



ESCUELA DE NEGOCIOS
Universidad Torcuato Di Tella

cif

centro de investigación
en finanzas

Documento de Trabajo 05/2010

Efficient Monte Carlo for Discrete Variance Contracts

Nicolás Merener

UTDT

and

Leonardo Vicchi

UTDT and UNSAM

Efficient Monte Carlo for Discrete Variance Contracts

NICOLAS MERENER LEONARDO VICCHI *

Business School, UTDT Center of Applied Mathematics, UNSAM
1010 Saenz Valiente, Buenos Aires, (1426), Argentina.
nmerener@utdt.edu lvicchi@unsam.edu.ar

November 24, 2010

Abstract

We develop an efficient Monte Carlo method for the valuation of a financial contract with payoff dependent on discretely realized variance. We assume a general model in which asset returns are random shocks modulated by a stochastic volatility process. Realized variance is the sum of squared daily returns, depending on the sequence of shocks to the asset and the realized path of the volatility process. The price of interest is the expected payoff, represented as a high dimensional integral over the fundamental sources of randomness. We compute it through the combination of deterministic integration over a two dimensional manifold defined by the sum of squared shocks to the asset and the path average of the modulating variance process, followed by exact conditional Monte Carlo sampling. The deterministic integration variables capture most of the variability in realized variance therefore the residual variance in our estimator is much smaller than that in standard Monte Carlo. We derive theoretical results that quantify the variance reduction achieved by the method. We test it for the Hull-White, Heston, and Double Exponential models and show that the algorithm performs significantly better than standard Monte Carlo for realistic computational budgets.

1 Introduction

We develop in this paper an efficient Monte Carlo method for the valuation of derivative contracts on realized variance. We are motivated by recent strong growth in the market for these of financial contracts. A variance derivative has a stochastic payoff that depends on the variance realized by an underlying asset over a certain time interval. Following market practice, we focus on discretely realized variance, defined as the sum of squared daily returns over N consecutive days. In absence of arbitrage the contract price equals the expected discounted payoff, which in

*We thank Jose Blanchet and attendees at ALIO INFORMS 2010 and SIAM FM 2010 for valuable comments. All remaining errors are ours.

turn is a function of a sequence of random daily returns. This dependence on a large number of random variables implies that the expectation of interest is a high dimensional integral.

Computing the expectation by a standard Monte Carlo simulation of all the shocks driving the payoff leads to a price estimator with potentially large variance. Alternatively, computing the expectation through deterministic integration is unfeasible due to the high dimensionality of the problem. We develop a numerical algorithm that combines deterministic numerical integration over a low dimensional manifold with repeated Monte Carlo simulation conditional on the values of the deterministically integrated variables. This approach generates an essentially unbiased price estimator with significantly lower variance than a straightforward Monte Carlo simulation for the same computational budget.

We assume in this paper that two independent Gaussian random vectors, Z and W , each of dimension N , are the sources of randomness behind a sample of realized variance. Actual returns are proportional to the components of Z multiplied by the square root of a modulating variance process V . We assume that V is autonomous in the sense that its evolution in time depends on its own history, but it is also sensitive to newly arrived shocks in Z and W , therefore introducing correlation between the increments of the variance process and returns of the underlying asset. We consider models in discrete time because we show, through an example, that the time discreteness of realized variance is important in the valuation of relatively short dated contracts. The class of models we consider includes, as special cases, Euler discretizations (Kloeden and Platen 1999) of popular continuous-time stochastic volatility models (Hull and White 1987), (Heston 1993), and (Hagan et al. 2002), models with heavy tailed conditional returns motivated by (Kou and Wang 2004), and autoregressive moving average dynamics.

In our general setting realized variance generally depends on the joint realization of the shocks to the asset price and the path of the variance process. This dual dependence typically rules out a closed form expression for the price of an arbitrary function of realized variance. Numerical methods are the natural tool for this problem.

We propose an algorithm that takes advantage of the fact that the value of a variance derivative arises from the stochastic nature of realized variance. This is largely driven by the

stochasticity in the sum of squared shocks to the asset and the stochasticity in the average level of the variance process over the period in which discrete variance is recorded. Accordingly, extensive computational effort is allocated to eliminating these sources of noise by deterministic integration. Specifically, we integrate deterministically over the norm of Z and over the path average of the modulating variance process V . Then sample Z and W conditional on each pair of values of the integrating variables, and compute the exact realization of the contract payoff. Finally, we average over many samples to construct our estimator.

We obtain a theoretical result that quantifies the efficiency gain in our algorithm by comparing the variance of our estimator with the variance of a standard Monte Carlo estimator. We find that the residual variance of our estimator, which is zero for constant V , can be expressed in general in terms of the distance between the modulating variance process V and a constant.

The practical implementation of our method requires the knowledge of the distributions of the conditioning variables, as well as a mechanism to conditionally sample Z and W . In order to handle this issue we replace the true average of the variance process by its linear projection on the space spanned by the components of W . A linear combination of Gaussian shocks is helpful because its distribution is normal, and because its components remain conditionally Gaussian. The algorithm has a deterministic error that arises from discrete numerical integration but that is negligibly small in our implementation, and statistical error that arises from the conditional random sampling.

We present numerical evidence of the performance of our method with realistic experiments under the (Hull and White 1987) and (Heston 1993) models and a model with Double Exponential returns motivated by (Kou and Wang 2004). We find that the variance reduction is significant when comparing the method against standard Monte Carlo for the same computational budget. We also find that gains are higher for models with low correlation between the asset and the modulating variance processes.

Deterministic integration along a particular direction is equivalent, in terms of variance reduction, to very fine *stratification* along the same direction (Chapter 4 in Glasserman 2004). Our main contribution is the identification of the relevant variables for stratification in the

pricing of variance derivatives, a quantification of the associated efficiency gains, a recipe for a concrete numerical implementation, and numerical evidence, under realistic parameters, that the method delivers significant gains under fixed computational budget.

The paper is structured as follows. Section 2 discusses the practical importance of contracts on realized variance and the relationship between this paper and the existing literature. It also shows, through an example, the importance of properly accounting for the discrete nature of the returns. Section 3 introduces the class of discrete time models with autonomous stochastic volatility. We assume that the models are driven by Gaussian shocks but we let actual returns be nonlinear transformations of the shocks, therefore possibly heavy tailed. Section 4 discusses the properties of an estimator of an expectation constructed as the combination of deterministic integration and conditional random sampling. We apply this to our problem in Section 5 by presenting and justifying our choice of conditioning variables tailored to decrease the noise of the estimator of realized variance. The detailed implementation of the algorithm is discussed in Section 6, and tested in realistic examples in Section 7. Our conclusions and plans for future research are in Section 8.

2 Motivation and Related Work

The market for financial contracts on realized variance has grown strongly in recent years. Its most popular contract is the variance swap, in which two parties agree to exchange a future payment proportional to realized variance against a fixed payment agreed upon today. The fair price of future realized variance for the SP500 index is the well known VIX index. Due to its high liquidity and ease of interpretation this index has become a widely followed indicator of uncertainty in the market in the aftermath of the financial crisis of 2008 (Whaley 2000), (Carr and Lee 2009), and (Carr and Wu 2009). After the development of the variance swap, which has a *linear* payoff in realized variance, the market has begun to trade more complex contracts. Generically, a contract on realized variance delivers a stochastic cashflow at T that is a function of the empirical variance ω , realized between 0 and T . In the absence of arbitrage, the price at inception can be represented as

$$C = B(0, T)E[G(\omega)] \tag{1}$$

where $G : \mathbb{R} \rightarrow \mathbb{R}$ is a payoff function, the expectation is taken with respect to a risk-neutral probability measure, and $B(0, T)$ is a discount factor.

For example, the valuation of a variance swap corresponds to a linear G in (1), a volatility swap is associated to a square root function G and a European call option takes $G(\omega) = \max\{\omega - k, 0\}$. The initial length of most traded contracts on realized variance is of several months.

In order to position our contribution relative to existing work on variance derivatives and Monte Carlo, we briefly comment on the literature in the field. Much work has been done in recent years on variance and volatility derivatives. As discussed in the review (Carr and Lee 2009), most of the literature assumes that the realized variance in (1) is recorded continuously in time. Notable exceptions are (Broadie and Jain 2008), (Windcliff et al. 2006) and (Carr et al. 2010). Moreover, if the asset price dynamics is assumed to have continuous paths, then realized variance depends only on the path of the instantaneous volatility process and is independent of individual shocks to the asset. Yet, in practice, realized variance is defined as the sum of squared asset returns realized discretely in time. In this paper we focus on applications in which the difference between continuously and discretely realized variance is potentially large. In our setting realized variance jointly depends on the path taken by the volatility process and by the shocks to the asset price.

