - \hat{P}_{l-1})\right| \le \mathbb{E}\left|\hat{P}_l - \hat{P}_{l-1}\right| \le \mathbb{E}^{1/2}\left(\hat{P}_l - \hat{P}_{l-1}\right)^2 = O(h_l^{\alpha/2}),$$

where $h_l = T/M^l$ and α is the convergence rate of the MLMC variance between two coupled paths, i.e., $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1}) = O(h_l^\alpha)$. It holds that

$$\left|\mathbb{E}(\hat{P}_l - P)\right| \le \sum_{i=1}^{n}\left|\mathbb{E}(\hat{P}_{l+i} - \hat{P}_{l+i-1})\right| + \left|\mathbb{E}(\hat{P}_{l+n} - P)\right|,$$

for all $n \in \mathbb{N}^+$. Suppose that $\lim_{l\to\infty}\mathbb{E}(\hat{P}_l) = \mathbb{E}(P)$, by taking the limit as $n \to \infty$, we obtain

$$\left|\mathbb{E}(\hat{P}_l - P)\right| = O(h_l^{\alpha/2}).$$

For the application of MLMC on the Heston model, we have shown that $\alpha = 2$ for path-independent options, and $\alpha = 1$ for path-dependent options. Hence, in either case, we could expect the weak convergence rate $\beta \geq 1/2$, satisfying the assumptions of Theorem 2.1. Now, the problem is reduced to investigating whether $\lim_{l\to\infty} \mathbb{E}(\hat{P}_l) = \mathbb{E}(P)$ and Theorem 3.6 provides a positive answer.

Theorem 3.6. *For the path-independent option, suppose that the payoff $P : [0, +\infty) \to [0, +\infty)$ satisfies the assumptions in Theorem 3.1, we have*

$$\lim_{h\to 0} \mathbb{E}(P(\tilde{S}_{T,h})) = \mathbb{E}(P(S_T));$$

For the path-dependent option, assuming that its payoff satisfies

$$|P(\hat{S}^h) - P(S)| \leq c \sup_{t\in[0,T]} |\ln \hat{S}^h_{\eta(t)} - \ln S_t|,$$

for all paths of S and its weak approximation $\hat{S}^h$, where $\eta(t) = \max\{lh : lh \leq t, l = 0, 1, 2, ..\}$, we obtain

$$\lim_{h\to 0} \mathbb{E}(P(\hat{S}^h)) = \mathbb{E}(P(S)).$$

Proof. For the path-independent simulation, we have demonstrated in the proof of Theorem 3.2 that $\lim_{h\to 0} \mathbb{E}\left(\ln \tilde{S}_{T,h} - \ln S_T\right)^2 = 0$, where the equation for S_T and that for $\tilde{S}_{T,h}$ have the same normal random variable N. Then it is straightforward to show that $\lim_{h\to 0} \mathbb{E}(P(\tilde{S}_{T,h})) = \mathbb{E}(P(S_T))$ as $|P(\tilde{S}_{T,h}) - P(S_T)| \leq c|\ln \tilde{S}_{T,h} - \ln S_T|$ by Theorem 3.1.

For the path-dependent simulation, we can rewrite the approximated path $\hat{S}^h_{ih}, i = 1, 2, ..., T/h$ as

$$\ln \hat{S}^h_{ih} = \ln S_0 + c_0 ih + c_1(V_{ih} - V_0) + c_2 \int_0^{ih} \bar{V}_{s,h} \mathrm{d}s + c_3 \int_0^{ih} \sqrt{\bar{V}_{s,h}} \mathrm{d}W_s,$$

and the exact path $(S_t)_{t\in[0,T]}$ as

$$\ln S_t = \ln S_0 + c_0 t + c_1(V_t - V_0) + c_2 \int_0^t V_s \mathrm{d}s + c_3 \int_0^t \sqrt{V_s} \mathrm{d}W_s,$$

where $\bar{V}_{s,h} = \frac{1}{2}(V_{\eta(s)} + V_{\eta(s)+h})$, if $s \in [0, T)$, and $\bar{V}_{T,h} = V_T$. Here, $c_j, j = 0, 1, 2, 3$, are some constant coefficients. To simplify the weak analysis, we let both equations keep the same path of the Brownian motion $(W_s)_{s\in[0,T]}$. Then we have

$$\begin{aligned}
\mathbb{E}(P(\hat{S}^h) - P(S))^2 &\leq \mathbb{E} \sup_{t\in[0,T]} \left(\ln \hat{S}^h_{\eta(t)} - \ln S_t\right)^2 \\
&\leq c\mathbb{E} \max_{i=1,2,...,T/h} \left(\ln \hat{S}^h_{ih} - \ln S_{ih}\right)^2 + c\mathbb{E} \sup_{t\in[0,T]} \left(\ln S_t - \ln S_{\eta(t)}\right)^2. \qquad (25)
\end{aligned}$$

It yields from the Burkholder-Davis-Gundy inequality and Lemma 3.2 that

$$\begin{aligned}
&\mathbb{E} \max_{i=1,2,...,T/h} \left(\ln \hat{S}^h_{ih} - \ln S_{ih}\right)^2 \\
&\leq c\mathbb{E} \max_{i=1,2,...,T/h} \left(\int_0^{ih} \left(\bar{V}_{s,h} - V_s\right) \mathrm{d}s\right)^2 + c\mathbb{E} \max_{i=1,2,...,T/h} \left(\int_0^{ih} \left(\sqrt{\bar{V}_{s,h}} - \sqrt{V_s}\right) \mathrm{d}W_s\right)^2 \\
&\leq c \int_0^T \mathbb{E}\left(\bar{V}_{s,h} - V_s\right)^2 \mathrm{d}s + c \int_0^T \mathbb{E}\left(\sqrt{\bar{V}_{s,h}} - \sqrt{V_s}\right)^2 \mathrm{d}s.
\end{aligned}$$

Due to that $\sup_{s\in[0,T]}\mathbb{E}(\bar{V}_{s,h}-V_s)^2$ and $\sup_{s\in[0,T]}\mathbb{E}(\sqrt{\bar{V}_{s,h}}-\sqrt{V_s})^2$ converge to 0 as $h\to 0$, we obtain

$$\lim_{h\to 0}\mathbb{E}\max_{i=1,2,\ldots,T/h}\left(\ln\hat{S}^h_{ih}-\ln S_{ih}\right)^2=0. \tag{26}$$

On the other hand, we find that for any $t\in[0,T]$, $\eta(t)$ converges to t as h approach zero. Thanks to the almost sure continuity of the process $(\ln S_t)_{t\in[0,T]}$, we have

$$\lim_{h\to 0}\sup_{t\in[0,T]}|\ln S_t-\ln S_{\eta(t)}|=0$$

almost surely. As $\sup_{t\in[0,T]}(\ln S_t-\ln S_{\eta(t)})^2\le 2\sup_{t\in[0,T]}(\ln S_t)^2$, where $\mathbb{E}\sup_{t\in[0,T]}(\ln S_t)^2<\infty$, the dominated convergence theorem implies that

$$\lim_{h\to 0}\mathbb{E}\sup_{t\in[0,T]}\left(\ln S_t-\ln S_{\eta(t)}\right)^2=0. \tag{27}$$

Finally, with (26) and (27) substituted into (25), together with the fact that convergence in the L_2 norm implies convergence in distribution, we prove the argument that $\lim_{h\to 0}\mathbb{E}(P(\hat{S}^h))=\mathbb{E}(P(S))$. □

Although the proof of Theorem 3.6 may not give a sharp upper bound for the weak error, it suffices for applications of MLMC. The problem of exact weak convergence rates for the Heston model is an open question, and literature on such a topic is quite limited. For some sufficiently smooth payoffs of the logarithmic Heston solution, Altmayer and Neuenkirch (2017) proved that a numerical method, that uses the drift-implicit Milstein scheme for the variance process and an Euler discretization for the log-Heston price, converges weakly with order 1; Zheng (2017) proved that the stochastic trapezoidal approximation converges weakly with order 2. For more complicated payoffs, such as those of path-dependent options, it is possible that their weak convergence rates could be lower.

