

# Technical report

## The computation of Greeks with Multilevel Monte Carlo

Sylvestre Burgos, M.B. Giles  
Oxford-Man Institute of Quantitative Finance  
University of Oxford

December 2010

### Abstract

We study the use of the multilevel Monte Carlo technique [2, 3] in the context of the calculation of Greeks. The pathwise sensitivity analysis [5] differentiates the path evolution and reduces the payoff's smoothness. This leads to new challenges: the inapplicability of pathwise sensitivities to non-Lipschitz payoffs often makes the use of naive algorithms impossible.

These challenges can be addressed in three different ways: payoff smoothing using conditional expectations of the payoff before maturity [5]; approximating the previous technique with path splitting for the final timestep [1]; using of a hybrid combination of pathwise sensitivity and the Likelihood Ratio Method [4]. We investigate the strengths and weaknesses of these alternatives in different multilevel Monte Carlo settings.

## 1 Introduction

In mathematical finance, Monte Carlo methods are used to compute the price of an option by estimating the expected value  $\mathbb{E}(P)$ .  $P$  is the payoff function that depends on an underlying asset's scalar price  $S(t)$  which satisfies an evolution SDE of the form

$$dS(t) = a(S,t)dt + b(S,t)dW_t, \quad 0 \leq t \leq T, \quad S(0) \text{ given}, \quad (1)$$

This is just one use of Monte Carlo in finance. In practice the prices are often quoted and used to calibrate our market models; the option's sensitivities to market parameters, the so-called Greeks, reflect the exposure to different sources of risk. Computing these is essential to hedge portfolios and is therefore even more important than pricing the option itself. This is why our research focuses on getting fast and accurate estimates of Greeks through Monte Carlo simulations.

## 1.1 Multilevel Monte Carlo

Let us consider a standard Monte Carlo method using a discretisation with first order weak convergence (e.g. the Milstein scheme). Achieving a root-mean square error of  $O(\varepsilon)$  requires a variance of order  $O(\varepsilon^2)$ , hence  $O(\varepsilon^{-2})$  independent paths. It also requires a discretisation bias of order  $O(\varepsilon^{-1})$ , thus  $O(\varepsilon^{-1})$  timesteps, giving a total computational cost  $O(\varepsilon^{-3})$ .

Giles' multilevel Monte Carlo technique reduces this cost to  $O(\varepsilon^{-2})$  under certain conditions. The idea is to write the expected payoff with a fine discretisation using  $2^{-L}$  uniform timesteps as a telescopic sum. Let  $\hat{P}_l$  be the simulated payoff with a discretisation using  $2^l$  uniform timesteps,

$$\mathbb{E}(\hat{P}_L) = \mathbb{E}(\hat{P}_0) + \sum_{l=1}^L \mathbb{E}(\hat{P}_l - \hat{P}_{l-1}) \quad (2)$$

We then use Monte Carlo estimators using  $N_l$  independent samples

$$\mathbb{E}(\hat{P}_l - \hat{P}_{l-1}) \approx \hat{Y}_l = \frac{1}{N_l} \sum_{i=1}^{N_l} (\hat{P}_l^{(i)} - \hat{P}_{l-1}^{(i)}) \quad (3)$$

The small corrective term  $\hat{P}_l^{(i)} - \hat{P}_{l-1}^{(i)}$  comes from the difference between a fine and a coarse discretisation of the same leading Brownian motion. Its magnitude depends on the strong convergence properties of the scheme used. Let  $V_l$  be the variance of a single sample  $\hat{P}_l^{(i)} - \hat{P}_{l-1}^{(i)}$ . The next theorem shows that what determines the efficiency of the multilevel approach is the convergence speed of  $V_l$  as  $l \rightarrow \infty$ .

To ensure a better efficiency we may modify (3) and use different estimators of  $\hat{P}$  on the fine and coarse levels of  $\hat{Y}_l$  as long as the telescoping sum property is respected, that is

$$\mathbb{E}(\hat{P}_L) = \mathbb{E}(\hat{P}_0) + \sum_{l=1}^L \mathbb{E}(\hat{P}_l^f - \hat{P}_{l-1}^c) \quad (4)$$

with

$$\mathbb{E}(\hat{P}_l^f) = \mathbb{E}(\hat{P}_l^c). \quad (5)$$

**Theorem 1.1.** *Let  $P$  be a function of a solution to (1) for a given Brownian path  $W(t)$ ; let  $\hat{P}_l$  be the corresponding approximation using the discretisation at level  $l$ , i.e. with  $2^l$  steps of width  $h_l = 2^{-l} T$ .*

*If there exist independent estimators  $\hat{Y}_l$  of computational complexity  $C_l$  based on  $N_l$  samples and there are positive constants  $\alpha \geq \frac{1}{2}, \beta, c_1, c_2, c_3$  such that*

1.  $\mathbb{E}(\hat{Y}_l) = \begin{cases} \mathbb{E}(\hat{P}_0) & \text{if } l = 0 \\ \mathbb{E}(\hat{P}_l - \hat{P}_{l-1}) & \text{if } l > 0 \end{cases}$
2.  $|\mathbb{E}(\hat{P}_l - P)| \leq c_1 h_l^\alpha$
3.  $\mathbb{V}(\hat{Y}_l) \leq c_2 h_l^\beta N_l^{-1}$
4.  $C_l \leq c_3 N_l h_l^{-1}$

*Then there is a constant  $c_4$  such that for any  $\varepsilon < e^{-1}$ , there are values for  $L$  and  $N_l$  resulting in a multilevel estimator  $\hat{Y} = \sum_{l=0}^L \hat{Y}_l$  with a mean-square-error  $MSE = \mathbb{E}((\hat{Y} - \mathbb{E}(P))^2) < \varepsilon^2$  with a complexity  $C$  bounded by*

$$C \leq \begin{cases} c_4 \varepsilon^{-2} & \text{if } \beta > 1 \\ c_4 \varepsilon^{-2} (\log \varepsilon)^2 & \text{if } \beta = 1 \\ c_4 \varepsilon^{-2-(1-\beta)/\alpha} & \text{if } 0 < \beta < 1 \end{cases} \quad (6)$$

*Proof.* See [3]. □

We usually know  $\alpha$  thanks to the literature on weak convergence. Results in [6] give  $\alpha = 1$  for the Milstein scheme, even in the case of discontinuous payoffs.  $\beta$  is related to strong convergence and is practically what determines the efficiency of the multilevel approach. Its value depends on the payoff shape and is usually not known *a priori*.

## 1.2 Monte Carlo Greeks

Let us briefly recall two classic methods used to compute Greeks in a Monte Carlo setting: the pathwise sensitivities and the Likelihood Ratio Method. More details can be found in [5].

### Pathwise sensitivities

Let  $\hat{S} = (\hat{S}_n)_{n \in [0, N]}$  be the simulated values of the asset at the discretisation times and  $\hat{W} = (\hat{W}_n)_{n \in [1, N]}$  be the corresponding set of Brownian increments. Assuming that the payoff  $P(\hat{S})$  is Lipschitz, we can use the chain rule and write

$$\frac{\partial \hat{V}}{\partial \theta} = \frac{\partial}{\partial \theta} \int P(\hat{S}(\theta, \hat{W})) p(\hat{W}) d\hat{W} = \int \frac{\partial P(\hat{S})}{\partial \hat{S}} \frac{\partial \hat{S}(\theta, \hat{W})}{\partial \theta} p(\hat{W}) d\hat{W}$$

where  $d\hat{W} = \prod_{k=1}^N d\hat{W}_k$  and  $p(\hat{W}) = \prod_{k=1}^N p(\hat{W}_k)$  is a multivariate Normal probability density function.

We obtain  $\frac{\partial \hat{S}}{\partial \theta}$  by differentiating the discretisation of (1) with respect to  $\theta$  and iterating the resulting formula. The limitation of this technique is that it requires the payoff to be Lipschitz and piecewise differentiable.

### Likelihood Ratio Method

The Likelihood Ratio Method consists in writing

$$V = \mathbb{E} [P(\hat{S})] = \int P(\hat{S}) p(\theta, \hat{S}) d\hat{S} \quad (7)$$

The dependence on  $\theta$  comes through the probability density function  $p(\theta, \hat{S})$ ; assuming some conditions discussed in [5], we can write

$$\frac{\partial V}{\partial \theta} = \int P(\hat{S}) \frac{\partial p(\hat{S})}{\partial \theta} d\hat{S} = \int P(\hat{S}) \frac{\partial \log p(\hat{S})}{\partial \theta} p(\hat{S}) d\hat{S} = \mathbb{E} \left[ P(\hat{S}) \frac{\partial \log p(\hat{S})}{\partial \theta} \right] \quad (8)$$

$$\text{with } d\hat{S} = \prod_{k=1}^N d\hat{S}_k \quad \text{and} \quad p(\hat{S}) = \prod_{k=1}^N p(\hat{S}_k | \hat{S}_{k-1})$$

The main limitation of the method is that the estimator's variance is  $O(N)$ , becoming infinite as we refine the discretisation.