The approach in (Carr et al. 2010) values variance swaps under a general time changed Levy process. This is a nonparametric approach, which is desirable from a practitioner's point of view, as it eliminates model risk, but is restricted to a specific payoff. In comparison, our numerical method can be used for an arbitrary European payoff on discretely realized variance, but it requires the specification of a particular stochastic volatility model from a conveniently wide class of models. In summary, a Monte Carlo implementation has the following desirable features.

- Discrete time simulation of the asset and a correlated variance processes can be handled

naturally by Monte Carlo.

- Monte Carlo implementations can easily handle non-Gaussian conditional returns or moving average (ARMA) dynamics.
- Analytical methods tend to be specific to the form of the payoff of the contract. Monte Carlo is very flexible in this regard.

We consider a Monte Carlo algorithm a complement to analytical techniques, recognizing that both approaches have advantages and disadvantages, and that in practice any pricing problem is likely to be solved in several different ways.

Some continuous time stochastic volatility models, including the (Heston 1993) model, can be simulated exactly (Broadie and Kaya 2006) and (Glasserman and Kim 2010). Exact simulation of discrete returns of the asset price can be used to compute discretely realized variance. These methods, however, are restricted to specific models, and their discretization rules are not driven by Gaussian shocks.

The emphasis in (Broadie and Kaya 2006) and (Glasserman and Kim 2010) is in finding a discrete simulation rule that preserves the exact law of the continuous time model. We explore a different question. We start from an already discrete model driven by Gaussian shocks, and develop an algorithm that generates essentially exact samples for the discrete model leading to an estimator with less variance than a standard Monte Carlo implementation. A special case is the Euler scheme for a model driven by Brownian motion. Its convergence properties are discussed in (Kloeden and Platen 1999). However, we are not concerned here with the convergence of an Euler scheme to a continuous time model. Our approach is that of a market participant that is aware that the Euler scheme differs in its law from the continuous time dynamics, yet consistently uses the same Euler discretization for pricing complex contracts and in the calibration to vanilla options. This user is effectively replacing the continuous time model by its discrete counterpart, which is a common practice in the financial industry. In this setting, we aim to develop a method that efficiently prices the uncertainty that arises from the contractual discreteness of realized variance.

We achieve variance reduction in the estimation of a high dimensional integral by combining deterministic integration along a low dimensional manifold with conditional Monte Carlo. The literature shows some examples of the combination of Monte Carlo and numerical integration. For example, in (Hull and White (1987)), pricing under a stochastic volatility model is implemented by simulation of the variance process and closed form pricing conditional on the path of realized volatility. This technique is called Conditional Monte Carlo and discussed in the review (Boyle et al. 1997). Our approach is different in the sense that we first perform numerical integration, and then run Monte Carlo simulation.

Related work on variance reduction for stochastic volatility models includes (Ben Ameur et al. 1999), but this is not tailored to discrete variance contracts. Variance reduction through optimal stratification for path dependent derivatives in models driven by Gaussian shocks is studied in (Glasserman et al. 1999) but with no emphasis on variance dependent payoffs. Radial stratification, similar to the deterministic integration over the norm of a vector of Gaussian shocks that we propose, is used in (Glasserman et al. 2000) in a Value-at-Risk application for portfolios of options.

2.1 Discrete vs. Continuous Variance

Variance related contracts are written, in practice, in terms of discretely realized variance with daily frequency. Continuously realized variance is popular in the academic literature as an approximation for discretely realized variance but this is a problematic assumption in some cases. A simple example shows the difference between discrete and continuous variance. Consider a continuous time lognormal model

$$\frac{dS_t}{S_t} = \sqrt{V_0}dZ_t \quad (2)$$

with Z_t a Brownian motion and constant variance parameter $V_0 = 0.09$, which corresponds to 30% instantaneous volatility for the asset price S . This is a realistic level for V_0 and it is typically chosen to match the prices of simple options such as calls and puts on S_T . We assume 252 trading days per year. In this model, the continuously realized variance over N days is

$\frac{1}{252} \int_0^N V_0 dt = V_0 N / 252$ a.s. Therefore, there is no uncertainty associated with continuously realized variance.

Discrete realized variance for the same model is the sum of a finite number of independent squared Gaussian variables, therefore distributed as a χ^2 random variable with N degrees of freedom. Its mean is $V_0 N / 252$, equal to the continuously realized variance, and its standard deviation is $V_0 \sqrt{2N} / 252$. For relatively small N , for example $N = 20$ days, this implies a mean equal to 0.0071 and a standard deviation equal to 0.0023.

Because the expected value of discrete realized variance equals the value of continuous realized variance, a *linear* payoff in realized variance has the same price under both discrete and continuous settings, consistent with (Broadie and Jain 2008) for more general models. However, for volatility swaps and variance options, which have payoffs that are *non-linear* in realized variance, the values do not agree. For example, consider an at-the-money option on realized variance, with $N = 20$ and $V_0 = 0.09$. The strike, for either discrete or realized variance, is 0.0071. The standard deviation of continuous realized variance is zero, therefore the price of an option on continuous variance is zero as well. Yet the standard deviation of discrete realized variance is 0.0023, or approximately 30% of the value of the mean. This is a very valuable option. Because variance derivatives are in practice written in terms of discrete variance this simple example shows that assuming continuous variance might severely affect the quality of pricing for contracts relatively close to expiration.

3 Models

3.1 Models driven by Gaussian shocks

In this paper we focus on discrete-time models driven by Gaussian shocks, with autonomous stochastic volatility. We assume that the evolution happens on a discrete time grid that coincides with the daily grid on which changes in the asset price are recorded for the computation of variance. The underlying asset price S follows

$$\begin{aligned}
S_{i+1} &= S_i + S_i r \Delta + \sigma(S_i) \sqrt{V_i} \sqrt{\Delta} h(Z_{i+1}) \\
V_{i+1} &= f(V_{\eta \leq i}) + \gamma(V_i) \sqrt{\Delta} [\rho h(Z_{i+1}) + \sqrt{1 - \rho^2} W_{i+1}]
\end{aligned} \tag{3}$$

under the risk neutral measure, with $i = 0, 1, \dots, N - 1$. Z_1, \dots, Z_N and W_1, \dots, W_N are the independent standard normal components of vectors $\{Z, W\}$ and Δ is the length of the time interval between consecutive observations. The history of V is $V_{\eta \leq i} \equiv \{V_0, V_1, \dots, V_i\}$. The deterministic function h is chosen to satisfy

Assumption 3.1 *We assume that*

- $E[h(Z_i)] = 0$
- $\text{Var}[h(Z_i)] = 1$
- $E[h(Z_i)^4] = \kappa$.

Conditionally Gaussian returns are generated with $h(z) = z$. In this case κ , which measures kurtosis, is equal to 3. Other choices of h distort the Gaussian shocks to generate Non-Gaussian returns in the process S , for example with heavy tails. For any choice of h it holds that

- $E[\rho h(Z_{i+1}) + \sqrt{1 - \rho^2} W_{i+1}] = 0$
- $\text{Var}[\rho h(Z_{i+1}) + \sqrt{1 - \rho^2} W_{i+1}] = 1$
- $\rho = \text{Corr} \{h(Z_{i+1}), \rho h(Z_{i+1}) + \sqrt{1 - \rho^2} W_{i+1}\}$.

The class of models defined by (3) is very wide. It includes Euler discretizations of many popular continuous time models driven by Brownian motion. It can also be taken as the starting point, without any reference to an underlying continuous model, to represent discrete models with Gaussian returns, heavy tailed conditional returns, or with auto regressive moving average features by taking $f(V_{\eta \leq i}) = \sum_{j=0}^k a_j V_{i-j}$. Some specific examples are discussed below.