3.6 Almost exact methods for the variance process

We have established the MLMC estimators for the Heston model, conditioned on that the variance process is simulated exactly. In the literature, the exact simulation method for such a process is generally time-consuming in comparison to sampling a normal random variable. Although the acceptance-rejection method by Marsaglia and Tsang (2000) can be relatively fast, it is generally inconvenient for financial applications, as the number of uniform random variables required to sample a chi-squared random variable depends on the model parameters, which makes it difficult for sensitivity analysis.

One approach for this problem is to use some almost exact method to replace the exact method, such as the NCI method by Van Haastrecht and Pelsser (2010). The 'almost exact' here means there is a good control of the approximation error to achieve any desired accuracy. The algorithm of the NCI method can be found in the Appendix, in which the main procedure is to employ the Monotone cubic Hermite spline interpolation technique to almost exactly sample from the chi-squared distribution.

It is important to mention that the QE scheme by Andersen (2008) is biased for simulation of the variance process, and thus it may generally not be suitable for our MLMC estimators. For a biased scheme, the approximated payoff $\hat{P}_l$ in $\mathbb{E}(\hat{P}_l-\hat{P}_{l-1})$ and in $\mathbb{E}(\hat{P}_{l+1}-\hat{P}_l)$ may not have the same expectation $\mathbb{E}(\hat{P}_l)$, where l is the level of MLMC.

	Case I	Case II	Case III	Case IV
k	1	0.5	0.3	6.2
θ	0.09	0.04	0.04	0.02
σ	1	1	0.9	0.6
ρ	-0.3	-0.9	-0.5	-0.7

Table 1: Parameters of the Heston model for standard European call options. In all cases, $T = 1$, $V_0 = \theta$, $r = 0$, and $S_0 = 100$.

4 Numerical results

In this section, we will carry out four numerical experiments. In the first, we evaluate the convergence rate of $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ as the level l increases. In the second, we compare the MLMC method with the standard Monte Carlo for the Heston model to test whether the computational savings are significant. In the third, we extend the result to exotic options, such as the Asian option and the lookback option, which have relatively complicated payoff functions. Finally, we combine the almost exact technique, the NCI method, with MLMC and compare it with some other numerical methods in the literature on the Heston model.

All the experiments here are conducted in MATLAB, and we use the function 'ncx2rnd' to exactly simulate the non-central chi-squared random variable. Since the exact price of the standard European call option can be calculated with very high accuracy, it is a good benchmark for our test, and we will examine the numerical performance of the path-independent estimator and path-dependent estimators by approximating the option price. In particular, when we approximate the option price based on the path-dependent estimator, we simulate the asset price pathwise. Four sets of parameters are considered, as shown in Table 1, all of which are interesting in practice. Specifically, Case I is for equity options, Case II is for long-dated FX options, and Case III is for long-term interest rate options, as mentioned in Andersen (2008). The last Case IV comes from Glasserman and Kim (2011), representing S&P 500 index options. Here, a unit of T is one year, and for computational convenience, we let $T = 1$ for all. The strike of the option is set to be equal to the initial value of the asset S_0, hence the option is called at-the-money. Recall that when $2k\theta \geq \sigma^2$, the boundary zero of the variance process is unattainable, whereas when $2k\theta < \sigma^2$, the origin is attainable and reflecting. The simulation concerning the latter is more challenging. As it is easy to verify, all of the cases in Table 1 fall into the latter category.

4.1 Numerical convergence rate

In the first experiment, we plot the MLMC variance

$$\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$$

to examine how rapidly it converges to zero as the level l grows. Here, $\hat{P}_l$ is the price of the standard European call option approximated by the stochastic trapezoidal discretization with step size M^{-l}, corresponding to the level l, and we set $M = 4$. Recall that our theoretical analysis is roughly for options with bounded Lipschitz payoffs. The theoretical convergence rate of the MLMC variance is 2 for path-independent simulation and 1 for path-dependent simulation, for all ranges of the model parameters. However, the analysis of the standard European call option is beyond our theoretical

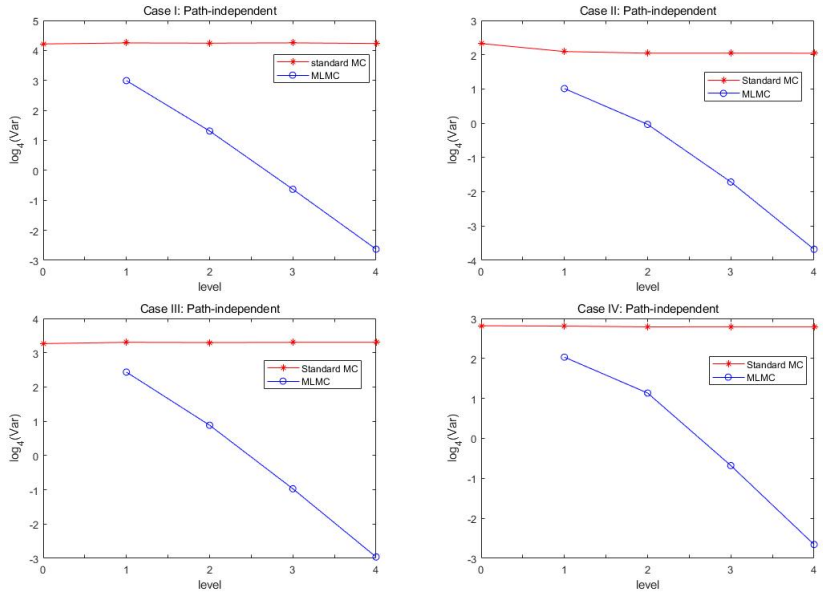

Figure 1: Comparison of $\mathrm{Var}(\hat{P}_l)$ and $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ for path-independent cases I-IV, M = 4. Red: standard Monte Carlo. Blue: multilevel Monte Carlo.

scope, and we will investigate its numerical convergence rate. The number of samples we simulate at each level is 1 million, hence the effect of the standard deviation of the estimator for $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ is negligible.

The result of the path-independent simulation is shown in Figure 1, which plots the logarithm of $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ (the blue curve) with base M versus different levels l. For comparison, we also plot the logarithm of $\mathrm{Var}(\hat{P}_l)$ (the red curve). As can be observed in all of the cases, $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ converges at rate 2, which is consistent with the theoretical rate. We find that when $l = 4$, $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ is generally less than 4^{-6} of $\mathrm{Var}(\hat{P}_l)$.

For the path-dependent simulation, we have two MLMC estimators for comparison: the standard estimator (15) and the weighted average estimator (16), with the former plotted in black and the latter plotted in blue, as displayed in Figure 2. They differ in the construction of the standard normal random variable at the coarse level using those standard normal random variables simulated at the fine level. Similar to the treatment of path-independent simulation, we illustrate the logarithm of $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ and $\mathrm{Var}(\hat{P}_l)$ with base M at the different level l. The numerical result shows that the MLMC variance converges at rate 1, consistent again with the theoretical convergence rate. In addition, the weighted average estimator is more efficient than the standard estimator, because the MLMC variance $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ of the former is significantly smaller at the same level, despite that there seems no improvement over the convergence rate.

4.2 MLMC versus standard Monte Carlo

In this experiment, we examine the efficiency of MLMC on the stochastic trapezoidal discretization for the Heston model, compared with the underlying scheme without applying MLMC.