### 1.3 Multilevel Monte Carlo Greeks

By combining the elements of sections 1.1 and 1.2 together, we write

$$\frac{\partial V}{\partial \theta} = \frac{\partial \mathbb{E}(P)}{\partial \theta} \approx \frac{\partial \mathbb{E}(\hat{P}_L)}{\partial \theta} = \frac{\partial \mathbb{E}(\hat{P}_0)}{\partial \theta} + \sum_{l=1}^L \frac{\partial \mathbb{E}(\hat{P}_l - \hat{P}_{l-1})}{\partial \theta} \quad (9)$$

As in (3), we define the multilevel estimators

$$\hat{Y}_0 = N_0^{-1} \sum_{i=1}^M \frac{\partial \hat{P}_0^{(i)}}{\partial \theta} \quad \text{and} \quad \hat{Y}_l = N_l^{-1} \sum_{i=1}^{N_l} \left( \frac{\partial \hat{P}_l^{(i)}}{\partial \theta} - \frac{\partial \hat{P}_{l-1}^{(i)}}{\partial \theta} \right) \quad (10)$$

where  $\frac{\partial \hat{P}_0}{\partial \theta}$ ,  $\frac{\partial \hat{P}_{l-1}}{\partial \theta}$ ,  $\frac{\partial \hat{P}_l}{\partial \theta}$  are computed with the techniques presented in section 1.2.

## 2 European call

We first consider a Lipschitz payoff. That of the European call is

$$P = (S_T - K)^+ = \max(0, S_T - K)$$

We illustrate the techniques by computing delta ( $\delta$ ) and vega ( $\nu$ ), the sensitivities to the asset's initial value  $S_0$  and to its volatility  $\sigma$ .

### 2.1 Pathwise sensitivities

We consider the Black-Scholes model: the asset's evolution is modelled by a geometric Brownian motion  $dS(t) = rS(t)dt + \sigma S(t)dW_t$ . We use the Milstein scheme for its good strong convergence properties. For timesteps of width  $h$ ,

$$\hat{S}_{n+1} = \hat{S}_n \cdot \left( 1 + rh + \sigma \Delta W_n + \frac{\sigma^2}{2} (\Delta W_n^2 - h) \right) := \hat{S}_n \cdot D_n \quad (11)$$

Since the payoff is Lipschitz, we can use pathwise sensitivities. The differentiation of equation (11) gives

$$\begin{cases} \frac{\partial \hat{S}_0}{\partial S_0} = 1, & \frac{\partial \hat{S}_{n+1}}{\partial S_0} = \frac{\partial \hat{S}_n}{\partial S_0} \cdot D_n \\ \frac{\partial \hat{S}_0}{\partial \sigma} = 0, & \frac{\partial \hat{S}_{n+1}}{\partial \sigma} = \frac{\partial \hat{S}_n}{\partial \sigma} \cdot D_n + \hat{S}_n (\Delta W_n + \sigma(\Delta W_n^2 - h)) \end{cases} \quad (12)$$

To compute  $\hat{Y}_l$  we use a fine and a coarse discretisation with  $N_f = 2^l$  and  $N_c = 2^{l-1}$  uniform timesteps respectively. We denote these by the superscripts  $(l)$  and  $(l-1)$ . We take  $N_l$  samples to compute

$$\hat{Y}_l = \frac{1}{N_l} \sum_{i=1}^{N_l} \left[ \left( \frac{\partial P}{\partial S_{N_f}} \frac{\partial \hat{S}_{N_f}^{(i)}}{\partial \theta} \right)^{(l)} - \left( \frac{\partial P}{\partial S_{N_c}} \frac{\partial \hat{S}_{N_c}^{(i)}}{\partial \theta} \right)^{(l-1)} \right]$$

We use the same leading Brownian motion for the fine and coarse discretisations: we first generate the fine Brownian increments  $\hat{W} = (\Delta W_1, \Delta W_2, \dots, \Delta W_{N_f})$  and then use  $\hat{W}^c = (\Delta W_1 + \Delta W_2, \dots, \Delta W_{N_f-1} + \Delta W_{N_f})$  as the coarse level's increments.

## Estimated complexity and analysis

Unless otherwise stated, the simulations used to illustrate this paper use the parameters  $S_0 = 100$ ,  $K = 100$ ,  $r = 0.05$ ,  $\sigma = 0.20$ ,  $T = 1$ .

In figure 1 we plot  $(\mathbb{V}(\hat{Y}_l))$  as a function of  $(1/h_l)$  in a log-log plot. We then measure the slopes for the different estimators: this gives a numerical estimate of the parameter  $\beta$  in theorem 1.1. Combining this with the theorem, we get an estimated complexity of the multilevel algorithm (table 1). This gives the following results :

Figure 1: Pathwise sensitivities, European call :  $\mathbb{V}(\hat{Y}_l)(l)$

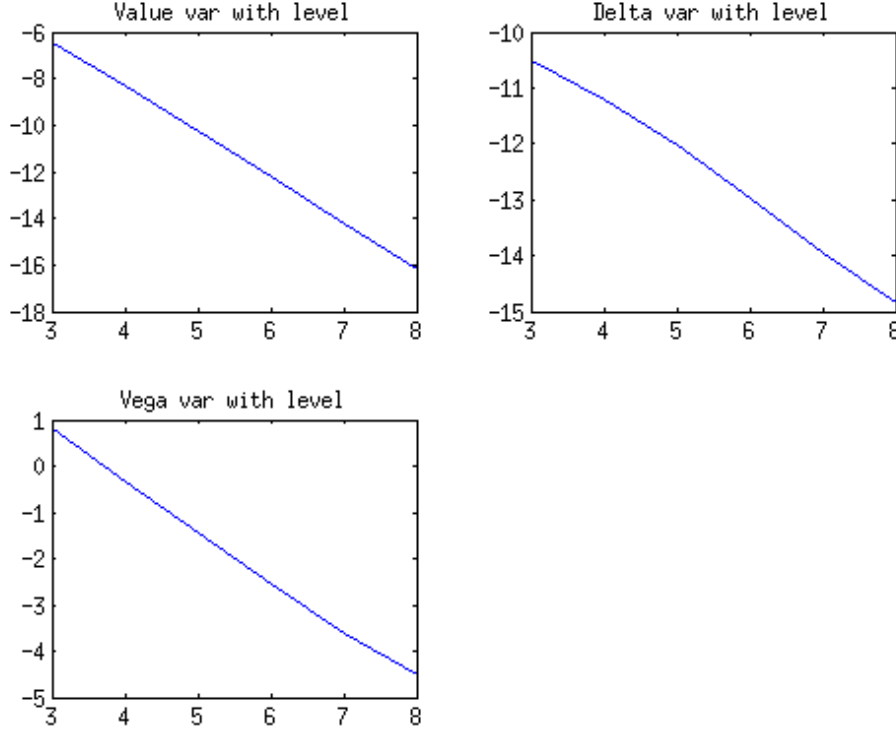


Table 1: Pathwise sensitivities, European call : estimated complexity

Estimator	$\beta$	MLMC Complexity
Value	$\approx 2.0$	$O(\varepsilon^{-2})$
Delta	$\approx 0.8$	$O(\varepsilon^{-2.2})$
Vega	$\approx 1.0$	$O(\varepsilon^{-2} \log \varepsilon^2)$

Giles has shown in [2] that  $\beta = 2$  for the value's estimator. For Greeks, the convergence is degraded by the discontinuity of  $\frac{\partial P}{\partial S} = \mathbf{1}_{S > K}$ : a fraction  $O(h_l)$  of the paths has a final value  $\hat{S}$  which is  $O(h_l)$  from the discontinuity  $K$ . For these paths, there is a  $O(1)$  probability that  $\hat{S}_{N_f}^{(l)}$  and  $\hat{S}_{N_c}^{(l-1)}$  are on different sides of the strike  $K$ , implying  $\left( \frac{\partial P}{\partial S_{N_f}} \frac{\partial \hat{S}_{N_f}}{\partial \theta} \right)^{(l)} - \left( \frac{\partial P}{\partial S_{N_c}} \frac{\partial \hat{S}_{N_c}}{\partial \theta} \right)^{(l-1)} = O(1)$ . Thus  $\mathbb{V}(\hat{Y}_l) = O(h_l)$ , and  $\beta = 1$  for the Greeks.

## 2.2 Pathwise sensitivities and Conditional Expectations

We have seen that the payoff's lack of smoothness prevents the variance of Greeks' estimators  $\hat{Y}_l$  from decaying quickly and limits the potential benefits of the multilevel approach. To improve the convergence speed, we can use conditional expectations [5]. Instead of simulating the whole path, we stop at the penultimate step and then for every fixed set  $\hat{W} = (\Delta W_k)_{k \in [1, N-1]}$ , we consider the full distribution of  $(\hat{S}_N | \hat{W})$ . With  $a_n = a(\hat{S}_{N-1}(\hat{W}), (N-1)h)$  and  $b_n = b(\hat{S}_{N-1}(\hat{W}), (N-1)h)$ , we can write

$$\hat{S}_N(\hat{W}, \hat{W}_N) = \hat{S}_{N-1}(\hat{W}) + a_n(\hat{W})h + b_n(\hat{W})\hat{W}_N \quad (13)$$

We hence get a normal distribution for  $(\hat{S}_N | \hat{W})$ .

$$p(\hat{S}_N | \hat{W}) = \frac{1}{\sigma_{\hat{W}} \sqrt{2\pi}} \exp\left(-\frac{(\hat{S}_N - \mu_{\hat{W}})^2}{2\sigma_{\hat{W}}^2}\right) \quad (14)$$

$$\text{with } \begin{cases} \mu_{\hat{W}} = \hat{S}_{N-1} + a(\hat{S}_{N-1}, (N-1)h)h \\ \sigma_{\hat{W}} = b(\hat{S}_{N-1}, (N-1)h)\sqrt{h} \end{cases}$$

We can thus compute  $\mathbb{E}[P(\hat{S}_N) | \hat{W}]$ . Using the chain rule, we get

$$\hat{V} = \mathbb{E}[P(\hat{S}_N)] = \mathbb{E}_{\hat{W}}[\mathbb{E}_{\Delta W_N}[P(\hat{S}_N) | \hat{W}]] \approx \frac{1}{M} \sum_{m=1}^M \mathbb{E}[P(\hat{S}_N^{(m)}) | \hat{W}^{(m)}] \quad (15)$$