3.2 Discretization of continuous time models driven by Brownian motion

Let z_t and w_t be independent Brownian Motions, and consider an underlying S following

$$\begin{aligned}
dS_t &= S_t r dt + \sigma(S_t) \sqrt{V_t} dz_t \\
dV_t &= \beta(V_t) dt + \gamma(V_t) [\rho dz_t + \sqrt{1 - \rho^2} dw_t]
\end{aligned} \tag{4}$$

under suitable technical conditions for σ, β and γ . This is a general stochastic volatility model in continuous time with autonomous stochastic volatility. Some special cases of (4) include the model in (Hull and White 1987),

$$\begin{aligned}
dS_t &= S_t r dt + S_t \sqrt{V_t} dz_t \\
dV_t &= \mu V_t dt + \sigma V_t [\rho dz_t + \sqrt{1 - \rho^2} dw_t],
\end{aligned} \tag{5}$$

the model in (Heston 1993)

$$\begin{aligned}
dS_t &= S_t r dt + S_t \sqrt{V_t} dz_t \\
dV_t &= \kappa(\theta - V_t) dt + \sigma \sqrt{V_t} [\rho dz_t + \sqrt{1 - \rho^2} dw_t],
\end{aligned} \tag{6}$$

and the SABR model (Hagan et al. 2002), which is usually expressed in terms of the dynamics of the underlying S and its instantaneous stochastic volatility σ

$$\begin{aligned}
dS_t &= S_t r + S_t^b \sigma_t dz_t \\
d\sigma_t &= \alpha \sigma_t [\rho dz_t + \sqrt{1 - \rho^2} dw_t].
\end{aligned} \tag{7}$$

The SABR model can be cast in the form of (4) by introducing $V_t \equiv \sigma_t^2$ and applying Ito's Lemma.

The Euler scheme is the simplest discretization rule. It is implemented over a deterministic time grid $t_0 < t_1 < \dots < t_N$, with $t_{i+1} - t_i = \Delta$ and leads to an approximate solution \hat{S}, \hat{V} defined computed recursively over the time grid

$$\begin{aligned}
\hat{S}_{i+1} &= \hat{S}_i + \hat{S}_i r \Delta + \sigma(\hat{S}_i) \sqrt{V_i} \sqrt{\Delta} Z_{i+1} \\
\hat{V}_{i+1} &= \hat{V}_i + \beta(\hat{V}_i) \Delta + \gamma(\hat{V}_i) \sqrt{\Delta} [\rho Z_{i+1} + \sqrt{1 - \rho^2} W_{i+1}],
\end{aligned} \tag{8}$$

with Z_1, \dots, Z_N and W_1, \dots, W_N the independent standard normal components of vectors $\{Z, W\}$. In particular, the one-step conditional return is Gaussian. The discrete time dynamics in (8) is a special case of (3) with linear h .

3.3 Models with non-Gaussian returns

We also consider discrete time models with Non-Gaussian conditional returns of S through non-linear transformations of Gaussian shocks. Let F be the desired distribution for the conditional return of S and $N(z)$ the standard normal distribution. Take h in (3) to be

$$h(z) = F^{-1}(N(z)).$$

This implies that $h^{-1}(z) = N^{-1}(F(z))$. Let Z be a standard normal random variable. It is clear that the probability of the event $\{h(Z) < a\}$ is

$$P(h(Z) < a) = P(Z < h^{-1}(a)) = P(Z < N^{-1}(F(a))) = F(a). \tag{9}$$

Therefore $h(Z) \sim F$. Moreover, (9) suggests a practical sampling methodology. For given F , in order to sample $X \sim F$, first sample $Z \sim N(0, 1)$ and then let $X = F^{-1}(N(Z))$. An example of this sampling mechanism motivated by the Double Exponential model in (Kou and Wang 2004) is shown next. We postulate that one-step conditional returns are proportional to samples from a Laplace distribution

$$\begin{aligned}
F(x) &= \frac{1}{2} e^{\sqrt{2}x} & \text{if } x \leq 0 \\
F(x) &= 1 - \frac{1}{2} e^{-\sqrt{2}x} & \text{if } x > 0
\end{aligned}$$

with inverse

$$\begin{aligned} F^{-1}(u) &= \frac{1}{\sqrt{2}} \ln(2u) \quad \text{for } u \leq 1/2 \\ F^{-1}(u) &= \frac{-1}{\sqrt{2}} \ln(2(1-u)) \quad \text{for } u > 1/2. \end{aligned} \quad (10)$$

We generate a discrete time model with Double Exponential returns and general stochastic volatility with (3) by taking $h(z) = F^{-1}(N(z))$ for F^{-1} from (10).

3.4 Realized variance

Our goal is to design an efficient numerical algorithm for the computation of the expectation of a function of discretely recorded variance. This is defined as

$$\omega \equiv \sum_{i=0}^{N-1} \left(\frac{S_{i+1} - S_i}{S_i} \right)^2, \quad (11)$$

where S_i evolves as in (3). Under this class of models (11) can be expressed as a deterministic function, RV , of the stochastic shocks $\{Z, W\}$:

$$\begin{aligned} RV(Z, W) &\equiv \sum_{i=0}^{N-1} \left(r\Delta + \frac{\sigma(S_i)}{S_i} \sqrt{V_i} \sqrt{\Delta} h(Z_{i+1}) \right)^2 \\ &= \sum_{i=0}^{N-1} \left\{ r^2 \Delta^2 + 2 \frac{\sigma(S_i)}{S_i} \sqrt{V_i} r \Delta^{3/2} h(Z_{i+1}) + \left(\frac{\sigma(S_i)}{S_i} \right)^2 V_i \Delta h(Z_{i+1})^2 \right\}, \end{aligned} \quad (12)$$

where V_i is some function of $\{Z, W\}$ specified by the choice of model and that we do not need to write explicitly. The price we want to compute is

$$C = B(0, T) E[G(RV(Z, W))] \quad (13)$$

The method that we will present later in this paper involves no approximations. But we will supplement the algorithm with some theoretical results that approximately quantify the variance reduction achieved by the method. To derive these results it will be useful to consider an approximation to RV that is based on the relative weights of the terms in (12).

For realistic parameters ($r = 0.05$, $\frac{\sigma(S_i)}{S_i} \sqrt{V_i} = 0.3$, $\Delta = 1/252$), the terms in (12) are of orders 10^{-8} , 10^{-6} , and 10^{-3} . Therefore discrete realized variance is essentially

$$\hat{RV}(Z, W) \equiv \sum_{i=0}^{N-1} \left(\frac{\sigma(S_i)}{S_i} \right)^2 V_i \Delta h(Z_{i+1})^2. \quad (14)$$

We stress, however, that the algorithm in this paper computes realized variance through the exact expression (12), therefore involves no approximation.

4 Combining Deterministic Integration with Random Sampling

In this section we introduce an algorithm to compute (13) that combines deterministic integration and Monte Carlo simulation. For proper context, we begin by considering the implementation of a simple Monte Carlo estimator for the expectation of a function of realized variance. This is straightforward (we omit the trivial discount factor): first generate M independent paths, each formed by an independent realization of the vectors Z and W , and then form an estimator as

$$\hat{C} = \frac{1}{M} \sum_{j=1}^M g(RV(Z^j, W^j)). \quad (15)$$

The fact that the estimator \hat{C} is unbiased, in the sense that $E[\hat{C}] = E[g(RV(Z, W))]$, and its variance is

$$\text{Var}[\hat{C}] = \frac{\text{Var}[g(RV(Z, W))]}{M}, \quad (16)$$

are standard properties of Monte Carlo under exact sampling.

An alternative way of computing (13) is as a deterministic integral over the $2N$ -dimensional space spanned by the Gaussian vectors Z and W . The joint distribution of independent Z and W is known in closed form, then we write

$$C = \int_{z_1, \dots, z_N, w_1, \dots, w_N} g(RV(z, w)) \prod_{j=1}^N n(z_j) n(w_j) dz_j dw_j \quad (17)$$

with $n(u)$ the univariate standard normal density. The result of (17) is the exact price, but (17) is an integral over a high dimensional space. Except for a very restrictive class of payoffs and

models, (17) can not be solved analytically. And its numerical solution as a deterministic integral suffers from the curse of dimensionality, becoming prohibitively expensive beyond dimension 3 or 4.

The method we develop in this paper lies essentially between (15) and (17) in the sense that numerical integration is performed over a few dimensions that substantially reduce the noise of the price estimator, and Monte Carlo samples are generated conditional on the deterministically integrated variables. Deterministic integration reduces estimator noise, while random sampling preserves the tractability of the computation.