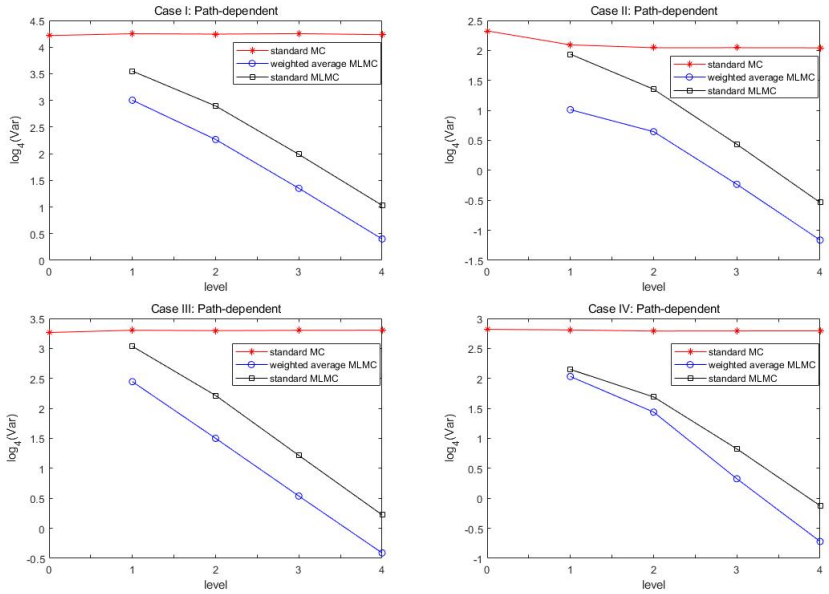

Figure 2: Comparison of $\mathrm{Var}(\hat{P}_l)$ and $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ for path-dependent cases I-IV, M = 4. Red: standard Monte Carlo. Blue: weighted average MLMC. Black: standard MLMC.

As discussed, the accuracy of the standard Monte Carlo is determined by the step size and the number of path simulations, whereas for MLMC, there is only one user-specified parameter ξ, the square of which is approximately the upper bound for the mean squared error. Furthermore, recall that in MLMC, we address path-independent simulation and path-dependent simulation separately using different MLMC estimators, so their accuracy and computational complexity are expected to be different. We use the same definition of the computational cost for MLMC as was employed in Giles (2008b), and the computational cost is

$$C = N_0 + \sum_{l=1}^{L} N_l (M^l + M^{l-1}),$$

where N_l is the number of sample paths specified by ξ and l, calculated through the MLMC algorithm in section 2.3. Here, L is the finest level, which is also from the MLMC algorithm. For standard Monte Carlo, we calculate it as

$$C^* = \sum_{l=0}^{L} N_l^* M^l,$$

where $N_l^* = 2\xi^{-2}\mathrm{Var}(\hat{P}_l)$, so the variance of the estimator to the option price is $\frac{1}{2}\xi^2$ as with MLMC.

In the numerical test, we let $M = 4$, and the weak convergence rate we use in the MLMC algorithm to control the bias is 2, which is consistent with the usual theory of the trapezoidal discretization. We compare the computational complexity of MLMC with that of standard Monte Carlo. The result is illustrated in Figure 3, which plots $\xi^2 C$ and $\xi^2 C^*$ versus ξ. For the path-dependent simulation, we only consider the weighted average estimator because it is more efficient

than the standard estimator, as discussed above. From Figure 3, we find that the computational savings from MLMC are significant in all cases, both for path-independent simulation and for path-dependent simulation. Specifically, the ratio ranges from 2.4 to 12.1 for path-independent simulation and from 2.2 to 6.7 for path-dependent simulation. As expected, the computational savings in path-independent simulation are generally higher than those in path-dependent simulation for the same ξ. For instance, when $\xi = 0.005$ in Case I, the computational saving is 12.1 for path-independent simulation, while it is 6.7 for path-dependent simulation. This is mainly because the MLMC estimator for path-independent simulation has a higher convergence rate, and thus the MLMC variance $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ is smaller when the level l is high.

Next, we analyse the computational complexity from a more theoretical perspective. Duffie and Glynn (1995) have shown that the mean squared error of standard Monte Carlo has an asymptotic form $\mathrm{MSE} = O(n^{-1}) + O(h^{2\beta})$, where n is the number of samples, β is the weak order, h is the step size of the underlying numerical scheme, and the optimal trade-off between n and h is $n = O(h^{-2\beta})$. Therefore, for a numerical scheme with $\beta = 2$, to obtain an MSE $O(\xi^2)$ for standard Monte Carlo, we need to set $n = O(\xi^{-2})$ and $h = O(\xi^{-1/2})$, and the overall cost would be $nh^{-1} = O(\xi^{-5/2})$. As can be observed, the red markers show that the Monte Carlo computational complexity is approximately $O(\xi^{-5/2})$, so that $\xi^2 C^* = O(\xi^{-1/2})$. In terms of the computational complexity for MLMC, the theory by Giles states that the computational cost is $O(\xi^{-2})$ for an MLMC estimator with $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1}) = O(h^2)$, and it is $O(\xi^{-2}(\ln \xi)^2)$ for that with $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1}) = O(h)$. Figure 3 illustrates that the numerical computational cost for path-independent simulation is slightly higher than $O(\xi^{-2})$, and for path-dependent simulation, it is roughly $O(\xi^{-2}(\ln \xi)^2)$. This means that the computational cost generally consistent with what the theory predicts.

4.3 Result for exotic options

In this part, we analyse the Asian option and the lookback option. The payoff of an Asian call option has the form $\left(\frac{1}{T}\int_0^T S_t \mathrm{d}t - K\right)^+$, in which the integration can usually be approximated as $\int_0^T S_t \mathrm{d}t \approx \sum_{i=1}^{T/\Delta} \frac{\hat{S}_{(i-1)\Delta} + \hat{S}_{i\Delta}}{2}\Delta$ based on the trapezoidal rule with step size Δ. The payoff of a lookback put option is $\left(K - \inf_{t\in[0,T]} S_t\right)^+$, where we approximate $\inf_{t\in[0,T]} S_t \approx \min_{i=0,1,\ldots,T/\Delta} \hat{S}_{i\Delta}$. For MLMC at the coarse level, we let $\Delta = Mh$, and at the fine level, we let $\Delta = h$.

The parameters of the Heston model for the Asian option are from Smith (2007), and we set $T = 1$ for convenience. The parameters for the lookback option are the same as those of Case I from Table 1. We investigate how rapidly $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ converges, as we did for the standard European option. The result is illustrated in Figure 4, which compares the relevant MLMC variances at different levels. As can be observed from the plots, for both types of options, the MLMC variances decay at the rate $O(h)$, and the weighted average estimator is significantly better than the standard estimator, which is again consistent with our theoretical analysis.

