Tesis Doctoral

Sobre la valuación de derivados de varianza a través de métodos de Monte Carlo eficientes

Vicchi, Leonardo H.

2012-03-27

UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Sobre la valuación de derivados de varianza a través de métodos de Monte Carlo eficientes

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires en el área Ciencias Matemáticas

Leonardo H. Vicchi

Director de tesis: Dr. Pablo Amster
Consejero de estudios: Dr. Pablo Amster

Lugar de trabajo: Departamento de Matemática, FCEyN, UBA

Buenos Aires, Marzo de 2012

On the Pricing of Variance Derivatives through Efficient Monte Carlo Methods

In this thesis, financial contracts on realized variance are studied. Efficient Monte Carlo method for their valuation is developed under 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 squared shocks to the asset and the realized path of the volatility process. The derivative price is represented as a high dimensional integral over the fundamental sources of randomness. We identify a low dimensional manifold, defined by the sum of squared shocks to the asset and the path average of the modulating variance process, that drives the uncertainty in realized variance. We compute the contract price by combining precise integration over this low dimensional manifold, implemented as fine stratification or deterministic quadrature, with conditional Monte Carlo sampling on the remaining dimensions. Focusing computational effort on the low dimensional manifold leads to an estimator with lower variance than standard Monte Carlo. We derive, under an independence assumption, approximate theoretical results that quantify this effect for a class of nonlinear payoffs. We verify numerically that for the Hull-White and Heston models the algorithm performs significantly better than standard Monte Carlo for fixed computational budgets.

Keywords: Stochastic Volatility, Variance Derivatives, Stratification, Conditional Monte Carlo.

Sobre la Valuación de Derivados de Varianza a través de métodos de Monte Carlo eficientes

En esta tesis, se estudian contratos financieros sobre varianza realizada. Un método eficiente de Monte Carlo es desarrollado bajo un modelo general en el que los retornos del activo considerado vienen dados por cambios aleatorios modulados por un proceso de volatilidad estocástica. La varianza realizada es la suma de los cuadrados de los retornos diarios, que dependen de la secuencia de la serie de cambios en el activo y del camino realizado por el proceso de volatilidad. El precio del derivado se ve representado como una integral en un elevado número de dimensiones sobre las fuentes fundamentales de incertidumbre. Identificamos una variedad de baja dimensión, definida por la suma de los cuadrados de los cambios en el activo y la volatilidad, que son los que conducen la estocasticidad en el proceso de varianza realizada. El precio del contrato es calculado por medio de una combinación de integración deteminística sobre esta variedad de baja dimensión (implmentado a través de una estratificación precisa o cuadratura), junto con un sampleo de Monte Carlo Condicional en las restantes dimensiones. Concentrar el esfuerzo computacional en la variedad de baja dimensión conduce a un estimador con menor varianza que en un Monte Carlo estándar. Bajo un supuesto de independencia, obtenemos resultados teóricos aproximados que cuantifican este efecto para una clase de funciones de pago no lineales. Verificamos numéricamente que para los modelos de Hull-White y Heston, el algoritmo funciona significativamente mejor que un Monte Carlo tradicional dado un costo computacional fijo.

Palabras claves: Volatilidad estocástica, Derivados de varianza, Estratificación, Monte Carlo condicional.

Contents

Chapter 1

Introduction

1.1 Why variance derivatives

A central theme in finance is risk, and variance or volatility are widely accepted as the standard ways to measure it. Formally, the volatility is the annualized standard deviation of the stock's returns over the period of interest. In the early 1970's Black and Scholes made a crucial breakthrough regarding risk management: the pricing of vanilla options on a stock. However, their model had the limitation of assuming constant volatility, a fact not satisfied in real world markets. In fact, it does not even fluctuates deterministically over time but stochastically. Since then, many issues regarding volatility have been addressed and studied in several manners.