Let π a vector-valued random variables with components π_1, \dots, π_p defined by the action of deterministic functions of $\mathbb{R}^{2N} \rightarrow \mathbb{R}$ on the Gaussian vectors $\{Z, W\}$. We propose to combine deterministic integration over the values of the components of π with Monte Carlo sampling conditional on π . The corresponding estimator is

$$\hat{C} = \int \frac{1}{M(u)} \sum_{j=1}^{M(u)} g(RV(Z^j, W^j)) \eta(u) du, \quad (18)$$

where η is the probability density associated to π and the vectors $\{Z^j, W^j\}$ are sampled conditional on $\pi = u$ and independently over j . The estimator (18) is unbiased. By the independence of samples the variance of \hat{C} is

$$\text{Var}[\hat{C}] = \int \text{Var}\left[\frac{1}{M(u)} \sum_{j=1}^{M(u)} g(RV(Z^j, W^j)) \eta(u) \mid \pi = u\right] du, \quad (19)$$

and then

$$\text{Var}[\hat{C}] = \int \frac{\eta(u)^2}{M(u)} \text{Var}[g(RV(Z^j, W^j)) \mid \pi = u] du. \quad (20)$$

As shown in (Glasserman 2004), it is convenient to make a proportional allocation of paths by taking $M(u) = M\eta(u)$, (which can be done because we assume that $\eta(u)$ is known). Then (20) becomes

$$\text{Var}[\hat{C}] = \frac{E[\text{Var}[g(RV(Z, W)) \mid \pi]]}{M} \quad (21)$$

for the variance of the estimator based on the combination of deterministic integration and conditional Monte Carlo.

It is informative to compare (21) with the variance of standard Monte Carlo estimator (16). By the variance decomposition formula it always holds that

$$\text{Var}[g(RV(Z, W))] = \text{Var}[E[g(RV(Z, W))|\pi]] + E[\text{Var}[g(RV(Z, W))|\pi]]. \quad (22)$$

The left side of (22) is independent of π and equal to M times (16). The two terms on the right are nonnegative and the second term coincides with M times (21). Therefore, for a total number of paths M , our Monte Carlo method with proportional allocation can never increase variance relative to the unconditional case, regardless of the specific choice of π . However, in practice we want to decrease the variance of an estimator subject to a finite computational budget. Therefore, conditioning on a bad choice of π might not be useful if the increase in computational time associated to the deterministic integration is larger than the speed gain associated to the decrease in the variance of the estimator.

These observations indicate that a successful implementation of (18) should attempt to identify π_1, \dots, π_p such that:

- The relevant densities associated to π are known analytically, allowing for deterministic numerical integration.
- The dimension of the conditioning set, p , is small enough to permit *fast* numerical integration.
- A practical method to sample W and Z conditional on π is available.
- The choice of π_1, \dots, π_p leads to small $E[\text{Var}(g(RV(Z, W))|\pi)]$, to achieve significant variance reduction.

In the next section we propose a choice of conditioning variables that satisfies these conditions.

5 The Algorithm

In this section we consider a numerical algorithm based on the combination of deterministic two-dimensional integration, performed on the norm of the shocks driving S and on the path average of the variance process, with Monte Carlo simulation conditional on the value of the integration variables. The motivation for this choice of integrating variables is (14), which already suggests that realized variance is determined, to a large extent, by the squared shocks $h(Z_{i+1})^2$, and the average of the modulating volatility process V . Before describing the implementation of the algorithm in detail we present some theoretical results that quantify its efficiency gains, relative to standard Monte Carlo simulation.

We introduce some notation. Let $\|V\|_1 \equiv \sum_{i=0}^{N-1} V_i$, $\|V\|_2^2 \equiv \sum_{i=0}^{N-1} V_i^2$ and $\|H\|_2^2 \equiv \sum_{i=0}^{N-1} h(Z_{i+1})^2$.

5.1 Monte Carlo conditional on $\|H\|_2^2$ and $\|V\|_1$

We investigate in this section the variance reduction achieved by performing Monte Carlo conditional on $\pi = \{\pi_1, \pi_2\} = \{\|H\|_2^2, \|V\|_1\}$. In order to see the reason behind this choice recall from (21) that a good choice of conditioning variables should lead to small

$$E[\text{Var}[g(RV(Z, W))|\pi]],$$

therefore, by the closeness of (12) and (14), small

$$E[\text{Var}[g(\hat{R}V(Z, W))|\pi]].$$

Consider first a special case, with linear g , linear $\sigma(S)$ in (3), and $V_i = V_0 \forall i < N$. Then (14) implies that

$$E[\text{Var}[\hat{R}V(Z, W)|\pi]] = E[\text{Var}[(\sum_{i=0}^{N-1} h(Z_{i+1})^2)(\frac{\sum_{i=0}^{N-1} V_i}{N})|\pi]] = E[\text{Var}[\pi_1 \frac{\pi_2}{N}|\pi]] = 0.$$

This choice of π is successful in eliminating variance because it contains two main sources of stochasticity in $\hat{R}V$: the sum of squared shocks in the underlying, and the path average of the

modulating variance process. Variance reduction is absolute in this case because the variance process is assumed constant. In general, this is not the case therefore variance is reduced partially. In this section we quantify this notion by deriving a result that lends support to the notion that the variability of $\hat{R}V$ is largely explained by that of the components of π .

Motivated by (22) define *residual variance* as

$$\frac{E[\text{Var}[g(\hat{R}V(Z, W))|\pi]]}{\text{Var}[g(\hat{R}V(Z, W))]} \quad (23)$$

This is the proportion of the variance of standard Monte Carlo that survives even after performing deterministic integration on some directions.

The derivation of a theoretical result quantifying variance reduction will rely on the following assumptions, which are not actually needed in the implementation of the algorithm.

Assumption 5.1 *We assume that:*

- V is independent of Z .
- The function σ in (3) is linear.
- The payoff function g is linear.

Assumption (i) states that the variance process is not only autonomous but also independent of shocks to the process S . A linear σ , as in Assumption (ii), is exactly the case in the original versions of the (Heston 1993) and (Hull and White 1987). Assumption (iii) is made for simplicity.

LEMMA 5.1 *It holds that*

$$\text{Var}[\hat{R}V] = (\kappa - 1)E[\|V\|_2^2] + \text{Var}[\|V\|_1]$$

Proof: $\text{Var}[\hat{R}V] = \text{Var}[\sum_{i=0}^{N-1} V_i h(Z_{i+1})^2]$, therefore

$$\text{Var}[\hat{R}V] = \sum_{i=0}^{N-1} \text{Var}[V_i h(Z_{i+1})^2] + \sum_{i,j=0, i \neq j}^{N-1} \text{Covar}[V_i h(Z_{i+1})^2, V_j h(Z_{j+1})^2]. \quad (24)$$

For the variance terms in (24) we have

$$\begin{aligned}\text{Var}[V_i h(Z_{i+1})^2] &= E[h(Z_{i+1})^4]E[V_i^2] - E^2[h(Z_{i+1})^2]E^2[V_i] \\ &= \kappa E[V_i^2] - E^2[V_i].\end{aligned}\tag{25}$$

For the covariance terms in (24), independence of V and Z implies that

$$\text{Covar}[V_i h(Z_{i+1})^2, V_j h(Z_{j+1})^2] = \text{Covar}[V_i, V_j].\tag{26}$$

From (24), (25) and (26) we obtain

$$\begin{aligned}\text{Var}[\hat{R}V] &= \sum_{i=0}^{N-1} \{\kappa E[V_i^2] - E^2[V_i]\} + \sum_{i,j=0, i \neq j}^{N-1} \text{Covar}[V_i, V_j] \\ &= \sum_{i=0}^{N-1} (\kappa - 1)E[V_i^2] + \sum_{i=0}^{N-1} \text{Var}[V_i] + \sum_{i,j=0, i \neq j}^{N-1} \text{Covar}[V_i, V_j] \\ &= (\kappa - 1)E[\|V\|_2^2] + \text{Var}[\|V\|_1]. \quad \square\end{aligned}\tag{27}$$

According to Lemma 5.1, the variance of $\hat{R}V$ has two parts: $\text{Var}[\|V\|_1]$ due to the variability of the variance process, and $(\kappa - 1)E[\|V\|_2^2]$, which is nonzero even for constant V , and which is due to the variability of the shocks Z . Lemma 5.1 will be used next to derive, under Assumption 5.1, an exact result on the magnitude of variance reduction obtained by conditioning on $\|H\|_2^2$ and $\|V\|_1$.