4.4 Comparison with existing numerical schemes

In the last experiment, we combine MLMC with the NCI method by Van Haastrecht and Pelsser (2010), which we call NCI-MLMC for convenience, and test its efficiency by comparing it with some existing methods in the literature on the Heston model. Specifically, in the path-independent simulation, we compare it with the gamma expansion method developed by Glasserman and Kim (2011), while in the path-dependent simulation, we compare it with the QE scheme developed by

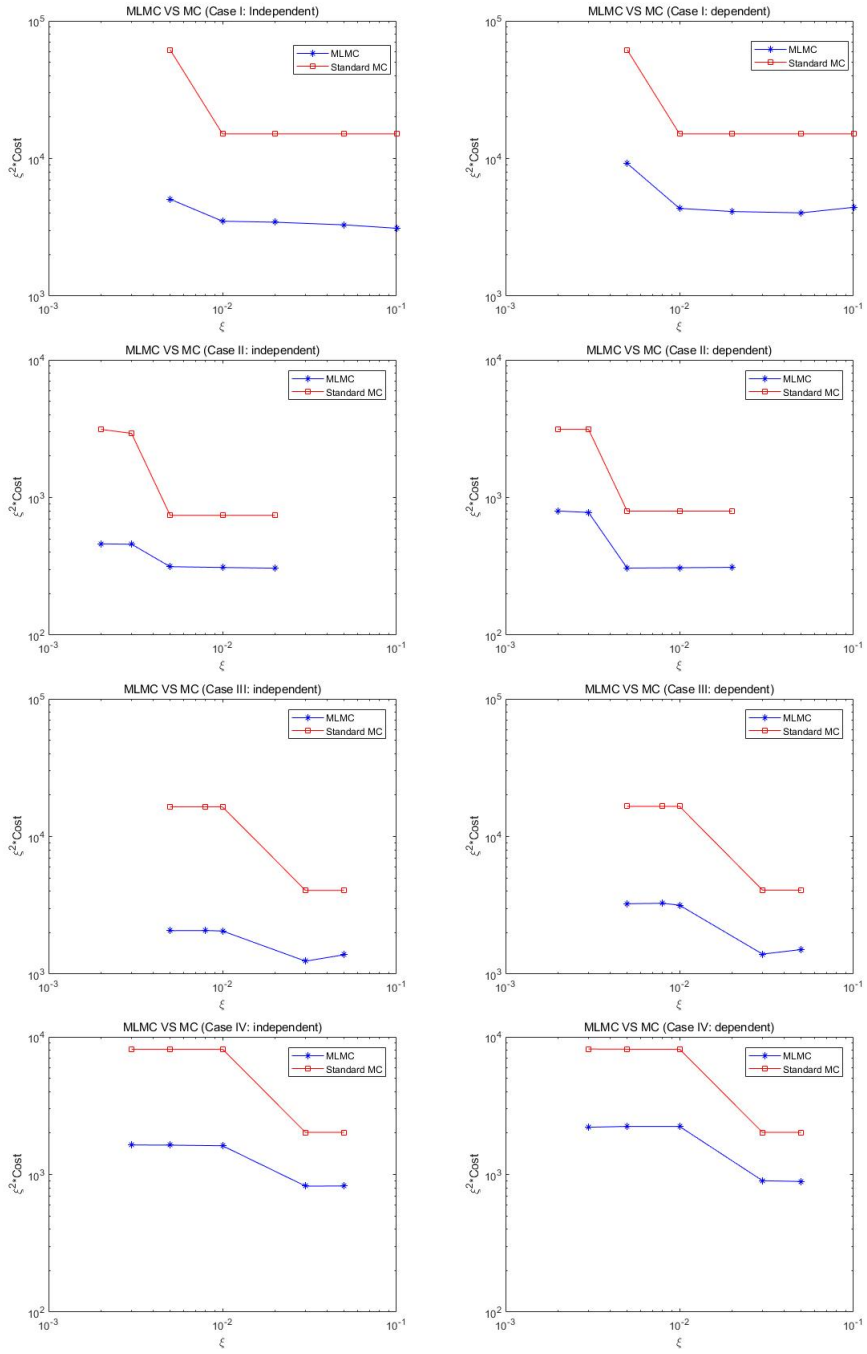

Figure 3: Comparison of the computational complexity between MLMC and the standard MC. Left: path-independent simulation. Right: path-dependent simulation.

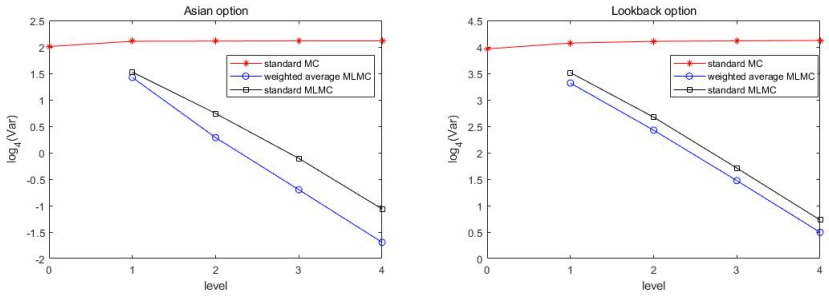

Figure 4: Convergence rate of $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ for exotic options. Left: the Asian call option ($k = 1.0407$, $\sigma = 0.5196$, $\theta = 0.0586$, $\rho = -0.6747$, $V_0 = 0.0194$, $r = 0$, $S_0 = K = 100$ and $T = 1$). Right: the lookback put option (the same parameters as in Case I from Table 1).

Andersen (2008). These two methods are among the most efficient methods in practice, so they are good benchmarks for our numerical test.

Before carrying out our numerical test, we will be careful about the 'bias' of the NCI method for the variance process. To obtain very high accuracy with the NCI method, we set the maximum Poisson number $N_{max} = 20$ and the grid size for interpolation $n = 10000$ in the NCI algorithm, which can be found in the Appendix. In addition, we adjust the weak convergence rate β in the MLMC algorithm to 1.8, which makes the MLMC algorithm work more efficiently. The experiment is conducted in MATLAB R2017b within a laptop of Intel Core i5 7200U CPU 2.5 GHz and 4 GB RAM.

First we compare NCI-MLMC with the gamma expansion method. The left panel of Figure 5 plots the relative computational time against the mean squared error, and the right shows their corresponding finest MLMC level L, where L means that the minimum step size of the stochastic trapezoidal discretization in MLMC is M^{-L} with $M = 4$. The relative computational time here is the ratio of the real computational time[1] to the time required to simulate 1 million samples from the Heston model by the QE method with step size $\frac{1}{8}$, for the same parameters of the Heston model. Glasserman and Kim [22] compared the gamma expansion method with the QE scheme. We use the result of the gamma expansion method with truncation level $K = 10$ from Table 7 of their paper and scale the computational time to that of the QE scheme also reported in Table 7. On the other hand, we compare NCI-MLMC with the QE scheme in our device, and record the relative computational time of NCI-MLMC. It is not uncommon to compare scaled results in computational finance (see, e.g., Lord, Koekkoek and Van Dijk (2010)). As shown in from Figure 5, NCI-MLMC is generally more efficient than the gamma expansion method. Specifically, both of the methods show the same rate of convergence for the computational complexity, and it is $O(\epsilon^{-2})$ as expected, where ϵ is the root mean square error[2]. However, the mean squared error of NCI-MLMC is smaller with the same computational time by a factor of 1.3-2.5. For example, in Case I, NCI-MLMC can be 2.5 times more efficient. Furthermore, we observe from the right panel of Figure 5 that the MLMC level L for the corresponding points is either 2 or 3, where $L = 2$ and $L = 3$ imply that the step

[1]The computational time required for tabulation in the NCI algorithm is excluded because this only has to be done once before running the Monte Carlo, which is also the case for the Gamma expansion method.

[2]The meaning of ϵ^2 here is different from that of ξ^2 used previously. The latter ξ^2 is an input in the MLMC algorithm, which is roughly the upper bound for the mean squared error.

size of the finest level in NCI-MLMC is $\frac{1}{16}$ and $\frac{1}{64}$ respectively, and when $L = 3$, the mean squared error is dominated by the variance since the bias is almost 0.