It is because of this stochasticity of volatility and the need of having an insight into its future level that investors have managed to trade volatility using as vehicles delta-hedged options. But as it is well known, this strategy shows imperfections in many aspects, such as market frictions, impossibility of continuous hedge (trading in the underlying), and lack of a continuum of option strikes. To overcome these difficulties, financial contracts whose value depend on future variance or volatility have been created. Here we study variance and volatility swaps and swaptions, motivated by recent strong growth in the market for this kind of financial contracts. The first two contracts are forward contracts on future realized variance or volatility. Both instruments provide an easy way for investors to gain exposure to the future level of variance. Variance swaps are now very actively traded on stock indices (and sometimes on individual stocks). They are also traded, but less commonly, in other asset classes such as FX. There are futures and options contracts on the CBOE VIX index which are now also very actively traded. The VIX index is the market price of a portfolio of vanilla options which (as we will show) replicates future

realized variance. The prices of vanilla options, variance swaps, VIX futures and VIX options are all closely linked - both practically and theoretically. It becomes clear then, that in order to effectively trade these contracts, we need to study them adequately.

1.2 Related work

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. Derman et al.[DDK] explain the theory and main properties of these derivatives. They derived an analytical formula for theoretical fair value in the presence of realistic volatility skews, and pointed out that volatility swaps can be replicated by dynamically trading the more straightforward variance swap. Carr and Lee [CL05] extend several known results to arbitrary functions of realized variance under zero-correlation hypothesis. In their work, no assumptions are made regarding the dynamics of volatility.

As discussed in the review by Carr and Lee [CL09], most of the literature assumes that the realized variance is recorded continuously in time. Notable exceptions are Broadie and Jain [BJ08], Windcliff et al. [WFV], Carr and Lee [CL10] and Keller-Ressel and Muhle-Karbe [KM] which preserve the fact that, in practice, realized variance is defined in terms of discrete time returns or logarithmic increments.

Because discrete and continuous expected variance are equal, a *linear* payoff in realized variance has the same price under both discrete and continuous settings, consistent with Broadie and Jain [BJ08] for more general models. However, for volatility swaps and variance options, with *nonlinear* payoffs in realized variance, the values do not agree.

The approach in Carr and Lee [CL10] 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 because it eliminates model risk, but is restricted to a specific payoff. Our numerical method can be used for an arbitrary European payoff on discretely realized variance, under a wide class of stochastic volatility models. As a disadvantage, it requires knowledge of the parameters of the model. We consider a Monte Carlo algorithm to be 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.

We achieve variance reduction in the estimation of a high dimensional expectation by integrating, or stratifying, along a low dimensional manifold followed by conditional Monte

Carlo. Our work is close in spirit to the stratification approach by Glasserman et al. [GHS] focused on path dependent derivatives, and to Glasserman et al. [GHS0], in which radial stratification, similar to the integration over the norm of a vector of Gaussian shocks that we propose, is used in a Value-at-Risk application for portfolios of options. Related work on variance reduction for stochastic volatility models includes Ben Ameur et al. [BLL], but this is not tailored to discrete variance contracts.

The literature also shows some examples of the combination of Monte Carlo and numerical integration in Conditional Monte Carlo. For example, in Hull-White [HW], 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 discussed in the review by Boyle et al. [BBG] but our approach is different in the sense that we first perform numerical integration, and then run many conditional Monte Carlo simulations.

Some continuous time stochastic volatility models, including the Heston [H] model, can be simulated exactly by techniques developed by Broadie and Kaya [BK08] and Glasserman and Kim [GK]. 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 not driven by Gaussian shocks. The emphasis in Broadie and Kaya [BK08] and Glasserman and Kim [GK] is in finding a discrete simulation rule that preserves the exact law of the continuous time model. We explore a different question. We assume a continuous time model driven by Gaussian shocks, adopt the Euler scheme with smallest bias as prescribed by Lord et al. [LKD], and develop an algorithm that generates samples for the discrete approximation leading to an estimator with less variance than a standard Monte Carlo implementation. The Euler scheme is helped by the very small time step implied by the daily recording frequency in the contract, but we are not concerned here with the exact matching of an Euler scheme to a continuous time model. Our approach is that of a market participant that is aware that an 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 use 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.

1.3 Our contribution

Our research can be divided in two parts: one more theoretical part in which we extend the treatment of [CL05] to allow for stochastic interest rates with arbitrary correlation with the underlying asset price and a second part in which numerical methods are performed.