THEOREM 5.1 *It holds that*

$$E[\hat{R}V|\pi] = \|H\|_2^2 \frac{\|V\|_1}{N}.\tag{28}$$

Therefore the residual variance after conditioning on $\|H\|_2^2$ and $\|V\|_1$ is:

$$\text{Residual Variance} = 1 - \frac{\text{Var}[\|H\|_2^2 \frac{\|V\|_1}{N}]}{\text{Var}[\hat{R}V(Z, W)]} = \frac{(\kappa - 1)(E[\|V\|_2^2] - \frac{1}{N}\|V\|_1^2)}{(\kappa - 1)E[\|V\|_2^2] + \text{Var}[\|V\|_1]}.\tag{29}$$

Proof: For (28) note that Z_1, \dots, Z_N i.i.d. implies $E[h(Z_i)^2|\|H\|_2^2] = \frac{\|H\|_2^2}{N}$ for all i . Using independence of Z and V together with the law of iterated expectation we get

$$\begin{aligned}
E(\hat{R}V|\pi) &= E\left[E\left[\sum_{i=0}^{N-1} V_i h(Z_{i+1})^2 | V, \|H\|_2^2\right] \middle| \pi\right] = E\left[\sum_{i=0}^{N-1} V_i E[h(Z_{i+1})^2 | V, \|H\|_2^2] \middle| \pi\right] \\
&= E\left[\sum_{i=0}^{N-1} V_i \frac{\|H\|_2^2}{N} \middle| \pi\right] = \frac{\|H\|_2^2}{N} E\left[\sum_{i=0}^{N-1} V_i \middle| \pi\right] = \frac{\|H\|_2^2}{N} \cdot \|V\|_1.
\end{aligned}$$

For (29), the decomposition of variance (22) and the definition of residual variance (23) imply that

$$\text{Residual Variance} = 1 - \frac{\text{Var}[E[\hat{R}V|\pi]]}{\text{Var}[\hat{R}V]}.$$

By (28),

$$\text{Var}[E[\hat{R}V|\pi]] = \text{Var}\left[\|V\|_1 \frac{\|H\|_2^2}{N}\right],$$

therefore Residual Variance = $1 - \frac{\text{Var}[\|H\|_2^2 \frac{\|V\|_1}{N}]}{\text{Var}[\hat{R}V(Z,W)]}$, which proves the first equality in (29).

Continuing we have

$$\begin{aligned}
\text{Var}\left[\|V\|_1 \frac{\|H\|_2^2}{N}\right] &= \frac{1}{N^2} [E[\|V\|_1^2] E[\|H\|_2^4] - E^2[\|V\|_1] E^2[\|H\|_2^2]] \\
&= \frac{1}{N^2} [(N(N-1) + N\kappa) E[\|V\|_1^2] - N^2 E^2[\|V\|_1]] \\
&= \text{Var}[\|V\|_1] + \frac{\kappa-1}{N} E[\|V\|_1^2].
\end{aligned} \tag{30}$$

Therefore, with (27) and the first equality in (29) we get that residual variance is

$$1 - \frac{\text{Var}[\|V\|_1] + \frac{(\kappa-1)}{N} E[\|V\|_1^2]}{(\kappa-1) E[\|V\|_2^2] + \text{Var}[\|V\|_1]} = \frac{(\kappa-1)(E[\|V\|_2^2] - \frac{1}{N} E[\|V\|_1^2])}{(\kappa-1) E[\|V\|_2^2] + \text{Var}[\|V\|_1]}. \quad \square \tag{31}$$

In order to interpret Theorem 5.1 consider the random variable

$$\Upsilon \equiv \frac{\sum_{i=0}^{N-1} V_i}{N} \sum_{i=0}^{N-1} h(Z_{i+1})^2 \tag{32}$$

This is $\Upsilon = \|H\|_2^2 \frac{\|V\|_1}{N}$. It is well known that $E[\hat{R}V] = E[E[\hat{R}V|\pi]] = E[\Upsilon]$. Theorem 5.1 states in the first equality in (29) that the efficiency of the proposed method in reducing variance depends on how close the second moment of Υ is to the second moment of $\hat{R}V$.

The final expression after the second equality in (29), has a geometrical interpretation. The quantity inside the expectation in the numerator is the square of the Euclidean distance of a point $V \in \mathbb{R}^N$ to the subspace $L = \{(x_1, \dots, x_N) \in \mathbb{R}^N : x_1 = x_2 = \dots = x_N\}$. Therefore the

residual variance is zero only if the discrete time process V is a constant a.s. Moreover, by continuity (the expression is a polynomial over the coordinates of V), it is clear that the closer the process is to a constant in a pathwise sense, the greater the variance reduction.

It is also useful in practice to quantify variance reduction in terms of the distributional properties of the process V , as in the following result.

THEOREM 5.2 *Fix N . Let μ_i and σ_i^2 be the mean and variance of V_i .*

Let $\mu_{\max} = \max\{\mu_0, \dots, \mu_{N-1}\}$, $\mu_{\min} = \min\{\mu_0, \dots, \mu_{N-1}\}$, $\sigma_{\max} = \max\{\sigma_0, \dots, \sigma_{N-1}\}$. It holds that

$$\text{Residual Variance} \leq \frac{\mu_{\max}^2 - \mu_{\min}^2 + \sigma_{\max}^2}{\mu_{\max}^2 + \sigma_{\max}^2}. \quad (33)$$

Proof: From the left side of (31) residual variance is

$$1 - \frac{\text{Var}[\|V\|_1] + \frac{(\kappa-1)}{N} E[\|V\|_1^2]}{(\kappa-1)E[\|V\|_2^2] + \text{Var}[\|V\|_1]} = 1 + \frac{\frac{1+(\kappa-1)}{N} \text{Var}[\|V\|_1] - \frac{(\kappa-1)}{N} E[\|V\|_1]^2}{(\kappa-1)E[\|V\|_2^2] + \text{Var}[\|V\|_1]}. \quad (34)$$

Let $\phi \equiv \text{Var}[\|V\|_1]$. Residual variance is

$$1 - \frac{\frac{1+(\kappa-1)}{N} \phi + \frac{(\kappa-1)}{N} (\sum_{i=0}^{N-1} \mu_i)^2}{(\kappa-1) \sum_{i=0}^{N-1} (\mu_i + \sigma_i)^2 + \phi}. \quad (35)$$

Assumption 3.1 and Jensen's inequality imply that $1 \leq \kappa$, therefore (35) is decreasing in ϕ . It is then bounded from above by its value at $\phi = 0$ because ϕ is nonnegative. We have

$$\text{Residual Variance} \leq 1 - \frac{\frac{(\kappa-1)}{N} (\sum_{i=0}^{N-1} \mu_i)^2}{(\kappa-1) \sum_{i=0}^{N-1} (\mu_i + \sigma_i)^2} = 1 - \frac{\frac{1}{N} (\sum_{i=0}^{N-1} \mu_i)^2}{\sum_{i=0}^{N-1} (\mu_i + \sigma_i)^2} \quad (36)$$

therefore

$$\text{Residual Variance} \leq 1 - \frac{\mu_{\min}^2}{\mu_{\max}^2 + \sigma_{\max}^2} = \frac{\mu_{\max}^2 + \sigma_{\max}^2 - \mu_{\min}^2}{\mu_{\max}^2 + \sigma_{\max}^2}. \quad \square \quad (37)$$

Residual variance in (33) is nonzero due to the fact that the process V is not a constant. In moment terms, this is due to a trend in V , represented in (37) by $\mu_{\max}^2 - \mu_{\min}^2$, and to uncertainty in V , represented by σ_{\max}^2 .

We have shown in this section that conditioning on the norm of $h(Z)$ and the path average of V , under the assumption of independence of Z and V , leads to quantifiable and possibly significant variance reduction. However, the implementation of an algorithm based on this choice of conditioning variables requires in practice the knowledge of the distributions of the norm of $h(Z)$ and the path average of V , and a mechanism to sample $\{Z, W\}$ conditional on the integrating variables. A general model (3) does not satisfy these requirements. Moreover, we do not want to restrict ourselves to models with independent Z and V . In order to address these issues we consider an alternative set of conditioning variables.

5.2 Monte Carlo conditional on $\|H\|_2^2$ and optimal $v'W$

We propose to combine deterministic integration and Monte Carlo in an estimator of the form (18), now taking as conditioning variables $\|H\|_2^2$ and a linear combination of shocks in W with weights defined by $v \in \mathbb{R}^N$. The new conditioning variable, $v'W$, is chosen to be close to the path average of V in a suitable sense, therefore preserving a substantial part of the efficiency gains achieved by taking the path average of V as conditioning variable. Yet, by being a linear combination of shocks W , this new conditioning variable has a well known Gaussian distribution that facilitates deterministic integration. Moreover, the components of W are normally distributed even after conditioning on a linear combination of them, therefore sampling is straightforward.