Then we compare NCI-MLMC with the QE scheme by calculating the standard European option price through the pathwise approximation, i.e., we simulate the asset price S_t at intermediate time $t = h, 2h, ..., T$. To illustrate the speed-accuracy tradeoff for the QE scheme, we plot the QE scheme with step sizes $\frac{1}{8}$ and $\frac{1}{32}$. From Figure 6, we see that for the QE scheme, the mean squared error is dominated by the bias as the number of simulations becomes large. This phenomenon is more obvious for the QE scheme with step size $\frac{1}{8}$, as its bias in sampling from the Heston model is relatively larger. The NCI-MLMC generally outperforms the QE scheme, particularly when the number of simulations is large, although with a large step size and a small number of stimulations, the QE scheme can be more efficient. In particular, with the same computational time, NCI-MLMC has a smaller mean squared error than the QE scheme with $h = \frac{1}{32}$ by a factor up to 6.3. At the point around $log_{10}(time) = 1$ in Case I, the mean squared error from the QE scheme with $h = \frac{1}{8}$ is approximately 7.9 times as large as that from NCI-MLMC, and this error from the QE scheme with $h = \frac{1}{32}$ is 2.0 times larger than its counterpart from NCI-MLMC. The computational complexity is approximately $O(\epsilon^{-2}(\ln \epsilon)^2)$ for NCI-MLMC, which coincides with the theory of MLMC. For the QE scheme, the result of Duffie and Glynn [12] implies its complexity is more than $O(\epsilon^{-2}(\ln \epsilon)^2)$ even if we increase the number of simulations asymptotically as h decreases, as discussed in section 4.2.

To include, we have demonstrated the efficiency of NCI-MLMC to compute the price of a standard European option. Hence, it is potentially useful to price options with more complicated payoffs. Moreover, as the MLMC estimators we develop can be combined with other exact or almost exact methods to simulate the variance process, some further development can be expected. However, the analysis of our MLMC estimators relies on the Lipschitz continuity of the payoff function. It is not clear to us whether our method remains efficient when the option payoff is non-Lipschitz continuous. The gamma expansion method has an advantage that it is exact for path-independent options. The QE scheme might be applicable for more general option payoffs.

5 Extensions

The multilevel Monte Carlo estimators we have defined for the Heston model can be extended for more Heston-type models, such as the Heston model with jump diffusion, the Heston model with CEV process and the Heston model with stochastic interest rate. These models generalize the Heston model, and have more flexibility to fit the implied volatility surface spotted in the financial market.

Heston Model with jump diffusion

The Heston model with jump diffusion, usually called the SVJ model, extends the Heston model by adding a jump diffusion. It was first used by Bates (1996) to address options of deutsche marks. The SDE of the SVJ model under the risk-neutral measure is as follows

$$\begin{aligned} \mathrm{d}S_t &= (r - \lambda\bar{\mu})S_t\mathrm{d}t + \sqrt{V_t}S_t(\rho\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2) + (e^{J_{N(t)}} - 1)S_{t-}\mathrm{d}N(t) \\ \mathrm{d}V_t &= k(\theta - V_t)\mathrm{d}t + \sigma\sqrt{V_t}\mathrm{d}W_t^1, \end{aligned}$$

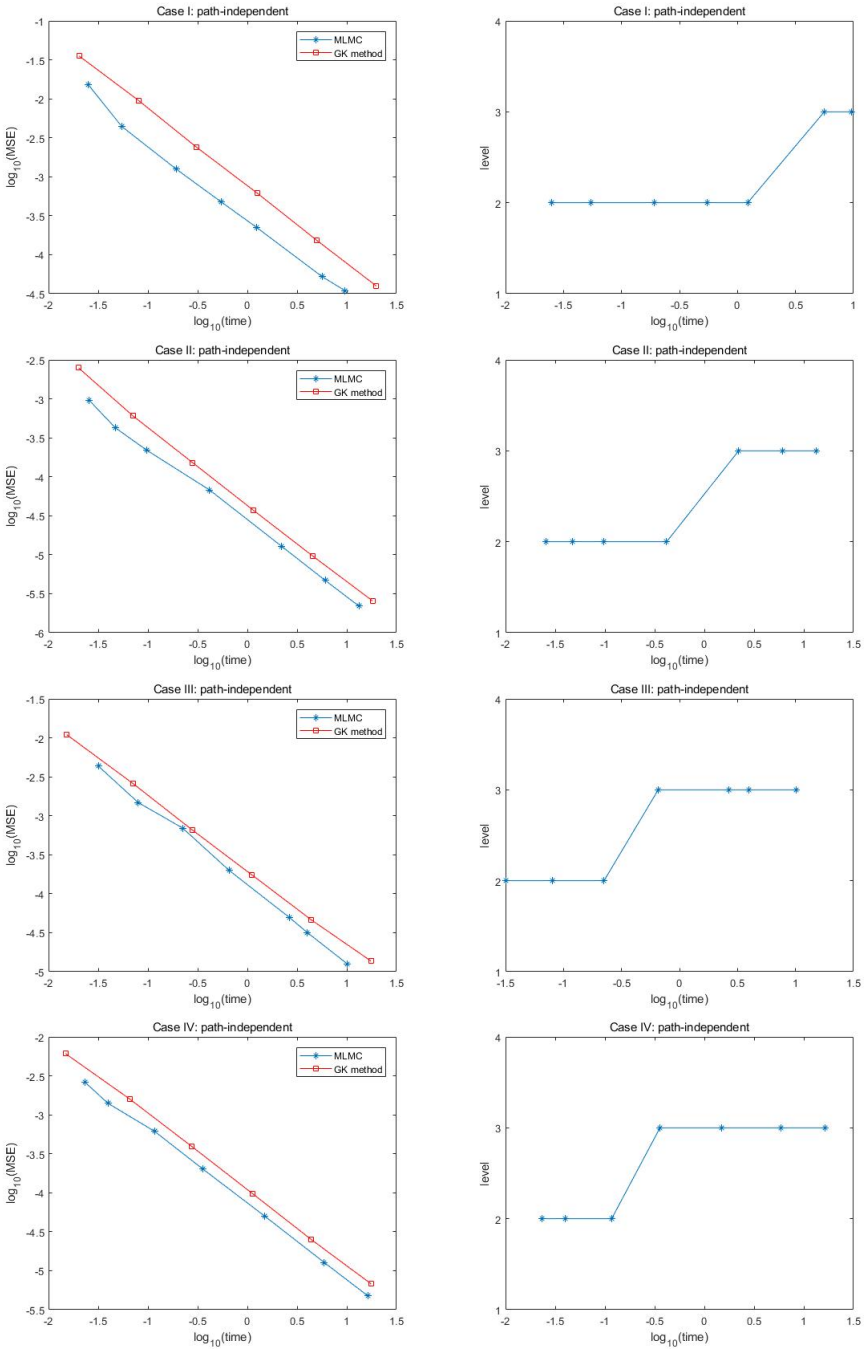

Figure 5: Comparison of NCI-MLMC and the gamma expansion method for path-independent simulation. Left: computational time versus MSE. Right: the finest levels for MLMC.