Here with  $\phi$  the normal probability density function,  $\Phi$  the normal cumulative distribution functions,  $\alpha = (1 + rh)\hat{S}_{N-1}(\hat{W})$  and  $\beta = \sigma \sqrt{h}\hat{S}_{N-1}(\hat{W})$ , we get

$$\mathbb{E}(P(\hat{S}_N) | \hat{W}) = \beta \phi\left(\frac{\alpha - K}{\beta}\right) + (\alpha - K) \Phi\left(\frac{\alpha - K}{\beta}\right) \quad (16)$$

This expected payoff is infinitely differentiable with respect to the input parameters. We can apply the pathwise sensitivities technique to this smooth function at time  $(N-1)h$ . The multilevel estimator for the Greek is then

$$\hat{Y}_l = \frac{1}{N_l} \sum_1^{N_l} \left[ \left( \frac{\partial \hat{P}_f^{(i)}}{\partial \theta} \right)^{(l)} - \left( \frac{\partial \hat{P}_c^{(i)}}{\partial \theta} \right)^{(l-1)} \right] \quad (17)$$

At the fine level we use (16) with  $h = h_f$  and  $\hat{W}_f = (\Delta W_1, \Delta W_2, \dots, \Delta W_{N_f-1})$  to get  $\mathbb{E}(P(\hat{S}_{N_f}) | \hat{W}_f)$ . We then use

$$\left( \frac{\partial \hat{P}_f}{\partial \theta} \right)^{(l)} = \frac{\partial \hat{S}_{N_f-1}}{\partial \theta} \frac{\partial \mathbb{E}(P(\hat{S}_{N_f}) | \hat{W}_f)}{\partial S_{N_f-1}} + \frac{\partial \mathbb{E}(P(\hat{S}_{N_f}) | \hat{W}_f)}{\partial \theta} \quad (18)$$

At the coarse level, directly using  $\mathbb{E}(P(\hat{S}_{N_c}) | \hat{W}_c)$  leads to an unsatisfactorily low convergence rate of  $\mathbb{V}(\hat{Y}_l)$ . As explained in (4) we use a modified estimator. The idea is to include the final fine Brownian increment in the computation of the expectation over the last coarse timestep. This guarantees that the two paths will be close to one another and helps achieve better variance convergence rates.

$\hat{S}$  still follows a simple Brownian motion with constant drift and volatility on all coarse steps. With  $\hat{W}_c = (\Delta W_1 + \Delta W_2, \dots, \Delta W_{N_f-3} + \Delta W_{N_f-2})$  and given that the Brownian increment on the first half of the final step is  $\Delta W_{N_f-1}$ , we get

$$p(\hat{S}_{N_c} | \hat{W}_c, \Delta W_{N_f-1}) = \frac{1}{\sigma_{\hat{W}_c} \sqrt{2\pi}} \exp\left(-\frac{(\hat{S}_{N_c} - \mu_{\hat{W}_c})^2}{2\sigma_{\hat{W}_c}^2}\right) \quad (19)$$

$$\text{with } \begin{cases} \mu_{\hat{W}_c} = \hat{S}_{N_c-1}(\hat{W}_c) + a(\hat{S}_{N_c-1}, (N_c-1)h_c)h_c + b(\hat{S}_{N_c-1}, (N_c-1)h_c)\Delta W_{N_f-1} \\ \sigma_{\hat{W}_c} = b(\hat{S}_{N_c-1}, (N_c-1)h_c)\sqrt{h_c/2} \end{cases}$$

From this distribution we derive  $\mathbb{E}[P(\hat{S}_{N_c}) | \hat{W}_c, \Delta W_{N_f-1}]$ , which leads to the same payoff formula as before with  $\alpha_c = (1 + rh_c + \sigma\Delta W_{N_f-1})\hat{S}_{N_c-1}(\hat{W}_c)$  and  $\beta_c = \sigma\sqrt{h_c}\hat{S}_{N_c-1}(\hat{W}_c)$ . Using it as the coarse level's payoff does not introduce any bias. Using the tower property we check that it satisfies condition (5),

$$\mathbb{E}_{\Delta W_{N_f-1}}[\mathbb{E}[P(\hat{S}_{N_c}) | \hat{W}_c, \Delta W_{N_f-1}] | \hat{W}_c] = \mathbb{E}[P(\hat{S}_{N_c}) | \hat{W}_c]$$

### Estimated complexity and analysis

Our numerical experiments show the benefits of the conditional expectation technique on the European call:

Figure 2: Pathwise sensitivities and conditional expectations, European call :  $\mathbb{V}(\hat{Y}_l)(l)$

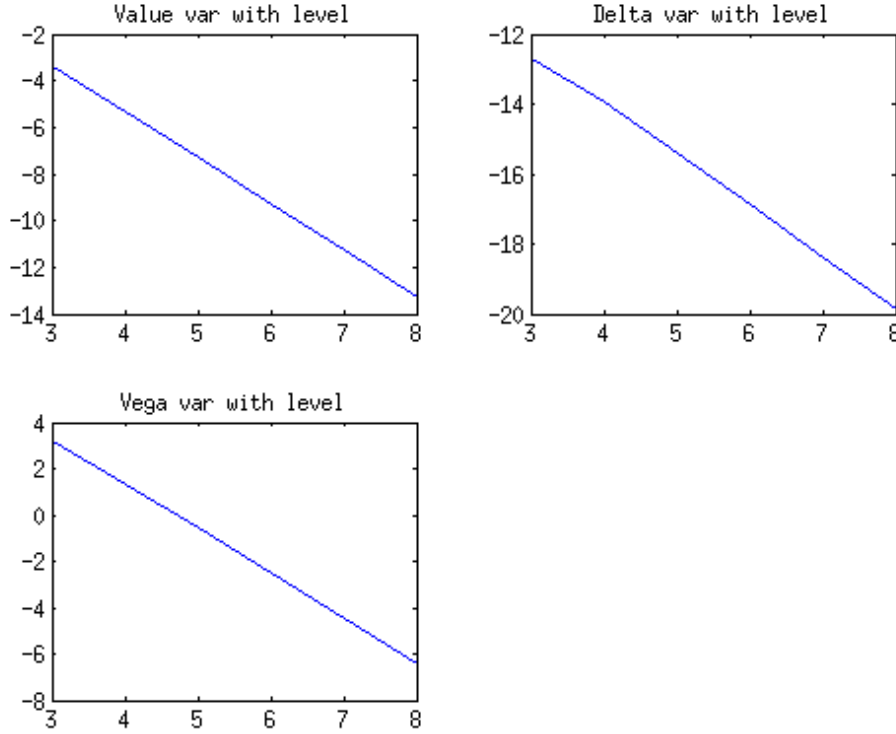


Table 2: Pathwise sensitivities and conditional expectations,  
European call : estimated complexity

Estimator	$\beta$	MLMC Complexity
Value	$\approx 2.0$	$O(\varepsilon^{-2})$
Delta	$\approx 1.5$	$O(\varepsilon^{-2})$
Vega	$\approx 2.0$	$O(\varepsilon^{-2})$

A fraction  $O(\sqrt{h_l})$  of the paths arrive in the area around the strike where the conditional expectation  $\frac{\partial \mathbb{E}(P(\hat{S}_N)|\hat{W})}{\partial \hat{S}_{N_f-1}}$  is neither close to 0 nor 1. In this area, its slope is  $O(h_l^{-1/2})$ . The coarse and fine paths differ by  $O(h_l)$ , we thus have  $O(\sqrt{h_l})$  difference between the coarse and fine Greeks' estimates. Reasoning as in [2] we get  $\mathbb{V}_{\hat{W}}(\mathbb{E}_{\Delta W_N}(\dots|\hat{W})) = O(h_l^{3/2})$  for the Greeks' estimators. This is the convergence rate observed for  $\delta$ ; the higher convergence rate of  $v$  is not explained yet by this rough analysis and will be investigated in our future research.

The main limitation of this approach is that in many situations it leads to complicated integral computations. Path splitting, to be discussed next, may represent a useful numerical approximation to this technique.

### 2.3 Split pathwise sensitivities

This technique is based on the previous one. The idea is to avoid the tricky computation of  $\mathbb{E}[P(\hat{S}_{N_f})|\hat{W}_f]$  and  $\mathbb{E}[P(\hat{S}_{N_c})|\hat{W}_c, \Delta W_{N_f-1}]$ . We get numerical estimates of these values by “splitting” every path simulation on the final timestep.

At the fine level: for every simulated path  $\hat{W}_f = (\Delta W_1, \Delta W_2, \dots, \Delta W_{N_f-1})$ , we simulate a set of  $d$  final increments  $(\Delta W_{N_f}^{(i)})_{i \in [1, d]}$  which we average to get

$$\mathbb{E}[P(\hat{S}_{N_f})|\hat{W}_f] \approx \frac{1}{d} \sum_{i=1}^d P(\hat{S}_{N_f}(\hat{W}_f, \Delta W_{N_f}^{(i)})) \quad (20)$$

At the coarse level we use  $\hat{W}_c = (\Delta W_1 + \Delta W_2, \dots, \Delta W_{N_f-3} + \Delta W_{N_f-2})$ . As before (still assuming a constant drift and volatility on the final coarse step), we improve the convergence rate of  $\mathbb{V}(\hat{Y}_l)$  by reusing  $\Delta W_{N_f-1}$  in our estimation of  $\mathbb{E}[P(\hat{S}_{N_c})|\hat{W}_c]$ . We can do so by constructing the final coarse increments as  $(\Delta W_{N_c}^{(i)})_{i \in [1, d]} = (\Delta W_{N_f-1} + (\Delta W_{N_f}^{(i)}))_{i \in [1, d]}$  and using these to estimate

$$\mathbb{E}(P(\hat{S}_{N_c})|\hat{W}_c) = \mathbb{E}[P(\hat{S}_{N_c})|\hat{W}_c, \Delta W_{N_f-1}] \approx \frac{1}{d} \sum_{i=1}^d P(\hat{S}_{N_c}(\hat{W}_c, \Delta W_{N_c}^{(i)}))$$

To get the Greeks, we simply compute the corresponding pathwise sensitivities.