In this section we allow for nonzero correlation ρ in (3), therefore the shocks Z affect the process V . For simplicity we continue to assume that σ and g are linear functions in the identification of an optimal v , but this assumptions will not be required in the implementation of the method.

The aim is to find v with $\|v\|_2 = 1$, such that by using $\pi_2 = v'W$ as conditioning variable the residual variance is minimal. Under the assumptions stated above, this is equivalent by (22) to maximizing

$$\text{Var}[E[\hat{R}V|\hat{\pi}]].$$

Stating explicitly the dependence of V on Z and W we are looking for

$$\arg \max_{\{v\}} \text{Var}[E[\sum_{i=0}^{N-1} V_i(Z, W)h(Z_{i+1})^2 \| \|H\|_2^2, v'W]]. \quad (38)$$

Solving (38) is a daunting task because, in principle, each V_i arising from (3) is a nonlinear function of Z and W . However, for relatively small N , which is indeed the setting of importance for this paper, the behavior of the coefficients of (3) as functions of V means that, for a wide class of models, V_i is approximately linear in the underlying shocks Z and W . Replacing $V_i(Z, W)$ by its linear projection on the space spanned by Z and W leads to

$$\arg \max_{\{v\}} \text{Var}[E[\sum_{i=0}^{N-1} (V_0 + b_z^i Z + b_w^i W)h(Z_{i+1})^2 \| \|H\|_2^2, v'W]],$$

where $b_z^i \equiv \nabla_z V_i(Z=0, W=0)$ and $b_w^i \equiv \nabla_w V_i(Z=0, W=0)$. By independence of Z and W we obtain

$$\arg \max_{\{v\}} \text{Var}[Y(Z) + Y(Z)E[\sum_{i=0}^{N-1} b_w^i W | v'W]],$$

for Y a generic function of Z . Recalling that $E[E[\sum_{i=0}^{N-1} b_w^i W | v'W]] = 0$, we get

$$\arg \max_{\{v\}} E[E^2[\sum_{i=0}^{N-1} b_w^i W | v'W]]. \quad (39)$$

Let $b \equiv \nabla_w \sum_{i=0}^{N-1} V_i(Z=0, W=0)$. Then (39) is equivalent to

$$\arg \max_{\{v\}} \text{Var}[E[b'W | v'W]], \quad (40)$$

and its solution is attained at $v = \frac{b}{\|b\|_2}$.

The characterization of optimal v as (40) is similar to the *optimal stratification direction* in (Glasserman et al. 1999). This analogy led us to consider an alternative optimal v defined as the solution of

$$\arg \max_{\{v\}} \text{Var}[E[b'W + \frac{1}{2}(W'H_f^i W) | v'W]] \quad (41)$$

which keeps terms up to order two in the expansion of V_i in powers of W , f is defined by $f(W) = \|V\|_1$ and H_f^i is the Hessian of f . The solution to (41) is characterized by the following proposition.

PROPOSITION 5.1 *Let $\hat{f}(W) = b'W + \frac{1}{2}(W'AW)$, for $b \in \mathbb{R}^N$ and $A \in \mathbb{R}^{N \times N}$. Then the optimal deterministic integration direction is the solution to*

$$\arg \max_{\{v\}} (b'v)^2 + \frac{1}{2}(v'Av)^2 \quad (42)$$

with $\|v\|_2 = 1$.

The proof of Proposition 5.1 is in the Appendix. There we also include an example of the steps to follow to implement a numerical solution of (42) for the Hull-White model (5). In our numerical experiments we found that including second order terms does not lead to significant variance reduction and that it has a cost in terms of overhead for the algorithm. For this reason we advocate adopting the linear approximation and to condition on $v'W$ with $v = \nabla_w \|V\|_1$ evaluated at $\{Z = 0, W = 0\}$. This is the prescription we follow in the remainder of the paper.

6 Implementation

We discuss in this section some implementation aspects for the proposed algorithm. The density of $\|H\|_2^2$, required for deterministic integration, and a mechanism to sample Z conditional on $\|H\|_2^2$ depend on the specific choice of function h in (3). In our numerical tests we take $\pi_1 = \|Z\|_2^2$, distributed as a χ^2 random variable, as conditioning variable. This coincides exactly with $\|H\|_2^2$ for models with conditionally Gaussian returns (Heston 1993) and (Hull and White 1987). In models with non-Gaussian H , conditioning on $\|Z\|_2^2$ is likely to be less effective in reducing variance but still exact as sampling mechanism. The method we propose is implemented as follows

- Compute $v = \nabla_w \|V\|_1$ evaluated at $\{Z = 0, W = 0\}$. An example of this computation is shown explicitly in the Electronic Companion

- Integrate deterministically over a two dimensional cartesian grid spanned by the value of $\|Z\|_2^2$ with χ^2 distribution of degree N and over the value of W_{opt} , a standard normal random variable. In our experiments we take 40 equally spaced steps over the normal random variable over a 4 standard deviation range and 30 steps over the χ^2 random variables. These parameters are the same for all experiments and models.
- Interpret each pair of values of $\|Z\|_2^2$ and W_{opt} as the square of the norm of the Gaussian vector Z and an independent Gaussian shock along the optimal direction v .
- Generate n samples of the vector Z , conditional on its norm, writing $Z = \|Z\|_2 \frac{\xi}{\|\xi\|_2}$, where $\xi \sim N(0, I)$. The number of paths, n , is proportional to the product of the χ^2 and Normal densities for the current values of $\|Z\|_2^2$ and W_{opt} .
- Generate n samples of W , conditional on the value of the shock along the stratification direction, using the fact that if $\xi \sim N(0, \Sigma)$ in \mathbb{R}^d with $Y = v'\xi$ stratified for some $v \in \mathbb{R}^d$, then $(\xi|Y = y) \sim N(\Sigma v y, \Sigma - \Sigma v v' \Sigma)$. Consequently, sample W as

$$W = vW_{opt} + (I - vv')\Lambda,$$

where $\Lambda \sim N(0, I)$ and $\|v\| = 1$.

- The average derivative payoff, dependent on realized variance, is computed over the n paths of $\{W, Z\}$ for each integration cell, and averaged over the domain of integration weighted by the cell probability.

The only source of bias in the method is the discretization of the deterministic integrals. There is no other approximation involved. The final estimator has statistical error due to the conditional random sampling component of the algorithm.

7 Numerical results

7.1 Hull-Wite model

In this section we call our method conditional Monte Carlo and compare it to standard Monte Carlo. The standard and conditional Monte Carlo algorithms are implemented in MATLAB.

Parameters	Function	Simple MC	Std. Err.	Cond. MC	Std. Err.
N=10, $\sigma = 0.001$	E[RV]	15.87	0.01	15.87	< 0.001
N=10, $\sigma = 0.001$	Var[RV]	50.21	0.13	50.39	< 0.001
N=10, $\sigma = 1$	E[RV]	15.86	0.01	15.87	0.01
N=10, $\sigma = 1$	Var[RV]	54.08	0.14	54.18	0.13
N=50, $\sigma = 0.001$	E[RV]	79.34	0.02	79.36	< 0.001
N=50, $\sigma = 0.001$	Var[RV]	251.52	0.52	251.85	< 0.001
N=50, $\sigma = 1$	E[RV]	79.36	0.03	79.36	0.01
N=50, $\sigma = 1$	Var[RV]	701.17	1.42	701.9	1.26

Table 1: Moment estimation under the Hull-White model

Tests are performed on a desktop PC running Windows XP Professional with an Intel Core 2 Duo CPU with 2.80 GHz and 1.87 GB of RAM. We first report results for an Euler discretization of the logarithm of S and V in the (Hull and White 1987) model (5). With this choice of variables the Euler rule for \hat{V} has no bias. A systematic difference could arise between standard Monte Carlo and conditional Monte Carlo from the discretization of the deterministic numerical integral. We first verify absence of bias in Table 1. These are long computations designed to obtain estimators with standard errors smaller than the economically relevant degree of precision (about 1 in 500 for contingent claims). All results assume $V_0 = 0.04$

We report in Table 1 the mean and standard error of standard Monte Carlo and conditional Monte Carlo estimators in a variety of experiments. We present two sets of results for $N = 10$: moments for almost zero volatility of variance ($\sigma = 0.001$) and economically large volatility of variance ($\sigma = 1$), which corresponds to a %100 annual volatility of variance. This is a very high stochastic volatility regime. The first moment is virtually identical in both cases: a higher σ widens the distribution of realized variance but does not change its expected value. Notably, from $\sigma = 0.001$ to $\sigma = 1$, the second moment only increases from 50.21 to 54.21. This is due to the fact that, for relatively small N , most of the *variance of variance* comes from the finite number of shocks in the construction of realized variance, and because in only ten days the variance process does not depart much from its initial value V_0 . Therefore RV is essentially a sum of equally weighted random Gaussian shocks, and the variance of RV is largely that already present in the case of $\sigma = 0.001$. This suggests that deterministic integration on $\|Z\|_2^2$ will be

helpful in reducing variance for small N .