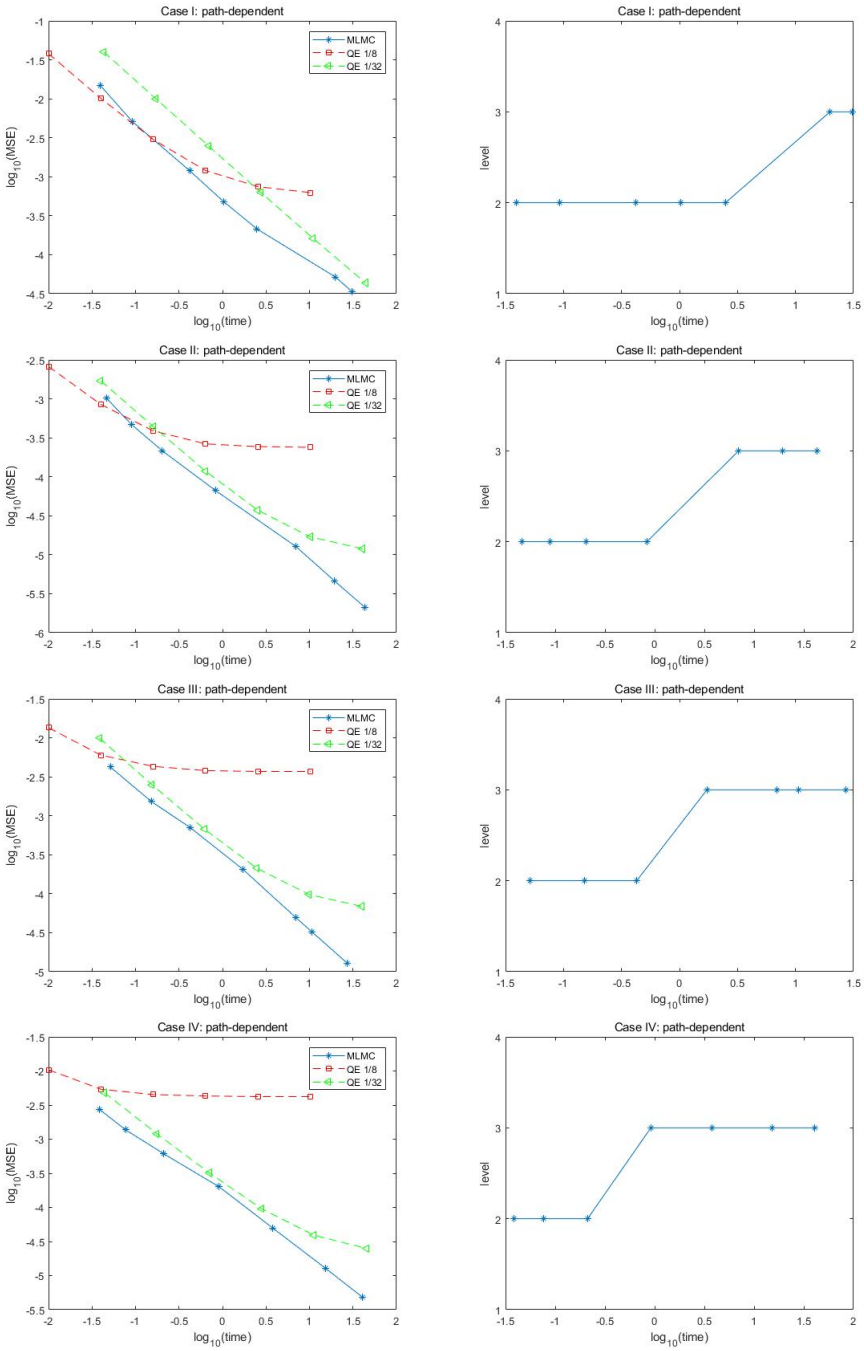

Figure 6: Comparison of NCI-MLMC and the QE scheme for path-dependent simulation. Left: computational time versus MSE. Right: the finest levels for MLMC.

where λ is the constant intensity of the Poisson process $(N(t))_{t\geq 0}$. The random variable J_i denotes the i-th jump, and it is normally distributed with constant mean μ_J and variance σ_J^2. The parameter $\bar{\mu}$ is specified to ensure that $e^{-rt}S_t$ is a martingale, and it satisfies $\ln(1+\bar{\mu}) = \mu_J + \frac{1}{2}\sigma_J^2$.

By applying Itô's Lemma, we have the SDE of the logarithmic asset price

$$\mathrm{d}\ln S_t = -\left(\frac{1}{2} + \lambda\bar{\mu}\right) V_t \mathrm{d}t + \rho\sqrt{V_t}\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\sqrt{V_t}\mathrm{d}W_t^2 + J_{N(t)}dN(t).$$

This SDE has two parts: one is the jump part $J_{N(t)}dN(t)$, and the other is the diffusion part. The diffusion part can be simulated in the same way as we did for the Heston model. The jump part is independent of the other elements of the SDE and can be simulated exactly. The MLMC estimators for the Heston model can be easily extended for the SVJ model, with an additional effort to address the jump part. Specifically, we keep the sample of $J_{N(t)}$ and that of $N(t)$ at the fine level the same as their counterparts at the coarse level. All convergence results we presented for the Heston model can be easily generalized for the SVJ model, since there is no error in simulating the jump part.

Heston model with CEV process

This model uses a mean-reverting constant elasticity of variance (CEV) process, discussed in Andersen and Piterbarg (2007). We have

$$\begin{aligned}\mathrm{d}S_t &= rS_t\mathrm{d}t + \sqrt{V_t}S_t(\rho\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2)\\ \mathrm{d}V_t &= k(\theta - V_t)\mathrm{d}t + \sigma V_t^{\gamma}\mathrm{d}W_t^1,\end{aligned}$$

where $\gamma \in [1/2, 1)$, and $S_0 > 0$, $V_0 > 0$. Note that when $\gamma = 1/2$, this is the Heston model. The CEV model has a unique strong solution. There are a number of time-discrete methods to simulate the CEV process, such as Andersen and Brotherton-Ratchie (2005), Lord et al. (2010), and Altmayer and Neuenkirch (2015). To price an option under the Heston model with the CEV process, we can use the approach from Jourdain and Sbai (2011). Let $F(x) = x^{\frac{3}{2}-\gamma}/(\frac{3}{2}-\gamma)$, and it follows by Itô's Lemma that

$$\mathrm{d}F(V_t) = \left(\left(k\theta + \frac{1}{4} - \frac{1}{2}\gamma\right)V_t^{\frac{1}{2}-\gamma} - kV_t^{\frac{3}{2}-\gamma}\right)\mathrm{d}t + \sigma\sqrt{V_t}\mathrm{d}W_t^1.$$

When $\gamma > \frac{1}{2}$, the CEV process is strictly positive with probability one and $\sup_{t\in[0,T]}\mathbb{E}(V_t^{-p}) < \infty$ for any non-negative p and T (Altmayer and Neuenkirch 2015). On the other hand, we have

$$\mathrm{d}\ln(S_t) = \left(r - \frac{1}{2}V_t\right)\mathrm{d}t + \sqrt{V_t}\left(\rho\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2\right).$$

Since the two SDEs above have the same term $\sqrt{V_t}\mathrm{d}W_t^1$, we combine them and obtain

$$\mathrm{d}\ln(S_t) = \left(r - \frac{1}{2}V_t - \frac{\rho}{\sigma}\left(k\theta V_t^{\frac{1}{2}-\gamma} + \left(\frac{1}{4} - \frac{\gamma}{2}\right)\sigma^2 V_t^{-\frac{1}{2}+\gamma} - kV_t^{\frac{3}{2}-\gamma}\right)\right)\mathrm{d}t + \frac{\rho}{\sigma}\mathrm{d}F(V_t) + \sqrt{1-\rho^2}\sqrt{V_t}\mathrm{d}W_t^2,$$

where the Brownian motion W^2 is independent of the CEV process V. Then the trapezoidal rule can be applied to approximate the time integral as we did for the Heston model, and the extension of the MLMC estimators for this model is trivial. To simulate the CEV process, we prefer an exact method (e.g., Lindsay and Brecher 2012) or a method that is a second-order weak scheme, to be consistent with the weak order of the stochastic trapezoidal rule.