#### Estimated complexity and choice of the number of splittings

As expected this method yields higher values of  $\beta$  than simple pathwise sensitivities: the convergence rates increase and tend to the rates offered by conditional expectations as  $d$  increases and the approximation gets more precise.



Figure 3: Pathwise sensitivities and path splitting,  
European call :  $\mathbb{V}(\hat{Y}_l)(l)$

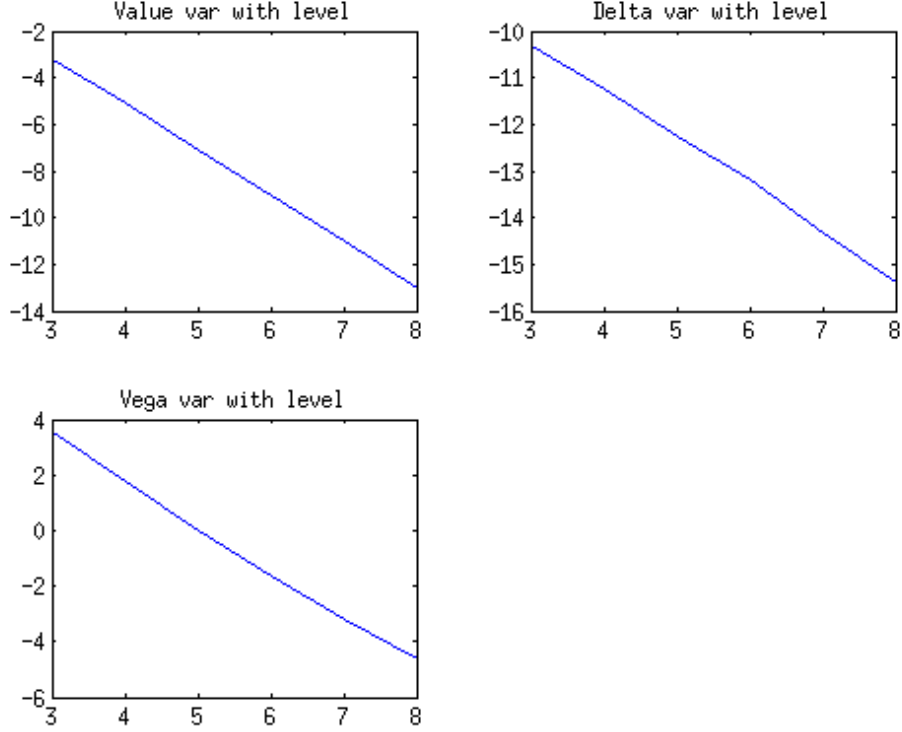


Table 3: Pathwise sensitivities and path splitting,  
European call : estimated complexity

Estimator	$d$	$\beta$	MLMC Complexity
Value	10	$\approx 2.0$	$O(\varepsilon^{-2})$
	500	$\approx 2.0$	$O(\varepsilon^{-2})$
Delta	10	$\approx 1.0$	$O(\varepsilon^{-2}(\log \varepsilon)^2)$
	500	$\approx 1.5$	$O(\varepsilon^{-2})$
Vega	10	$\approx 1.6$	$O(\varepsilon^{-2})$
	500	$\approx 2.0$	$O(\varepsilon^{-2})$

Taking a constant number of splittings  $d$  for all levels is actually not optimal; For Greeks, we can write the variance of the estimator as

$$\begin{aligned} \mathbb{V}(\hat{Y}_l) &= \frac{1}{N_l} \mathbb{V}_{\hat{W}_f}(\mathbb{E}(\left(\frac{\partial \hat{P}_f}{\partial \theta}\right)^{(l)} - \left(\frac{\partial \hat{P}_c}{\partial \theta}\right)^{(l-1)} \mid \hat{W}_f)) \\ &\quad + \frac{1}{N_l d} \mathbb{E}_{\hat{W}_f}(\mathbb{V}(\left(\frac{\partial \hat{P}_f}{\partial \theta}\right)^{(l)} - \left(\frac{\partial \hat{P}_c}{\partial \theta}\right)^{(l-1)} \mid \hat{W}_f)) \end{aligned} \quad (21)$$

As explained in section 2.2 we have  $\mathbb{V}_{\hat{W}_f}(\mathbb{E}(\dots \mid \hat{W}_f)) = O(h_l^{3/2})$  for the Greeks. We also

have  $\mathbb{E}_{\hat{W}_f}(\nabla(\dots|\hat{W}_f)) = O(h_l)$  for similar reasons. We optimise the variance at a fixed computational cost by choosing  $d$  such that the two terms of the sum are of similar order. Taking  $d = O(h_l^{-1/2})$  is therefore optimal.

## 2.4 Vibrato Monte Carlo

Since the previous method uses pathwise sensitivity analysis, it is not applicable when payoffs are discontinuous. To address this limitation, we use the Vibrato Monte Carlo method introduced by Giles [4]. This hybrid method combines pathwise sensitivities and the Likelihood Ratio Method.

We consider again equation (15). We now use the Likelihood Ratio Method on the last timestep and with the notations of section 2.2 we get

$$\frac{\partial \hat{V}}{\partial \theta} = \mathbb{E}_{\hat{W}} \left[ \mathbb{E}_{\Delta W_N} \left[ P(\hat{S}_N) \frac{\partial(\log p(\hat{S}_N|\hat{W}))}{\partial \theta} \Big| \hat{W} \right] \right] \quad (22)$$

We can write  $p(\hat{S}_N|\hat{W})$  as  $p(\mu_{\hat{W}}, \sigma_{\hat{W}})$ . This leads to the estimator

$$\begin{aligned} \frac{\partial \hat{V}}{\partial \theta} \approx \frac{1}{N_l} \sum_{m=1}^{N_l} & \left( \frac{\partial \mu_{\hat{W}^{(m)}}}{\partial \theta} \mathbb{E}_{\Delta W_N} \left[ P(\hat{S}_N) \frac{\partial(\log p)}{\partial \mu_{\hat{W}}} \Big| \hat{W}^{(m)} \right] \right. \\ & \left. + \frac{\partial \sigma_{\hat{W}^{(m)}}}{\partial \theta} \mathbb{E}_{\Delta W_N} \left[ P(\hat{S}_N) \frac{\partial(\log p)}{\partial \sigma_{\hat{W}}} \Big| \hat{W}^{(m)} \right] \right) \end{aligned} \quad (23)$$

We compute  $\frac{\partial \mu_{\hat{W}^{(m)}}}{\partial \theta}$  and  $\frac{\partial \sigma_{\hat{W}^{(m)}}}{\partial \theta}$  with pathwise sensitivities.

With  $\hat{S}_N^{(m,i)} = \hat{S}_N(\hat{W}^{(m)}, \Delta W_N^{(i)})$ , we substitute the following estimators into (23)

$$\left\{ \begin{array}{l} \mathbb{E}_{\Delta W_N} \left[ P(\hat{S}_N) \frac{\partial(\log p)}{\partial \mu_{\hat{W}}} \Big| \hat{W}^{(m)} \right] \approx \frac{1}{d} \sum_{i=1}^d \left( P(\hat{S}_N^{(m,i)}) \frac{\hat{S}_N^{(m,i)} - \mu_{\hat{W}^{(m)}}}{\sigma_{\hat{W}^{(m)}}^2} \right) \\ \mathbb{E}_{\Delta W_N} \left[ P(\hat{S}_N) \frac{\partial(\log p)}{\partial \sigma_{\hat{W}}} \Big| \hat{W}^{(m)} \right] \approx \frac{1}{d} \sum_{i=1}^d P(\hat{S}_N^{(m,i)}) \left( -\frac{1}{\sigma_{\hat{W}^{(m)}}} + \frac{(\hat{S}_N^{(m,i)} - \mu_{\hat{W}^{(m)}})^2}{\sigma_{\hat{W}^{(m)}}^3} \right) \end{array} \right.$$

In a multilevel setting: at the fine level we can use (23) directly. At the coarse level, for the same reasons as in section 2.3, we reuse the fine brownian increments to get efficient estimators. We take

$$\left\{ \begin{array}{l} \hat{W}_c = (\Delta W_1 + \Delta W_2, \dots, \Delta W_{N_f-3} + \Delta W_{N_f-2}) \\ (\Delta W_{N_c}^{(i)})_{i \in [1,d]} = (\Delta W_{N_f-1} + (\Delta W_{N_f}^{(i)}))_{i \in [1,d]} \end{array} \right. \quad (24)$$