A few papers have explored variance derivatives under stochastic interest rates. Merener [Me] has priced and hedged generalized variance swaps on forward swap rates under the assumption of perfect correlation between the underlying swap rate on which variance is recorded and the short rate used to reinvest dynamically accrued gains. Hölfert and Torne [HT] explored equity variance swaps under stochastic interest rates and derived bounds on the price under the assumption of a correlation input between the asset price and the interest rate.

We assume, however, that the instantaneous volatility process is independent from the interest rate and the asset price itself. We find that the stochasticity of interest rates has two effects: discounting and drift of S. By assuming that the correlation between the volatility process and the interest rate is independent we are able to compute a contract price that is independent of the properties of the interest rate. In particular, it is independent of its volatility, and of the correlation between the asset price and the interest rate. But even under this assumption, we find that hedging is not trivial, and highly sensitive to the volatility of the interest rate.

In the second part of the thesis, we develop an efficient Monte Carlo method for the valuation of derivative contracts on realized variance. 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 random payoff, in symbols,

$$C(t) = P(t,T)\mathbb{E}^Q(g(w)|\mathcal{F}_t),$$

where $g : \mathbb{R} \to \mathbb{R}$ is a payoff function, the expectation is taken with respect to a risk-neutral probability measure Q, and $P(t,T)$ is a discount factor. For example, the valuation of a variance swap corresponds to a linear g, 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.

We focus on realized variance ω, defined as the sum of squared daily returns over n consecutive days. Actual contracts in the market are written in terms of the very closely related sum of squared daily increments of the logarithm of the asset price, a feature

that we preserve in our numerical experiments. Each return is random, therefore the expectation of interest can be expressed as a high dimensional integral.

We assume in this paper that asset returns are generated by a random shocks modulated by an autonomous, but possibly correlated, stochastic volatility process. The class of models we consider includes, as special cases, the Hull-White [HW], Heston [H], and SABR [HKL] models among others. We focus on realized variance based on discrete time returns because we show, through an example, that this contractual property is important in the valuation of relatively short dated contracts. We approximate the returns in discrete time implied by the continuous time models through Euler discretizations [KPl] driven by Gaussian shocks. 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 payoff of realized variance. Numerical methods are the natural tool for this problem.

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 method that exploits a structural property of payoffs that depend on realized variance. We identify a two dimensional manifold, defined by the sum of squared shocks to the asset price and the average level of the variance process over the life of the contract, that largely drives the uncertainty in realized variance. We use this fact to propose a numerical algorithm that combines precise numerical integration over this low dimensional manifold, implemented as finely stratified Monte Carlo or deterministic quadrature, with repeated Monte Carlo simulation conditional on the values of the variables at the low dimensional manifold. Allocating extensive computational effort to eliminating the noise associated with the low dimensional manifold leads to a price estimator with significantly lower variance than a straightforward Monte Carlo simulation for the *same* computational budget.

Assuming that shocks to the asset price and the stochastic volatility process are independent we obtain an approximate theoretical result that quantifies the variance reduction in our algorithm for a class of nonlinear payoffs, relative to standard Monte Carlo. We find that the residual variance of our estimator, which is zero for constant stochastic volatility process, can be expressed in general in terms of the distance between the modulating variance process and a constant.

We adapt our method to handle correlation between the asset price and the stochastic variance process by conditioning on the component of the linearized average variance

process that is orthogonal to the asset price shocks. We present numerical evidence of the performance of our method, in their quadrature and stratification implementations, in experiments with realistic parameters under the Hull-White [HW] and Heston [H] models. We find that the variance reduction is significant when comparing our implementations against standard Monte Carlo for fixed computational budgets. We also find that efficiency gains are higher for models with lower correlation between the asset and the modulating variance processes.

1.4 Structure of the thesis

This thesis is structured as follows:

Chapter 2 introduces some background material and results in the literature, such as contingent claims, market completeness, arbitrage, risk-neutral valuation and Black-Scholes model.

No proof of any of these results is included as they are considered standard.