Heston model with stochastic interest rate

This model extends the Heston model by adding a stochastic evolution of the interest rate. We can write the model as

$$\begin{aligned}\mathrm{d}S_t &= r_t S_t \mathrm{d}t + \sqrt{V_t} S_t (\rho \mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2)\\ \mathrm{d}V_t &= k(\theta - V_t)\mathrm{d}t + \sigma\sqrt{V_t}\mathrm{d}W_t^1,\end{aligned}$$

where the interest rate $(r_t)_{t\geq 0}$ is stochastic. We refer readers to Brigo and Mercurio (2006) for an overview of interest rate models. Among them, the Hull-White model and the CIR model (Cox, Ingersoll and Ross 1985) are of particular interest to industrial applications. Grzelak and Oosterlee (2011) approximated the price of the standard European option under the Heston model with the interest rate following the Hull-White model and the CIR model respectively. Let

$$\mathrm{d}r_t = \alpha(\beta - r_t)\mathrm{d}t + \gamma r_t^{\eta}\mathrm{d}W_t^3,$$

where $(W_t^3)_{t\geq 0}$ is a Brownian motion process independent of $(W_t^1)_{t\geq 0}$ and $(W_t^2)_{t\geq 0}$. When $\eta = 0$, it is the Hull-White model, and when $\eta = 1/2$, it is the CIR model. The parameters α, β and γ are set to be deterministic functions of time t for the Hull-White model, and are constant for the CIR model. Then, we have

$$\mathrm{d}\ln(S_t) = \left(\left(r_t - \frac{k\rho\theta}{\sigma}\right) + \left(\frac{k\rho}{\sigma} - \frac{1}{2}\right)V_t\right)\mathrm{d}t + \frac{\rho}{\sigma}\mathrm{d}V_t + \sqrt{1-\rho^2}\sqrt{V_t}\mathrm{d}W_t^2.$$

Since the interest rate process is independent of the variance process V and the Brownian motion W^2, the MLMC estimators for the Heston model can be easily extended to fit this model.

The next question is how to simulate the interest rate process. For the Hull-White model, the interest rate at any time is normally distributed, and thus, it is easy to simulate it exactly. For the CIR model, Alfonsi (2010) commented that 'In finance, such large values of σ ($\sigma^2 \gg 4k\theta$) do not occur when the CIR diffusion is used to represent the short interest rate'. Then a time-discrete scheme is generally preferred to approximate the solution, and we refer to Alfonsi (2005) for a comparison of a number of discretization schemes.

6 Conclusion

We apply the multilevel Monte Carlo technique to the stochastic trapezoidal rule for the Heston model, where the variance process is simulated exactly or almost exactly. The efficiency of MLMC relies on the convergence rate of $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ as the level l goes to infinity, and a high order is preferred. We recommend one MLMC estimator for path-independent options and two MLMC estimators for path-dependent options. The theoretical convergence rates of $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$, under some bounded Lipschitz assumptions on the option payoffs, are derived for all parameter regimes. Specifically, the convergence rate is 2 for path-independent options and 1 for path-dependent options, which is confirmed by our numerical tests. For practical applications, we combine the MLMC estimators with the NCI method. We compare NCI-MLMC with the gamma expansion method and the QE scheme in our numerical test, and there is some improvement in the efficiency of simulation. Since the MLMC estimators we develop are applicable for more general methods that simulate the variance process exactly or almost exactly, some further development can be expected.

Although our theoretical analysis is based on bounded Lipschitz assumptions, we find that the MLMC estimators applied on the standard European call option with unbounded Lipschitz payoff can also have good numerical performance. However, it is not clear to us whether the method is still efficient when the payoff is non-Lipschitz, in particular, when there are some discontinuities in the payoff function. This indicates a direction for future research.

Appendix

The NCI Method

We revisit the NCI method by Van Haastrecht and Pelsser (2010). Let N_{max} be a positive integer, $\mathcal{N} := \{0, 1, 2, ..., N_{max}\}$ be a set of some Poisson-values, and $\mathcal{U} := \{0, ..., 1-\delta\}$ be a sorted set of uniform variables, in which δ is a minimal machine number such as 10^{-15}. The NCI method starts with a pre-calculation of the inverse of the chi-squared distributions, i.e., it calculates

$$H_N^{-1}(U) := G^{-1}_{\mathcal{X}^2_{d+2N}}(U), \quad \forall N \in \mathcal{N}, \forall U \in \mathcal{U} \tag{28}$$

where $G^{-1}_{\mathcal{X}^2_{d+2N}}$ is the inverse of the chi-squared distribution with degree of freedom $d+2N$, and here $d = \frac{4\theta k}{\sigma^2}$. Recall that for the variance process, V_t at time t follows the scaled non-central chi-squared distribution

$$C_0 \chi^2_d(\lambda)$$

given V_u for any $u \leq t$, with $C_0 = \frac{\sigma^2(1-\mathrm{e}^{-k(t-u)})}{4k}$ and $\lambda = \frac{4k\mathrm{e}^{-k(t-u)}}{\sigma^2(1-\mathrm{e}^{-k(t-u)})} V_u$. The NCI method then generates one sample N from the Poisson distribution with the mean $\frac{\lambda}{2}$ and another sample U from the uniform distribution. A sample of V_t is generated by

$$F_N^{-1}(U) := \begin{cases} C_0 J(U), & N \leq N_{max} \\ C_0 F^{-1}_{\mathcal{X}^2_{d+2N}}(U), & N > N_{max} \end{cases} \tag{29}$$

where $J(.)$ represents an interpolation rule, and $F^{-1}_{\mathcal{X}^2_{d+2N}}(.)$ is the exact cumulative distribution function of the non-central chi-squared random variable.

The authors suggested two interpolations: the linear interpolation and the monotone cubic Hermite spline interpolation. Here, we prefer the latter, since it is much more accurate with the same number of points in the cache. For any $U_i < U < U_{i+1}$, where $U_i, U_{i+1} \in \mathcal{U}$, the monotone cubic Hermite spline interpolation can be defined as follows

$$J(U) := h_{00}(t) H_N^{-1}(U_i) + h_{01}(t) H_N^{-1}(U_{i+1}) + \Delta_i (m_i h_{10}(t) + m_{i+1} h_{11}(t))$$

where $t = \frac{U-U_i}{U_{i+1}-U_i}$, and

$$\begin{aligned} h_{00}(t) &= 2t^3 - 3t^2 + 1, \quad & h_{10}(t) &= t^3 - 2t^2 + t \\ h_{01}(t) &= -2t^3 + 3t^2, & h_{11}(t) &= t^3 - t^2 \end{aligned}$$

are the basis functions of the interpolation rule, and the coefficients Δ_i, m_i, m_{i+1} can be specified by the Fritsch-Carlson algorithm (Fritsch and Carlson 1980). For given data $(x_i, y_i) := (U_i, H_N^{-1}(U_i)), i = 0, 1, ..., n-1$, the Fritsch-Carlson algorithm is as follows

1. Set $m_0 = \frac{y_1 - y_0}{x_1 - x_0}$, $m_{n-1} = \frac{y_{n-1} - y_{n-2}}{x_{n-1} - x_{n-2}}$ and $\Delta_0 = \frac{y_1 - y_0}{x_1 - x_0}$;
2. Set $k = 1$, while $k < n - 1$
   (a) Set $m_k = \frac{1}{2}\left(\frac{y_k - y_{k-1}}{x_k - x_{k-1}} + \frac{y_{k+1} - y_k}{x_{k+1} - x_k}\right)$, $\Delta_k = \frac{y_{k+1} - y_k}{x_{k+1} - x_k}$;
   (b) Let $k = k + 1$;
3. Set $k = 1$, while $k < n - 1$
   (a) If $\Delta_k = 0$, set $m_k = m_{k+1} = 0$;
   (b) Let $k = k + 1$;
4. Set $k = 0$, while $k < n - 1$
   (a) Let $a_k = \frac{m_k}{\Delta_k}$, and $b_k = \frac{m_{k+1}}{\Delta_k}$;
   (b) If $a_k^2 + b_k^2 > 9$, define $t_k = \Delta_k \frac{3}{\sqrt{a_k^2 + b_k^2}}$, and set $m_k = t_k a_k$, $m_{k+1} = t_k b_k$;
   (c) Let $k = k + 1$.