We use the chain rule to verify that condition (5) is verified on the last coarse step. With

the notations of equation (19) we derive the following estimators

$$\begin{aligned}
\mathbb{E}_{\Delta W_{N_c}} \left[ P(\hat{S}_{N_c}) \frac{\partial(\log p_c)}{\partial \mu_{\hat{W}_c}} | \hat{W}_c^{(m)} \right] &= \mathbb{E} \left[ \mathbb{E} \left[ P(\hat{S}_{N_c}) \frac{\partial(\log p_c)}{\partial \mu_{\hat{W}_c}} | \hat{W}_c^{(m)}, \Delta W_{N_f-1} \right] | \hat{W}_c^{(m)} \right] \\
&\approx \frac{1}{d} \sum_{i=1}^d \left( P(\hat{S}_{N_c}^{(m,i)}) \frac{\hat{S}_{N_c}^{(m,i)} - \mu_{\hat{W}_c^{(m)}}}{\sigma_{\hat{W}_c^{(m)}}^2} \right) \\
\mathbb{E}_{\Delta W_{N_c}} \left[ P(\hat{S}_{N_c}) \frac{\partial(\log p)}{\partial \sigma_{\hat{W}_c}} | \hat{W}_c^{(m)} \right] &= \mathbb{E} \left[ \mathbb{E} \left[ P(\hat{S}_{N_c}) \frac{\partial(\log p)}{\partial \sigma_{\hat{W}_c}} | \hat{W}_c^{(m)}, \Delta W_{N_f-1} \right] | \hat{W}_c^{(m)} \right] \\
&\approx \frac{1}{d} \sum_{i=1}^d P(\hat{S}_{N_c}^{(m,i)}) \left( -\frac{1}{\sigma_{\hat{W}_c^{(m)}}} + \frac{(\hat{S}_{N_c}^{(m,i)} - \mu_{\hat{W}_c^{(m)}})^2}{\sigma_{\hat{W}_c^{(m)}}^3} \right)
\end{aligned} \tag{25}$$

### Estimated complexity

Our numerical experiments show the following convergence rates for  $d = 10$ :

Figure 4: Vibrato Monte Carlo, European call :  $\mathbb{V}(\hat{Y}_l)(l)$

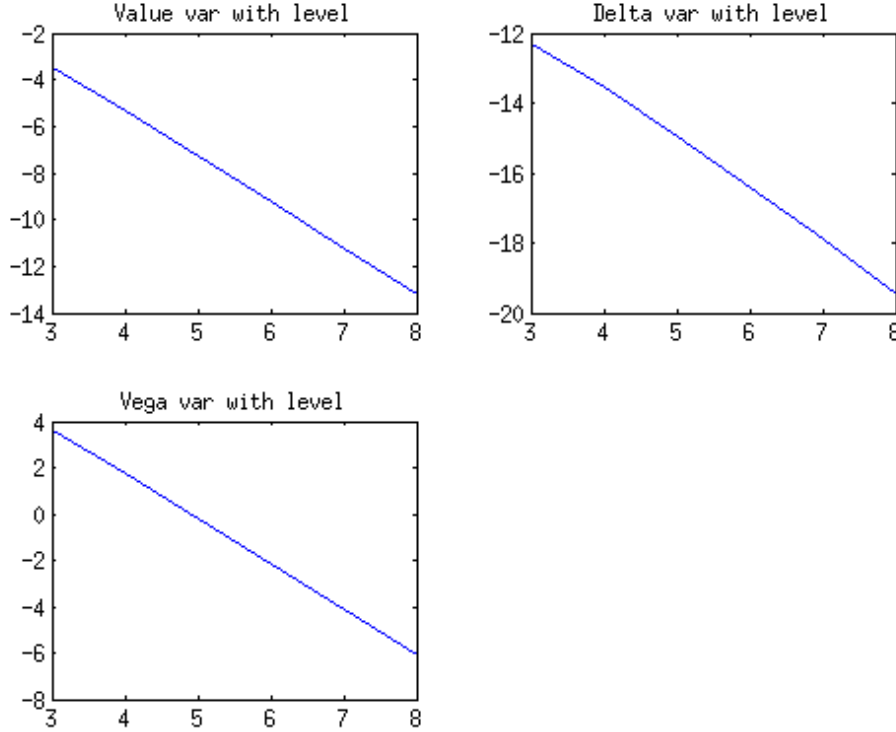


Table 4: Vibrato Monte Carlo, European call : estimated complexity

Estimator	$\beta$	MLMC Complexity
Value	$\approx 2.0$	$O(\varepsilon^{-2})$
Delta	$\approx 1.5$	$O(\varepsilon^{-2})$
Vega	$\approx 2.0$	$O(\varepsilon^{-2})$

As in section 2.3, this is an approximation of the conditional expectation technique, and so getting the same convergence rates was expected.

### 3 European digital call

The European digital call's payoff is  $P = \mathbf{1}_{S_T > K}$ . The discontinuity of the payoff makes the computation of Greeks more challenging. We cannot apply pathwise sensitivities, and so we use conditional expectations or Vibrato Monte Carlo.

#### 3.1 Pathwise sensitivities and conditional expectations

With the same notation as in section 2.2 we compute the conditional expectations of the digital call's payoff.

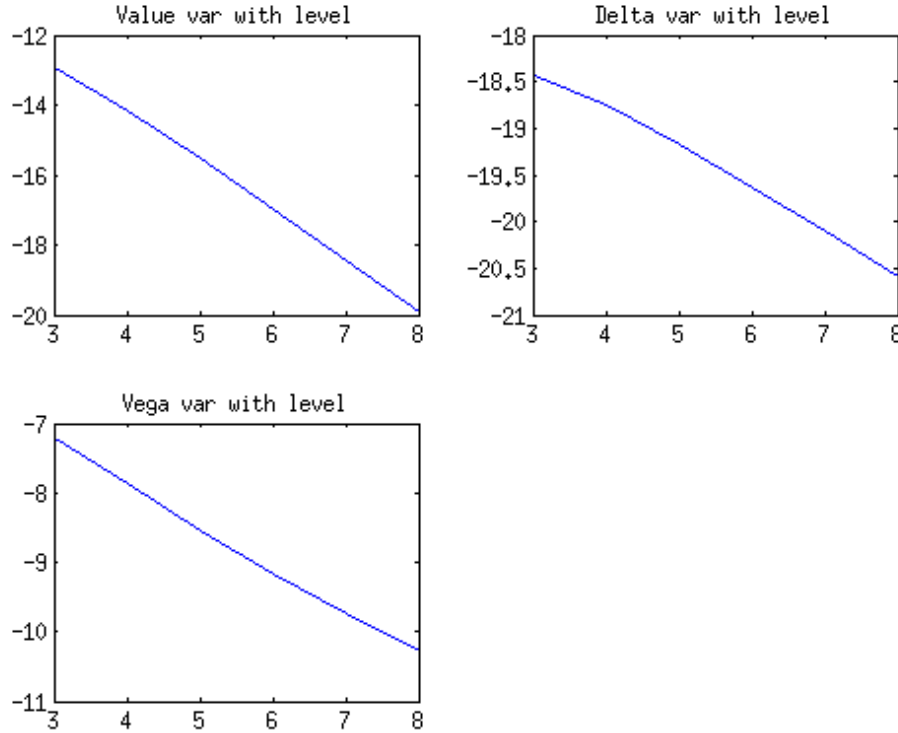
$$\mathbb{E}(P(\hat{S}_{N_f})|\hat{W}) = \Phi\left(\frac{\alpha - K}{\beta}\right) \quad \mathbb{E}(P(\hat{S}_{N_c})|\hat{W}_c, \Delta W_{N_f-1}) = \Phi\left(\frac{\alpha_c - K}{\beta_c}\right)$$

The simulations give figure 5 and table 5.

Table 5: Pathwise sensitivities and conditional expectations, digital call : estimated complexity

Estimator	$\beta$	MLMC Complexity
Value	$\approx 1.4$	$O(\varepsilon^{-2})$
Delta	$\approx 0.5$	$O(\varepsilon^{-2.5})$
Vega	$\approx 0.6$	$O(\varepsilon^{-2.4})$

Figure 5: Pathwise sensitivities and conditional expectations,  
digital call:  $\mathbb{V}(\hat{Y}_l)(l)$



### 3.2 Vibrato Monte Carlo

The Vibrato technique can be applied in the same way as with the European call. We get figure 6 and table 6.

Table 6: Vibrato Monte Carlo, digital call : estimated complexity

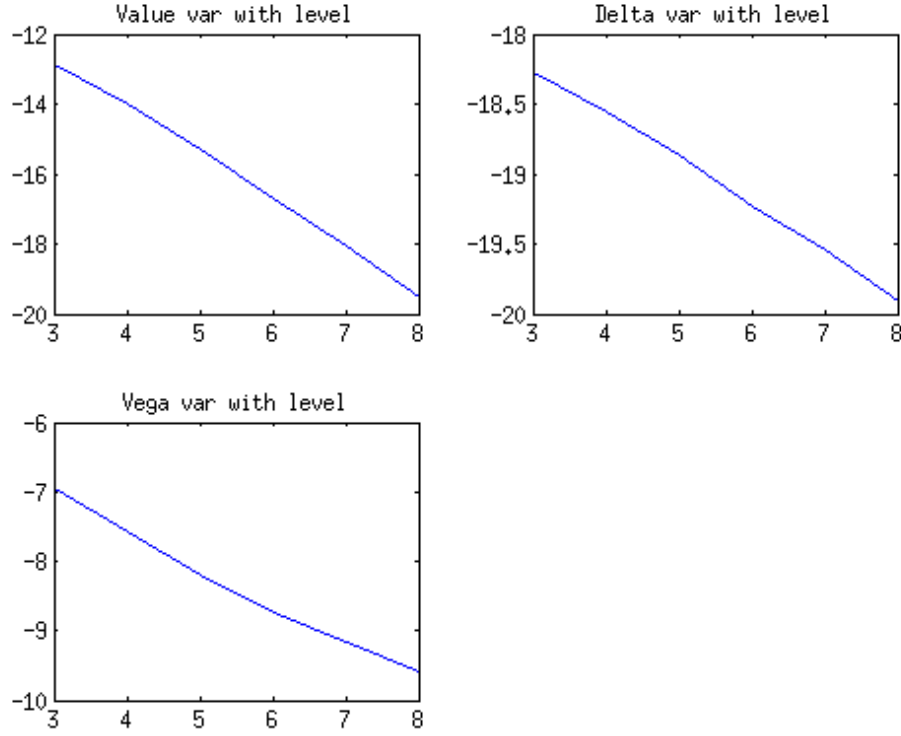
Estimator	$\beta$	MLMC Complexity
Value	$\approx 1.3$	$O(\varepsilon^{-2})$
Delta	$\approx 0.3$	$O(\varepsilon^{-2.7})$
Vega	$\approx 0.5$	$O(\varepsilon^{-2.5})$

### 3.3 Analysis

The analysis presented in section 2.2 explains why we expected  $\beta = 3/2$  for the value's estimator.