For $N = 50$ the opposite situation happens. Table 1 shows that the second moment of RV is 251.52 for $\sigma = 0.001$ and 701.17 for $\sigma = 1$. A larger number of days increases the relative importance of large σ because V has more time to diffuse away from V_0 and also because, as N increases, the variance of a χ_N^2 random variable decreases relative to its mean. This suggests that deterministic integration on the path average of V will be helpful in reducing variance for larger N .

All experiments in Table 1 show that the estimators of first and second moments by standard Monte Carlo are close to those obtained by our method. There is no discernible bias.

The next set of results, in Table 2, show the variance reduction achieved by conditional Monte Carlo over simple Monte Carlo in terms of the square of the ratio of standard errors. A theoretical result (22) guarantees that, for same number of paths and proportional allocation, the variance of conditional Monte Carlo can not be larger than the variance of standard Monte Carlo. However, this result is silent about the additional computational cost required by conditional Monte Carlo. A fair comparison, in practical terms, is achieved by running both methods for the same computational budget. We choose one second, as it is representative of the needs in an actual financial industry setting. We price ATM calls on variance, with strike at $E[RV]$ and OTM calls on variance, with strike at one standard deviation above the ATM strike. In all cases we take $\sigma = 1$, which corresponds to a high volatility of volatility regime. Experiments with low σ are not a fair test, as they are too close to the computation that the conditional Monte Carlo method performs with zero variance.

We see in Table 2 that the variance reduction is significant for low values of the correlation parameter ρ , and that even for a fixed computational budget the variances in our method are smaller than those from standard Monte Carlo.

7.2 Heston model

Experiments not shown in the paper indicate that there is no economically meaningful difference in the computation of moments of the discretized (Heston 1993) model under standard and

Parameters	Payoff	Simple MC	Std. Err.	Cond. MC	Std. Err.	Var. Red.
N=10, $\rho = 0$	ATM Call	2.83	0.04	2.85	0.005	64.0
N=10, $\rho = 0$	OTM Call	0.885	0.026	0.855	0.004	42.3
N=10, $\rho = 0.25$	ATM Call	2.88	0.04	2.87	0.06	44.4
N=10, $\rho = 0.75$	ATM Call	2.83	0.04	2.87	0.015	7.1
N=50, $\rho = 0$	ATM Call	9.92	0.28	10.2	0.05	31.4
N=50, $\rho = 0$	OTM Call	2.96	0.16	3.03	0.06	7.1
N=50, $\rho = 0.25$	ATM Call	10.31	0.27	10.19	0.11	6.0
N=50, $\rho = 0.75$	ATM Call	10.12	0.28	10.11	0.26	1.2

Table 2: Variance option pricing under Hull-White model

Parameters	Payoff	Simple MC	Std. Err.	Cond. MC	Std. Err.	Var. Red.
$\kappa = 10, \sigma = 0.2$	E[RV]	42.57	0.15	42.85	0.02	56.25
$\kappa = 10, \sigma = 0.2$	Var[RV]	208.18	3.84	208.77	0.83	21.40
$\kappa = 10, \sigma = 0.2$	ATM Call	5.7	0.11	5.68	0.02	30.25
$\kappa = 10, \sigma = 0.2$	OTM Call	1.69	0.06	1.57	0.01	36.00
$\kappa = 20, \sigma = 0.001$	E[RV]	39.71	0.14	39.84	0.02	49.00
$\kappa = 20, \sigma = 0.001$	Var[RV]	158.87	2.81	159.53	0.58	23.47
$\kappa = 20, \sigma = 0.001$	ATM Call	4.97	0.09	5.00	0.011	66.94
$\kappa = 20, \sigma = 0.001$	OTM Call	1.33	0.048	1.35	0.009	28.44

Table 3: Variance option pricing under the Heston model

conditional Monte Carlo. The results in Table 3 show the variance reduction achieved for computational budgets of 1 second, and $N = 20$ days. The initial variance is $V_0 = 0.06$, significantly above the mean variance $\theta = 0.04$. We take $\rho = 0$ and consider two cases: one with high volatility of variance, roughly equal to 100% per year, and mild mean reversion, and other with almost zero volatility of variance and higher mean reversion rate. In the first case the variability of V is due to its stochasticity, and in the second case the path of V is very close to deterministic but trending, therefore far from constant.

Results in Table 3 show significant variance reduction by conditional Monte Carlo relative to naive Monte Carlo for all the experiments considered.

7.3 Hull-White with double exponential increments

Experiments not shown in the paper indicate that there is no bias in the computation of moments of the discretized (Hull and White 1987) model with conditional Laplace shocks in the underlying asset (10). The results in Table 4 show the variance reduction achieved for computational

Parameters	Payoff	Simple MC	Std. Err.	Cond. MC	Std. Err.	Var. Red.
N=20, $\eta = 1$	ATM Put	6.37	0.08	6.30	0.03	7.1
N=20, $\eta = 1$	OTM Put	1.82	0.04	1.84	0.014	8.2
N=20, $\eta = 1$	OTM Call	2.76	0.11	2.64	0.065	2.9
N=50, $\eta = 0.5$	ATM Put	12.68	0.25	12.82	0.102	6.0
N=50, $\eta = 0.5$	OTM Put	4.03	0.13	3.97	0.052	6.3
N=50, $\eta = 0.5$	OTM Call	4.33	0.25	4.43	0.15	2.8

Table 4: Variance option pricing under Hull-White model with Double Exponential returns

budgets of 1 second, and $N = 20$ days and $N = 50$ days. The initial variance is $V_0 = 0.04$. The results show that the variance reduction in this case is smaller than that achieved in the model with pure Gaussian increments. This is due to the fact that we are integrating on the norm of the Gaussian shocks, which is not the norm of the actual increments due to the transformation of Gaussian shocks into Laplace increments.

Lemma 5.1 shows that the variance of realized variance is higher in models with large kurtosis. Therefore, realized variance in this model is noisier than in the Gaussian driven (Hull and White 1987) model. This is the influence of heavy tails, which translate into rare but very important large shocks, which have a large impact in realized variance. Because realized variance is a nonnegative random variable the effect of heavy tails in the underlying has a larger effect for OTM calls than for OTM puts. The table below shows that the variance reduction is less effective in high strike calls than in low strike puts.

8 Conclusions and Future Research

We have developed an efficient numerical algorithm for the computation of the expectation of an arbitrary function of discretely realized variance. We have assumed discrete time dynamics driven by Gaussian shocks with autonomous stochastic volatility. The algorithm combines deterministic integration over a two dimensional manifold that approximately spans the norm of the shocks to the underlying asset and the path average of the modulating variance process. Conditional on these variables, or their approximations, exact discrete realized variance is randomly sampled. The combination of deterministic integration and random sampling leads to efficiency gains in terms of variance reduction. We show examples of the effectiveness of our

method through the pricing of options on realized variance. The method is exact, except for economically insignificant discretization bias in the deterministic integration. Efficiency gains are particularly significant for models with conditionally Gaussian returns and low correlation between the underlying asset price and the variance process.

A main theme in this paper is the tension between the variance reduction that is achieved by partial deterministic integration with the increase in computational cost that such integration demands. It might be advantageous to consider alternative choices of variables. For example, Theorem 5.1 indicates that $\|Z\|_2^2$ and the path average of V reduce conditional variance because the variability of their product captures a significant portion of the variability of $\hat{R}V$. This suggests a conditional Monte Carlo based on a single conditioning variable distributed as the product of a chi-square random variable and an independent Gaussian random variable, to be interpreted as the product of $\|Z\|_2^2$ and a shock along the optimal stratification direction for the path average of V . Preliminary work on this type of random variable shows that its density can be computed through an application of the Mellin transform and a series expansion. Its proper implementation, as well as a mechanism for conditional sampling of Z and W are topics for other paper. In order to optimize variance reduction for models with non-Gaussian increments, it would be interesting to explore non-Gaussian vector random variables that have an analytical distribution for the norm of the vector and easily implementable conditional sampling of its components. Spherically contoured distributions, discussed in (Glasserman 2004) are a natural starting point.