This algorithm above, to calculate the coefficients of the monotone cubic Hermite spline interpolation, only has to be done once before running the Monte Carlo.

Now, we summarize the algorithm of the NCI method below

1. Set the maximum Poisson value N_{max} and the uniform grid $\mathcal{U}$. For any $N = 0, 1, ..., N_{max}$, use (28) to pre-compute the inverse of the chi-squared distribution on $\mathcal{U}$;
2. Generate a sample N from the Poisson distribution with the mean $\frac{\lambda}{2}$, and a sample U from the uniform distribution;
3. Use (29) to sample from the variance process.

As we can see, the accuracy of the NCI method relies on the parameter N_{max} and the size of the uniform grid $\mathcal{U}$. In our numerical test, for very high precision, we set $N_{max} = 20$ and $n = 10000$, where n is the size of $\mathcal{U}$, such that $\mathcal{U} = \{0, 1/n, 2/n., .., (n-1)/n, 1 - \delta\}$, with $\delta = 10^{-15}$.

Acknowledgement

Part of this research was done when I was at Heriot-Watt University. I would like to express my gratitude to Dr Anke Wiese and Dr Simon Malham, for their guidance and for polishing the article, and to Prof Mike Giles for very constructive discussions on multilevel Monte Carlo. Moreover, I also thank the referees of the previous submissions for their valuable comments and suggestions, which have led to a significant improvement on the quality of this article. In particular, one referee generously shared with me an approach to simplify the proof in the convergence analysis.

References

[1] Alfonsi, A. (2005). On the discretization schemes for the CIR (and Bessel squared) processes. Monte Carlo Methods and Applications, 11(4), 355-384.

[2] Alfonsi, A. (2010). High order discretization schemes for the CIR process: application to affine term structure and Heston models. Mathematics of Computation, 79(269), 209-237.

[3] Altmayer, M. and Neuenkirch, A. (2015). Multilevel Monte Carlo Quadrature of Discontinuous Payoffs in the Generalized Heston Model using Malliavin Integration by Parts. SIAM Journal of Financial Mathematics, 6(1), 22-52.

[4] Altmayer, M. and Neuenkirch, A. (2017). Discretizing the Heston Model: An Analysis of the Weak Convergence Rate. IMA Journal of Numerical Analysis, 37(4), 1930-1960.

[5] Andersen, L (2008). Simple and efficient simulation of the Heston stochastic volatility model. Journal of Computational Finance, 11(3), 1-41.

[6] Andersen, L and Brotherton-Ratchie, R (2005). Extended libor market models with stochastic volatility. Journal of Computational Finance, 9(1), 1-40.

[7] Andersen, L and Piterbarg, V (2007). Moment explosions in stochastic volatility models. Finance and Stochastics, 11(1), 29-50.

[8] Bates, D (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options, 9, 69-107.

[9] Broadie, M and Kaya, O (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes. Operations Research, 54(2), 217-231.

[10] Brigo, D and Mercurio, F (2006). Interest rate models–theory and practice: with smile, inflation and credit, Springer Verlag.

[11] Cox, J. Ingersoll, J. and Ross, S (1985). A theory of term structure of interest rates. Econometrica, 53(2), 385-407.

[12] Duffie, D and Glynn, P (1995). Efficient Monte Carlo simulation of security prices. Annals of Applied Probability, 5(4), 897-905.

[13] Dereich, S. Neuenkirch, A. and Szpruch, L (2012). An Euler-type method for the strong approximation of the Cox-Ingersoll-Ross process. Proceedings of the Royal Society A, 468, 1105-1115.

[14] Dufresne, D (2001). The integrated square-root process. Working Paper, University of Montreal. https://minerva-access.unimelb.edu.au/handle/11343/33693

[15] Fritsch, F and Carlson, R (1980). Monotone piecewise cubic interpolation. SIAM Journal on Numerical Analysis, 17(2), 238-246.

[16] Fischer, M and Nappo, G (2009). On the moments of the modulus of continuity of Itô processes. Stochastic Analysis and Applications, 2009, 28(1), 103-122.

[17] Giles, M and Szpruch, L (2014). Antithetic Multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. Annals of Applied Probability, 24(4), 1585-1620.

[18] Giles, M (2008). Improved multilevel Monte Carlo convergence using the Milstein scheme. Monte Carlo and Quasi-Monte Carlo Methods, 343-358.

[19] Giles, M (2008). Multilevel Monte Carlo Path Simulation. Operations Research, 56(3), 607-617.

[20] Giles, M. Higham, J. and Mao, X (2009). Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff. Finance and Stochastics, 13(3), 403-413.

[21] Giles, M (2015). Multilevel Monte Carlo methods. Acta Numerica, 24, 259-328.

[22] Glasserman, P and Kim, K (2011). Gamma expansion of the Heston stochastic volatility model. Finance and Stochastics, 15, 267-296.

[23] Glasserman, P (2003). Monte Carlo Methods in Financial Engineering. Springer Sciences and Business media, New York.

[24] Grzelak, L.A and Oosterlee, C.W (2011). On the Heston model with stochastic interest rates. SIAM Journal on Financial Mathematics, 2(1), 255-286.

[25] Hefter, M and Jentzen, A (2019). On arbitrarily slow convergence rates for strong numerical approximations of Cox-Ingersoll-Ross processes and squared Bessel processes. Finance and Stochastics 23, 139-172.

[26] Heston, S (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies, 6(2), 327 - 343.

[27] Higham, D and Mao, X (2005). Convergence of the Monte Carlo simulations involving the meanreverting square root process. Journal of Computational Finance, 8(3), 35-62.

[28] Hutzenthaler, M., Jentzen, A and Noll, M (2014). Strong convergence rates and temporal regularity for Cox-Ingersoll-Ross processes and Bessel processes with accessible boundaries. Preprint. https://arxiv.org/abs/1403.6385

[29] Lindsay, A and Brecher, D (2012). Simulation of the CEV process and the local martingale property. Mathematics and Computers in Simulation, 82(5), 868-878.

[30] Jourdain, B and Sbai, M (2013). High order discretization schemes for stochastic volatility models, Journal of Computational Finance 17, 113-165.

[31] Kahl, C and Jäckel, P (2006). Fast strong approximation Monte Carlo schemes for stochastic volatility models. Quantitative Finance, 6, 513-536.

[32] Kloeden, P and Neuenkirch, A (2012). Convergence of numerical methods for stochastic differential equations in mathematical finance. Recent Developments in Computational Finance.

[33] Kloeden, P and Platen, E (1999). Numerical Solution of Stochastic Differential Equations, 3rd edition, Springer Verlag, New York.

[34] Lord, R. Koekkoek, R and Van Dijk, D (2010) A comparison of biased simulation schemes for stochastic volatility models. Quantitative Finance, 10(2), 177-194.

[35] Malham, S.J.A. and Wiese, A (2013). Chi-square simulation of the CIR process and the Heston model. International Journal of Theoretical and Applied Finance, 16(3), 1-38.

[36] Marsaglia, G and Tsang, W (2000). A simple method for generating gamma variables. Transactions on Mathematical Software, 26(3), 363-372.

[37] Smith, R.D (2007). An almost exact simulation method for the Heston model. Journal of Computational Finance, 11(1), 115-125.

[38] Van Haastrecht, A and Pelsser, A (2010). Efficient, almost exact simulation of the Heston stochastic volatility model. International Journal of Theoretical and Applied Finance, 13(1), 1-41.

[39] Zheng, C (2017). Weak convergence rate of a time-discrete scheme for the Heston stochastic volatility model. SIAM Journal on Numerical Analysis, 55(3), 1243-1263.