A fraction  $O(\sqrt{h})$  of all paths arrive in the area around the payoff where  $(\partial \mathbb{E}(P(\hat{S}_N)|\hat{W})/\partial \hat{S}_{N-1})$  is not close to 0 ; there its derivative is  $O(h_l^{-1})$  and we have  $|\hat{S}_{N_f} - \hat{S}_{N_c}| = O(h_l)$ . For these paths, we thus have  $O(1)$  difference between the fine and coarse Greeks' estimates. This explains the experimental  $\beta \approx 1/2$ .

Figure 6: Vibrato Monte Carlo, digital call :  $\mathbb{V}(\hat{Y}_l)(l)$



## 4 European lookback call

The lookback call's value depends on the values that the asset takes before expiry. Its payoff is  $P(T) = (S_T - \min_{t \in [0, T]}(S_t))$ .

As explained in [2], the natural discretisation  $\hat{P} = (\hat{S}_N - \min_n \hat{S}_n)$  is not satisfactory. To regain good convergence rates, we approximate the behaviour within each fine timestep  $[t_n, t_{n+1}]$  of width  $h_f$  as a simple Brownian motion with constant drift  $a_n^f$  and volatility  $b_n^f$  conditional on the simulated values  $\hat{S}_n^f$  and  $\hat{S}_{n+1}^f$ . As shown in [5] we can then simulate the local minimum

$$\hat{S}_{n, \min}^f = \frac{1}{2} \left( \hat{S}_n^f + \hat{S}_{n+1}^f - \sqrt{(\hat{S}_{n+1}^f - \hat{S}_n^f)^2 - 2(b_n^f)^2 h_f \log U_n} \right) \quad (26)$$

with  $U_n$  a uniform random variable on  $[0, 1]$ . We define the fine level's payoff this way choosing  $b_n^f = b(\hat{S}_n^f, t_n)$  and considering the minimum over all timesteps to get the global minimum of the path.

At the coarse level we still consider a simple Brownian motion on each timestep of width  $h_c = 2h_f$ . To get high strong convergence rates, we reuse the fine increments by defining a midpoint value for each step

$$\hat{S}_{n+1/2}^c = \frac{1}{2} (\hat{S}_n^c + \hat{S}_{n+1}^c - b_n^c (\Delta W_{n+1} - \Delta W_{n+1/2})) \quad (27)$$

Where  $(\Delta W_{n+1} - \Delta W_{n+1/2})$  is the difference of the corresponding fine Brownian increments on  $[t_{n+1/2}, t_{n+1}]$  and  $[t_n, t_{n+1/2}]$ . Conditional on this value, we then define the minimum over the whole step as the minimum of the minimum over each half step, that is

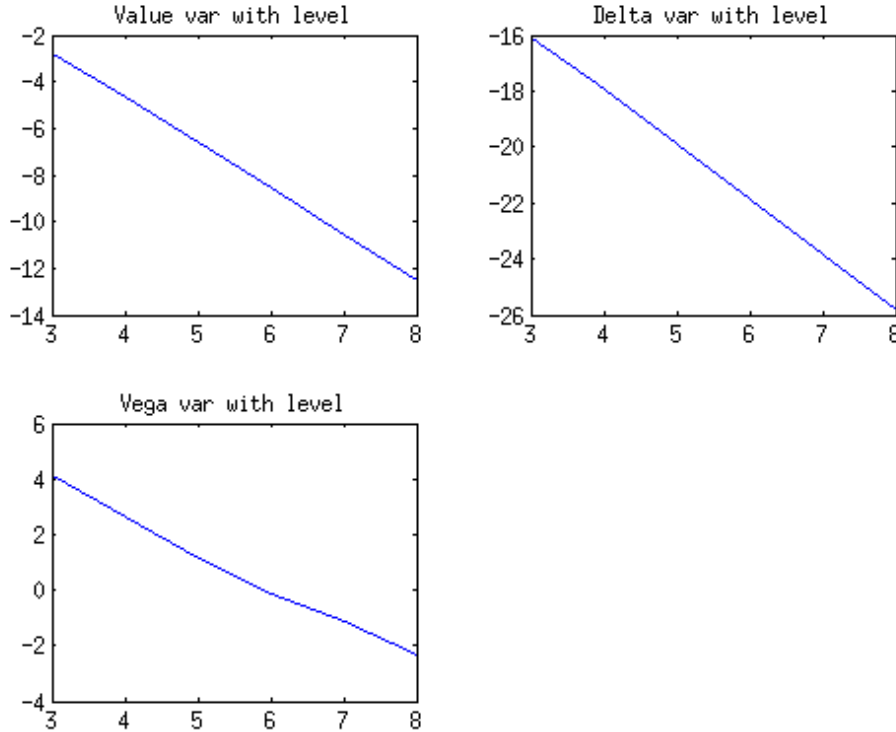
$$\hat{S}_{n,min}^c = \min \left[ \frac{1}{2} \left( \hat{S}_n^c + \hat{S}_{n+1/2}^c - \sqrt{\left( \hat{S}_{n+1/2}^c - \hat{S}_n^c \right)^2 - (b_n^c)^2 h_c \log U_{1,n}} \right), \right. \\ \left. \frac{1}{2} \left( \hat{S}_{n+1/2}^c + \hat{S}_{n+1}^c - \sqrt{\left( \hat{S}_{n+1}^c - \hat{S}_{n+1/2}^c \right)^2 - (b_n^c)^2 h_c \log U_{2,n}} \right) \right] \quad (28)$$

where  $U_{1,n}$  and  $U_{2,n}$  are the values we sampled to compute the minima of the corresponding timesteps at the fine level. Once again we use the tower property to check that condition (5) is verified and that this coarse-level estimator is adequate.

## 4.1 pathwise sensitivities

Using the treatment described above, we can then apply straightforward pathwise sensitivities to compute the multilevel estimator. This gives the following results:

Figure 7: Pathwise sensitivities, lookback call :  $\nabla(\hat{Y}_l)(l)$



Giles has proved that for the value's estimator,  $\beta = 2 - \eta$  for all  $\eta > 0$ . In the Black & Scholes model, we can prove that  $V = S_0 \delta$ . We therefore expected  $\beta \approx 2$  for  $\delta$  too. The strong convergence speed of  $v$ 's estimator cannot be derived that easily and will be analysed in our future research.

Table 7: Pathwise sensitivities, lookback call : estimated complexity

Estimator	$\beta$	MLMC Complexity
Value	$\approx 1.9$	$O(\varepsilon^{-2})$
Delta	$\approx 1.9$	$O(\varepsilon^{-2})$
Vega	$\approx 1.3$	$O(\varepsilon^{-2})$

## 4.2 Conditional Expectations, path splitting or Vibrato Monte Carlo

Unlike the regular call option, the payoff of the lookback call is perfectly smooth and so therefore there is no benefit from using conditional expectations and associated methods.

## 5 European barrier call

Barrier options are contracts which are activated or deactivated when the underlying asset  $S$  reaches a certain barrier value  $B$ . We consider here the down-and-out call for which the payoff can be written as

$$P = (S_T - K)^+ \mathbf{1}_{\min_{t \in [0, T]} (S_t) > K} \quad (29)$$

Both the naive estimators and the approach used with the lookback call are unsatisfactory here: the discontinuity induced by the barrier results in a higher variance than before. Therefore we use the approach developed in [2] where we compute the probability  $p_n$  that the minimum of the interpolant crosses the barrier within each timestep. This gives the conditional expectation of the payoff conditional on the Brownian increments of the fine path:

$$\hat{P}^f = (\hat{S}_{N_f}^f - K)^+ \prod_{n=0}^{N_f-1} (1 - \hat{p}_n^f) \quad (30)$$

with

$$\hat{p}_n^f = \exp \left( \frac{-2(\hat{S}_n^f - B)^+ (\hat{S}_{n+1}^f - B)^+}{(b_n^f)^2 h_f} \right)$$

At the coarse level we define the payoff similarly: we first simulate a midpoint value  $\hat{S}_{n+1/2}^c$  as before and then define  $\hat{p}_n^c$  the probability of not hitting  $B$  in  $[t_n, t_{n+1}]$ , that is the probability of not hitting  $B$  in  $[t_n, t_{n+1/2}]$  and  $[t_{n+1/2}, t_{n+1}]$ . Thus

$$\hat{P}^c = (\hat{S}_{N_c}^c - K)^+ \prod_{n=0}^{N_c-1} (1 - \hat{p}_n^c) = (\hat{S}_{N_c}^c - K)^+ \prod_{n=0}^{N_c-1} ((1 - \hat{p}_{n,1})(1 - \hat{p}_{n,2})) \quad (31)$$

with

$$\begin{cases} \hat{p}_{n,1} = \exp \left( \frac{-2(\hat{S}_n^c - B)^+ (\hat{S}_{n+1/2}^c - B)^+}{(b_n^c)^2 h_f} \right) \\ \hat{p}_{n,2} = \exp \left( \frac{-2(\hat{S}_{n+1/2}^c - B)^+ (\hat{S}_{n+1}^c - B)^+}{(b_n^c)^2 h_f} \right) \end{cases}$$