Chapter 3 reads about general theoretical properties of variance derivatives. Section 1 describes the general setting we propose for the variables involved. Section 2 makes a comparison between the two class of variances, continuous and discrete. In Section 3 variance contracts are defined and discussed. Section 4 shows a theoretical way of replicating continuously realized variance through options and forward. In Section 4 we present our original contribution, which consists in extending the main results of Carr-Lee theory ([CL05]) to a context of stochastic interest rates.

In Chapter 4 we explain all the Monte Carlo tools we implemented for our algorithm. Section 1 briefly describes the goal of Monte Carlo. Simulation efficiency and how to measure it is discussed in Section 2. In Section 3 we focus on the variance reduction techniques used: Conditional Monte Carlo and Stratification. The chapter is closed with an original characterization of the optimal stratification direction of a general quadratic form.

The last chapter, the most important of this essay, reads about numerical simulations. Section 5.1 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 5.2 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 5.3.2 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 5.3, and tested in realistic examples in Section 5.4. An original characterization of a density function involved in the algorithm is given.

Our conclusions are included after Chapter 5.

Also, an appendix with some technical elements about the Mellin transform are included apart to give the reading more fluency.

Chapter 2

Preliminary Notions

Throughout this work, several concepts of stochastic processes will be needed. In this chapter we give a summarized review of the fundamental concepts and results that will be central in our subsequent study. As the aim of this essay is the pricing of those assets known as derivatives, we can say that the most relevant result here is the Fundamental Theorem of Asset Pricing. This theorem will give us the key characterization of the value of the instruments under study. The description presented in this chapter is essentially based on books of stochastic processes. Proofs and further details on the subject may be found, for example, in Klebaner [Kl] and Björk [Bj].

2.1 Some background material on stochastic processes

Fields of Events and Filtration

Define by $\mathcal{F}_t$ the set of information available to investors at time t, which consists of stock prices before and at time t. At time t investors know which part of the sample space Ω contains the true state of the world. $\mathcal{F}_t$ is called a *field* or *algebra of sets*.

Definition 2.1 *(Field) A family of sets $\mathcal{F}$ is called a field if it satisfies:*

1. $\emptyset, \Omega \in \mathcal{F}$
2. *Given $A, B \in \mathcal{F}$, then $A \cup B, A \cap B, A \setminus B \in \mathcal{F}$.*

A generalization of the previous concept is the Σ-field, which is a field closed with respect to countable unions and countable intersections of its members.

Definition 2.2 *(Σ-fields) A collection $\mathcal{F}$ of subsets of Ω is called a Σ-field if it satisfies*

1. $\emptyset, \Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \Longrightarrow A^c \in \mathcal{F}$
3. $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F} \Longrightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Any subset B of Ω that belongs to $\mathcal{F}$ is called a measurable set.

The concept of filtration is used to model a flow of information. As time passes, an observer knows more and more detailed information. In the case of the price of stock, it describes how the information about prices is revealed to investors.

Definition 2.3 *(Filtration) A filtration $\mathfrak{F} = (\mathcal{F}_t)_{t\geq 0}$ is a family of Σ-fields such that $\mathcal{F}_0 \subset \mathcal{F}_s \subset \mathcal{F}_t$, for $s \leq t$.*

$\mathfrak{F}$ specifies how the information is revealed in time. The property that a filtration is increasing corresponds to the fact the information is not forgotten.

If we have a set Ω, a Σ-field of subsets of $\Omega, \mathcal{F}$, a probability P defined on elements of $\mathcal{F}$, and a filtration $\mathfrak{F}$ such that

$$\mathcal{F}_0 \subseteq \mathcal{F}_t \subseteq ... \subseteq \mathcal{F}_T = \mathcal{F},$$

then $(\Omega, \mathcal{F}, \mathfrak{F}, P)$ is called a *filtered probability space.*

In an abuse of notation we will drop $\mathfrak{F}$ and simply write $(\Omega, \mathcal{F}, P)$.

Σ-field Generated by a Random Variable

The Σ-field generated by a random variable X is the smallest Σ-field containing sets of the form $\{\omega : X(\omega) \in A\}$, for any set A in the Borel Σ-field $\mathcal{B}$. It will be denoted by $\mathcal{F}_X$, $\mathcal{F}^X$ or $\Sigma(X)$.