References

- Ben Ameur, H., L'Ecuyer, P., C. Lemieux. 1999. Variance reduction of Monte Carlo and randomized quasi-Monte Carlo estimators for stochastic volatility models in finance. *Proceedings of the 1999 Winter Simulation Conference*.
- Boyle, P., M. Broadie, P. Glasserman. 1997. Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control* **21** 1267–1321.
- Broadie, M., A. Jain. 2008. The effect of jumps and discrete sampling on volatility and variance swaps. *International Journal of Theoretical and Applied Finance* **11** No.8 761–797.
- Broadie, M., O.Kaya. 2006. Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research* **54** No.2 217–231.
- Carr, P., R. Lee. 2009. Volatility derivatives. *Annual Review of Financial Economics* **1**.

- Carr, P., R. Lee, L. Wu. 2010. Variance Swaps on Time-Changed Levy Processes. forthcoming in *Finance and Stochastics*.
- Carr, P., L. Wu. 2009. Variance risk premiums. *Review of Financial Studies* **22** 3 1311-1341.
- Dupire, B. 1993. Model Art. *Risk* **9** 118-120.
- Glasserman, P., P. Heidelberger, P. Shahabuddin. 1999. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance* **9**(2) 117-152.
- Glasserman, P., P. Heidelberger, P. Shahabuddin. 2000. Importance sampling and stratification for value-at-risk. in Computational Finance 1999 (Proceedings of the Sixth International Conference on Computational Finance) Y.S. Abu-Mostafa, B. LeBaron, A.W. Lo, and A.S. Weigend, eds., MIT Press, Cambridge, Mass. 2000. pp 7-24.
- Glasserman, P. 2004. Monte Carlo Methods in Financial Engineering. Springer-Verlag, New York. pp 209-228.
- Glasserman, P., K. Kim. 2010. Gamma expansion of the Heston stochastic volatility model. Forthcoming in *Finance and Stochastics*
- P. Hagan, D. Kumar, A. S. Lesniewski. Managing smile risk. *Wilmott* 84-108 September 2002.
- Heston, S.L. 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies* **6** 2 327-343.
- Hull, J., A. White. 1987. The pricing of options on assets with stochastic volatilities. *Journal of Finance* **42** (2) 281-300.
- Kloeden, P.E., E. Platen. Numerical Solution of Stochastic Differential Equations. Springer Verlag, New York, 1999. pp 305-310.
- Kou, S. G., H. Wang. 2004. Option pricing under a double exponential jump diffusion model. *Management Science* **50** 1178-1192.
- Whaley, R.E. 2000. The investor fear gauge. *The Journal of Portfolio Management* **26** 3 12-17.
- Windcliff, H., P.A. Forsyth, K.R. Vetzal. 2006. Pricing methods and hedging strategies for volatility derivatives. *Journal of Banking and Finance* **30** 409-431.

9 Appendix: Proof of Proposition 5.1.

Let $f(w) = b'w + \frac{1}{2}(w'Aw)$, for $b \in \mathbb{R}^N$ and $A \in \mathbb{R}^{N \times N}$. Then the optimal deterministic integration direction is the solution to

$$\arg \max_{\{v\}} (b'v)^2 + \frac{1}{2}(v'Av)^2 \quad (43)$$

with $\|v\|_2 = 1$.

Proof of Proposition 5.1.

The aim is to find $v \in \mathbb{R}^N$, $\|v\|_2 = 1$, that maximizes $\text{Var}[E(f(W)|v'W)]$. The variance in (40) is

$$\text{Var}[E[b'W|v'W]] + \text{Var}\left[\frac{1}{2}E[W'AW|v'W]\right] + 2\text{Covar}[E[b'W|v'W], \frac{1}{2}E[W'AW|v'W]].$$

The first term in the previous expression can be computed as follows: decompose W as a sum in the directions determined by v and $v^\perp \in \langle v \rangle^\perp$ (with $\|v^\perp\|_2 = 1$),

$$W = (v'W)v + ((v^\perp)'W)v^\perp,$$

multiply by b and take expectation conditional on $v'W$,

$$E[b'W|v'W] = (v'W)b'v + b'v^\perp E[(v^\perp)'W] = (v'W)b'v,$$

where the last equality holds because $v'W$ and $(v^\perp)'W$ are independent standard normals.

Finally, take variance to get

$$\text{Var}[E[b'W|v'W]] = (b'v)^2 \text{Var}[v'W] = (b'v)^2 \|v\|^2 = (b'v)^2.$$

For the remaining terms, note that

$$E[W'AW|v'W] = (v'W)^2 v'Av + (v^\perp)'Av^\perp,$$

using again that $v'W$ and $(v^\perp)'W$ are independent standard normals. Therefore,

$$\text{Var}\left[\frac{1}{2}E[W'AW|v'W]\right] = \frac{1}{2}(v'Av)^2,$$

and

$$\text{Covar}[E[b'W|v'W], \frac{1}{2}E[W'AW|v'W]] = \frac{1}{2}(b'v)E[(v'W)E[W'AW|v'W]] = (b'v)(v'Av)E[(v'W)^3] = 0.$$

In summary, the solution to (41) is equivalent to

$$\arg \max_{\{v\}} (b'v)^2 + \frac{1}{2}(v'Av)^2$$

under $\|v\|_2 = 1$. \square

10 Appendix: Computation of optimal integration direction for the Hull-White (5) model

We solve (43) through the Lagrange's multiplier method. For symmetric A , the derivative with respect to v_i of the function to maximize is

$$b_i(b'v) + 2(v'Av)((A_Dv)_i + (A_Sv)_i),$$

where $A_D + A_S + A'_S = A$, A_D being the diagonal matrix of the diagonal elements of A . Hence, the multiplier equations read

$$b_i(b'v) + (2(A_D - \lambda I)v)_i + (v'Av)(A_Sv)_i = 0 \quad (i = 1, 2, \dots, N). \quad (44)$$

Consider the Hull-White model (5) with $\rho = 0$ (the case of $\rho \neq 0$ follows exactly along the same lines). The Euler discretization of the logarithm of V leads to

$$\|V\|_1 = V_0 * (1 + e^{(\mu - \frac{1}{2}\sigma^2)\Delta + \sigma\sqrt{\Delta}W_1} + e^{2(\mu - \frac{1}{2}\sigma^2)\Delta + (\sigma\sqrt{\Delta})(W_1 + W_2)} + \dots + e^{(N-1)(\mu - \frac{1}{2}\sigma^2)\Delta + \sigma\sqrt{\Delta}(W_1 + \dots + W_{N-1})}),$$

which leads to

$$\partial_{w_i} \|V\|_1 = (\sigma\sqrt{\Delta}) \sum_{k=i}^{N-1} e^{k(\mu - \frac{1}{2}\sigma^2)\Delta + (\sigma\sqrt{\Delta})(W_1 + \dots + W_k)},$$

$$\nabla_i \|V\|_1(W=0) = (\sigma\sqrt{\Delta}) \left(\frac{e^{(\mu - \frac{1}{2}\sigma^2)\Delta N} - 1}{e^{(\mu - \frac{1}{2}\sigma^2)\Delta} - 1} - \frac{e^{i(\mu - \frac{1}{2}\sigma^2)\Delta} - 1}{e^{(\mu - \frac{1}{2}\sigma^2)\Delta} - 1} \right) = (\sigma\sqrt{\Delta}) \left(\frac{e^{(\mu - \frac{1}{2}\sigma^2)\Delta N} - e^{(\mu - \frac{1}{2}\sigma^2)\Delta i}}{e^{(\mu - \frac{1}{2}\sigma^2)\Delta} - 1} \right).$$

A similar computation shows that,

$$\partial_{w_i w_j} \|V\|_1 = b^2 \sum_{k=\max(i,j)}^{N-1} e^{(\mu - \frac{1}{2}\sigma^2)\Delta k + (\sigma\sqrt{\Delta})(W_1 + \dots + W_k)},$$

therefore the $\{i, j\}$ entry of the Hessian of f at $W = 0$ becomes

$$(\sigma^2 \Delta) \left(\frac{e^{(\mu - \frac{1}{2}\sigma^2)\Delta N} - e^{(\mu - \frac{1}{2}\sigma^2)\Delta \max(i,j)}}{e^{(\mu - \frac{1}{2}\sigma^2)\Delta} - 1} \right).$$

Numeric implementation becomes necessary at this stage to solve (44).