## 5.1 Pathwise sensitivities

The multilevel estimators  $\hat{Y}_l = (\hat{P}^f)^{(l)} - (\hat{P}^c)^{(l-1)}$  are Lipschitz with respect to all  $(\hat{S}_n^f)_{n=1\dots N_f}$  and  $(\hat{S}_n^c)_{n=1\dots N_c}$ , so we can use pathwise sensitivities to compute the Greeks. Our numerical simulations give

Figure 8: Pathwise sensitivities, barrier call :  $\mathbb{V}(\hat{Y}_l)(l)$

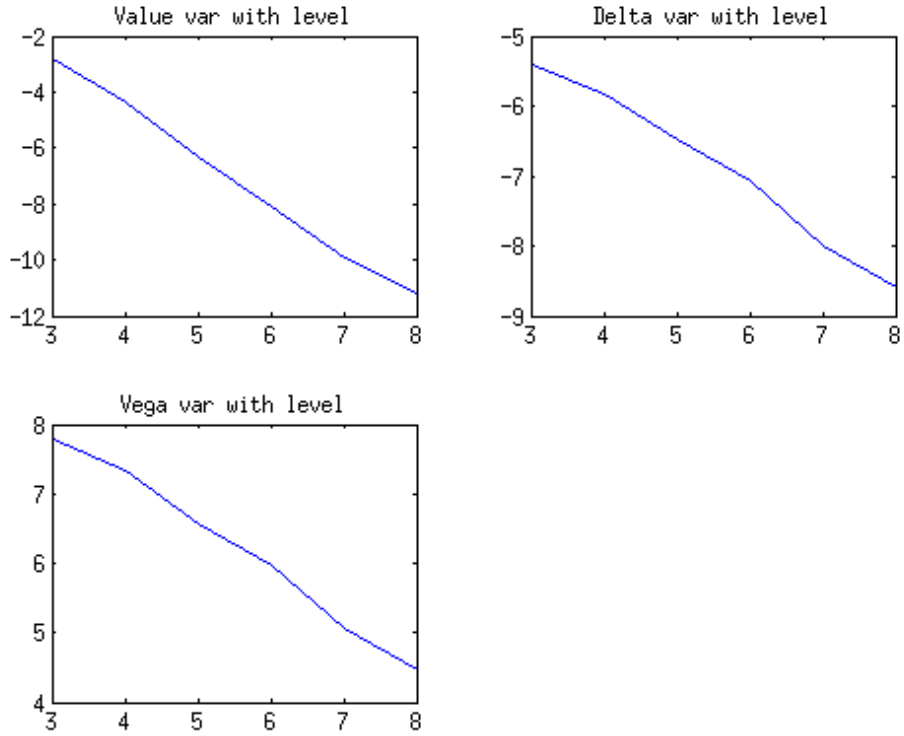


Table 8: Pathwise sensitivities, barrier call : estimated complexity

Estimator	$\beta$	MLMC Complexity
Value	$\approx 1.6$	$O(\varepsilon^{-2})$
Delta	$\approx 0.6$	$O(\varepsilon^{-2.4})$
Vega	$\approx 0.6$	$O(\varepsilon^{-2.4})$

Giles proved  $\beta = \frac{3}{2} - \eta$  ( $\eta > 0$ ) for the value's estimator. We are currently working on a numerical analysis supporting the observed convergence rates for the Greeks.

## 5.2 Conditional Expectations

The low convergence rates observed in the previous section come from from both the discontinuity at the barrier and from the lack of smoothness of the call around  $K$ . To address the latter, we can use the techniques described in section 1. Since path splitting and Vibrato Monte Carlo offer rates that are at best equal to those of conditional expectations, we implement conditional expectations to see the maximum benefits we can get.

Computing conditional expectations is slightly trickier than in section 2. We must indeed take into account the probability that the path will hit the barrier  $B$  during the final timestep. Reusing the notations of part 2.2 and defining

$$\begin{cases} \tilde{\alpha} = 2B + (-1 + rh) \hat{S}_{N-1}(\hat{W}) \\ L = \max(K, B) \\ D(\sigma, \hat{S}_{N-1}) = \exp\left(\frac{2r(B - \hat{S}_{N-1})}{\sigma^2 \hat{S}_{N-1}}\right) \end{cases} \quad (32)$$

at the fine level we get

$$\begin{aligned} \mathbb{E}(P(\hat{S}_N | \hat{W})) &= (\alpha - K) \Phi\left(\frac{\alpha - L}{\beta}\right) + \frac{\beta}{\sqrt{2\pi}} \exp\left(-\frac{(L - \alpha)^2}{2\beta^2}\right) \\ &\quad - D(\sigma, \hat{S}_{N-1}) \left[ (\tilde{\alpha} - K) \Phi\left(\frac{\tilde{\alpha} - L}{\beta}\right) + \frac{\beta}{\sqrt{2\pi}} \exp\left(-\frac{(L - \tilde{\alpha})^2}{2\beta^2}\right) \right] \end{aligned} \quad (33)$$

As before we then adapt this formula to the coarse level to compute

$$\mathbb{E}(P(\hat{S}_{N_c}) | \hat{W}_c) = \mathbb{E}(\mathbb{E}(P(\hat{S}_{N_c}) | \hat{W}_c, \Delta W_{N_f-1}) | \hat{W}_c) \quad (34)$$

Doing so actually leads to long impractical formulae, especially when computing the Greeks. The idea of the conditional expectation technique is to smoothen the payoff. We can quickly estimate the method's maximum benefits by replacing the true payoff by a smooth Lipschitz approximation: this introduces a bias but also eliminates all the problems due to the lack of regularity around the strike.

For example we can replace the payoff  $P = (S_T - K)^+ \mathbf{1}_{\min_{t \in [0, T]} (S_t) > K}$  by the smooth approximation

$$\tilde{P} = \left( \frac{\beta}{\sqrt{2\pi}} \exp\left(-\frac{(K - S_T)^2}{2\beta^2}\right) + (S_T - K) \Phi\left(\frac{S_T - K}{\beta}\right) \right) \mathbf{1}_{\min_{t \in [0, T]} (S_t) > K} \quad (35)$$

where  $\beta = \sigma \sqrt{h^*} S_T$  for some arbitrary  $h^*$  that controls the width of the smoothing. For example we take  $h^* = 1/64$  and we obtain the following results:

We see in figure 9 and table 9 that the maximum benefits of these techniques are only marginal. The barrier appears to be responsible for most of the variance of the multilevel estimators.

Figure 9: Pathwise sensitivities and payoff smoothing,  
barrier call :  $\mathbb{V}(\hat{Y}_l)(l)$

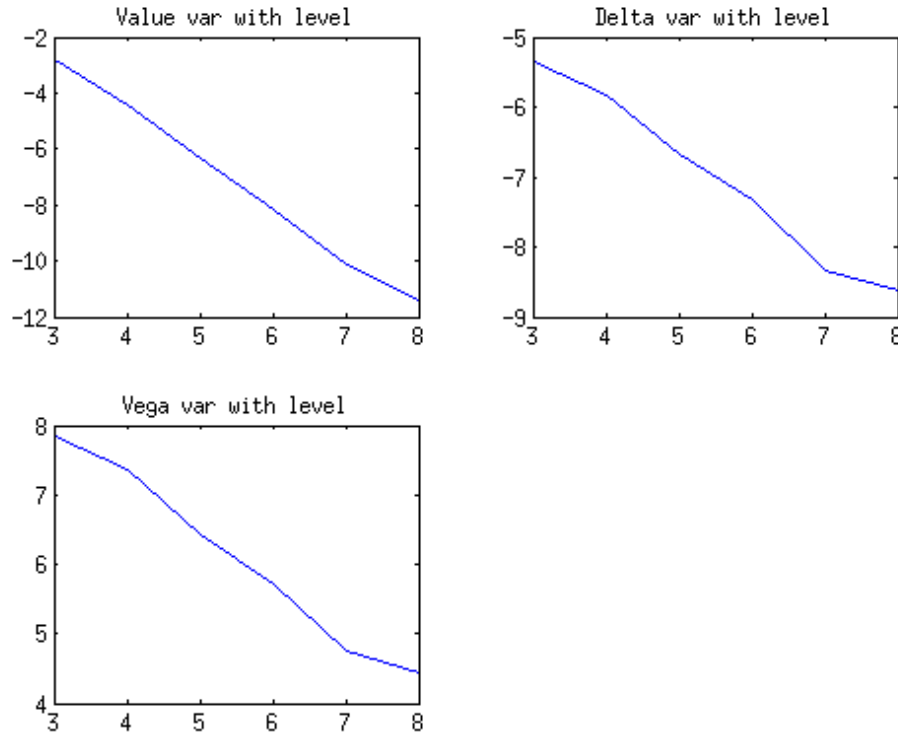


Table 9: Pathwise sensitivities and payoff smoothing,  
barrier call : estimated complexity

Estimator	$\beta$	MLMC Complexity
Value	$\approx 1.7$	$O(\varepsilon^{-2})$
Delta	$\approx 0.7$	$O(\varepsilon^{-2.3})$
Vega	$\approx 0.7$	$O(\varepsilon^{-2.3})$

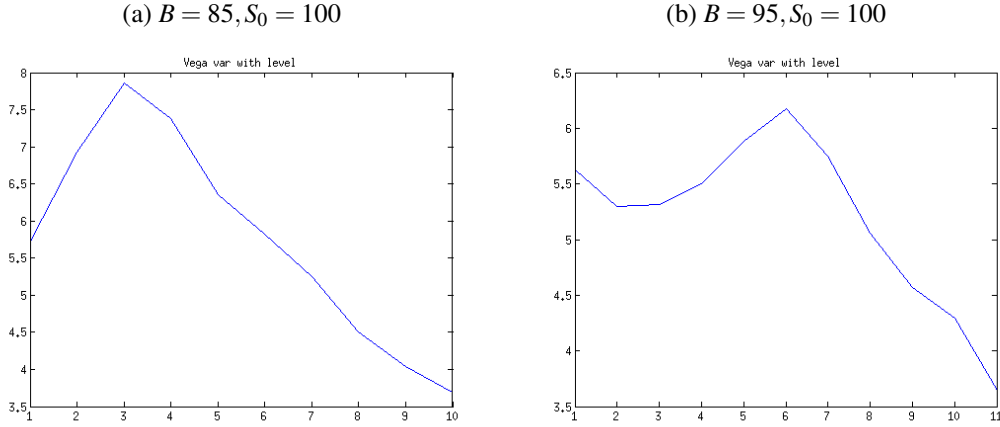
### 5.3 Non-constant timestepping

As illustrated in figure 10, the level at which  $\mathbb{V}(\hat{Y}_l)$  reaches its asymptotic convergence speed depends on the value of  $B$ . When  $B$  is far from  $S_0$ , the regime appears quickly (figure 10a), when  $B$  gets closer to  $S_0$ , it takes longer (figure 10b). Practically this can be a problem when  $B \approx S_0$  as the simulations may not reach the very fine levels at which the complexity analysis based on the asymptotic value of  $\beta$  is relevant.

In the case represented in figure 10b the variance first increases before eventually converging towards 0. This illustrates the fact that in some cases it may be interesting to start the multilevel algorithm at a level  $l > 0$  where problems related specifically to coarseness do not appear and where the variance is low enough for our application (it must be at least lower than than the variance of the equivalent monolevel estimator).

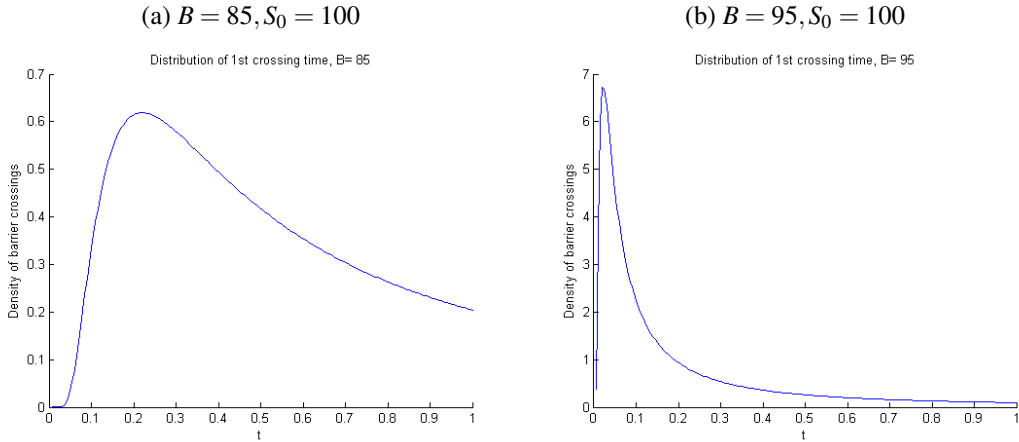
The variance's bad behaviour is related to the distribution of paths leaking out of the

Figure 10: Pathwise sensitivities (Vega), barrier call :  $\mathbb{V}(\hat{Y}_l)(l)$



barrier over time: we plot in figure 11 the density of first barrier crossings on the time interval  $[0, T]$ . We can show analytically that the width of the observed “peak” of the

Figure 11: Density of first barrier crossings on  $[0, 1]$



crossing density function is

$$\tau = O\left(\frac{\log(S_0/B)}{\sigma}\right)^2 \quad (36)$$

This means that as  $B$  tends to  $S_0$ , almost all paths going “down and out” do so in a very short time interval  $[0, \tau]$ . On timesteps  $[t_n, t_{n+1}]$  outside of this interval, most paths are far away from the barrier and both  $\frac{\partial \hat{p}_n}{\partial \theta}$  and  $\frac{\partial p_n}{\partial \theta}$  are close to 0 and hardly contribute to the variance of the multilevel estimator.

Morally the problem at the low levels is that the timesteps are much too large compared to the characteristic time  $\tau$ : the interval that is responsible for most of  $\mathbb{V}(\hat{Y}_l)$  is covered by only one step as long as  $h_l \geq \tau$ .

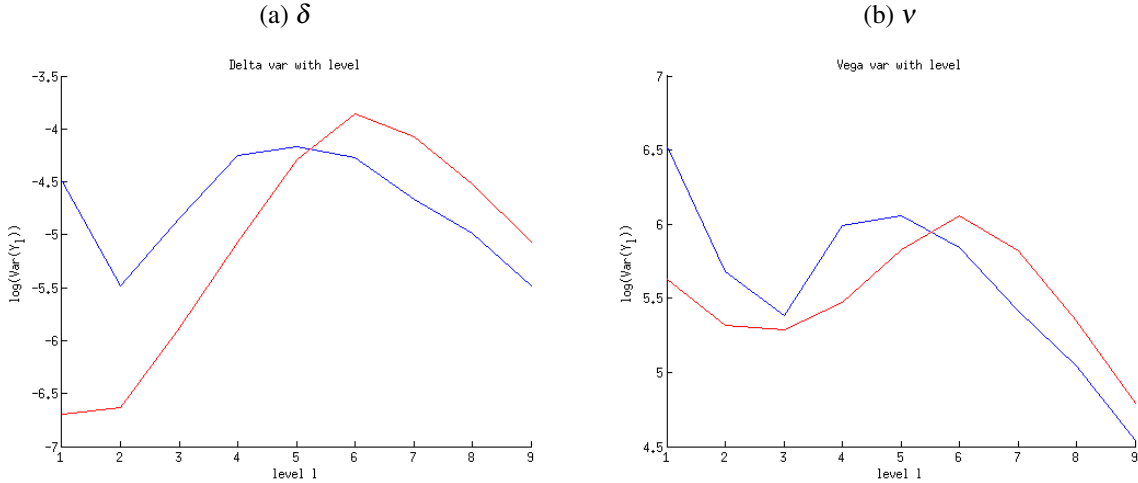
We hope to address this issue with adapted timestepping. Instead of taking constant timesteps of width  $h = T/N$  we use power timesteps to refine the discretisation in the time

interval  $[0, \tau]$ . We write  $t = u^\gamma$  and then split  $[0, T^{(1/\gamma)}]$  into equal steps of  $u$ . We want more steps in  $[0, \tau]$ . Taking half of all timesteps in  $[0, \tau]$  means that we must choose  $\gamma$  such that  $u = 1/2$  corresponds to  $\tau$ , that is

$$\left(\frac{1}{2}\right)^\gamma = \left(\frac{\log(S_0/B)}{\sigma}\right)^2 \Rightarrow \gamma = \frac{2}{\log 2} [\log \sigma - \log(\log(S_0/B))] \quad (37)$$

Figure 12 shows for  $B = 95$  the evolution of  $\mathbb{V}(\hat{Y}_l)$  for  $\delta$  (fig. 12a) and for  $v$  (fig. 12b) with constant timesteps (red) and with power timesteps (blue). We see that with power timesteps, the variance is lower at fine levels and reaches its asymptotic convergence speed faster than before. Nevertheless we note that these benefits may be practically cancelled by the higher variance of the method at the coarsest levels, especially if we need rough estimates and stay at very low levels.

Figure 12:  $\mathbb{V}(\hat{Y}_l)$ ,  $B = 95$ , power (blue) and constant (red) steps



## Conclusion and future work

In this paper we have shown for a range of cases how multilevel techniques could be used to reduce the computational complexity of Monte Carlo Greeks.

Smoothing a Lipschitz payoff with conditional expectations reduces the complexity down to  $O(\varepsilon^{-2})$ . From this technique we derive the Path splitting and Vibrato methods: they offer the same efficiency and avoid intricate integral computations. Payoff smoothing and Vibrato also enable us to extend the computation of Greeks to discontinuous payoffs where the pathwise sensitivity approach is not applicable. Numerical evidence shows that with well-constructed estimators these techniques provide computational savings even with exotic payoffs.

So far we have mostly relied on numerical estimates of  $\beta$  to estimate the complexity of the algorithms. Our current analysis is somewhat crude ; this is why our current research now focuses on getting a rigorous numerical analysis of the algorithms' complexity.

## References

- [1] A. Asmussen and P. Glynn. *Stochastic Simulation*. Springer, New York, 2007.
- [2] M.B. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In A. Keller, S. Heinrich, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 343–358. Springer-Verlag, 2007.
- [3] M.B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3), pages 607–617, 2008.
- [4] M.B. Giles. Vibrato Monte Carlo sensitivities. In P. L'Ecuyer and A. Owen, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2008*, pages 369–382. Springer, 2009.
- [5] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, New York, 2004.
- [6] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1992.