Stochastic Processes

Definition 2.4 *A stochastic process is a collection of random variables $(X(t))_{t\geq 0}$. For any fixed $t, X(t)$ is a random variable on $(\Omega, \mathcal{F}_T)$. A stochastic process is called adapted to filtration $\mathfrak{F}$ if for all $t \geq 0$, $X(t)$ is a random variable on $\mathcal{F}_t$, that is, if $X(t)$ is $\mathcal{F}_t$-measurable.*

A stochastic process is adapted if for any t, $\mathcal{F}_t$ contains all the information about $X(t)$ (and may contain extra information).

Definition 2.5 *(Martingale) A stochastic process M_t is a martingale with respect to the filtered probability space $(\Omega, \mathcal{F}, P)$ if it satisfies the following properties:*

1. *M_t is adapted to the filtration, i.e., M_t is $\mathcal{F}_t$-measurable for any $t > 0$,*
2. *$\mathbb{E}(|M_t|) < \infty$ for any $t \geq 0$,*
3. *For any $t \leq s$, we have $\mathbb{E}(M_t|\mathcal{F}_s) = M_s$.*

A standard notation for the conditional expectation of a $\mathcal{F}_t$-measurable process is $\mathbb{E}_t(\cdot) := \mathbb{E}(\cdot|\mathcal{F}_t)$.

The main process in the calculus of continuous processes is the *brownian motion*. It serves as a basic model for the cumulative effect of pure noise.

Definition 2.6 *(Brownian motion) A brownian motion $W(t)$ is a stochastic process satisfying the following conditions*

1. *W starts at 0, i.e. $W(0) = 0$*
2. *$W(t)$ is almost surely continuous*
3. *$W(t)$ has independent increments $W(t) - W(s) \sim \mathcal{N}(0, t - s)$. That is, $W(t) - W(s), W(t') - W(s')$ are independent, with $0 \leq s \leq t \leq s' \leq t'$.*

Property 3 above determines all the finite-dimensional distributions and it is possible to show that all of them are Gaussian. We don't prove here that a brownian motion exists, it can be found in many books on stochastic processes, and one construction is outlined in [Kl], Section 5.7.

Definition 2.7 *(Multivariate brownian motion) We define the N-dimensional brownian motion in as a random vector $\mathbf{W}(t) = (W^1(t), W^2(t), ..., W^N(t))$ with all coordinates $W^i(t)$ being independent one-dimensional brownian motions.*

2.1.1 Girsanov Theorem

Change of Measure

When we wish to compare two measures P and P^*, we don't want either of them simply to throw information away; since when they are positive they can be related by the Radon-Nikodym derivative. This motivates the following definition of equivalence of two measures.

Definition 2.8 *Two measures P and P^* are said to be equivalent if they operate on the same sample space, and if A is any event in the sample space, then*

$$P(A) > 0 \Longleftrightarrow P^*(A) > 0.$$

In other words, P is absolutely continuous with respect to P^*; and P^* is absolutely continuous with respect to P.

Theorem 2.9 *(Girsanov theorem for the brownian motion) Let $\mathbf{W}(t)$ be a m-dimensional brownian motion with respect to the filtered probability space $(\Omega, \mathcal{F}, P)$. Let $X(t) \in \mathbb{R}^n$ be the stochastic process*

$$d\mathbf{X}(t) = \mathbf{a}(t,\omega)dt + \mathbf{b}(t,\omega)d\mathbf{W}(t) \tag{2.1}$$

where $\mathbf{a}(t,\omega) \in \mathbb{R}^n$ and $\mathbf{b}(t,\omega) \in \mathbb{R}^{n\times m}$. Assume there exists processes $\lambda_t \in \mathbb{R}^m$ and $\mu_t \in \mathbb{R}^n$ such that

$$\mathbf{b}(t,\omega)\lambda_t(\omega) = \mathbf{a}(t,\omega) - \mu_t(\omega), \qquad \omega \in \Omega. \tag{2.2}$$

Let

$$M_t(\omega) = e^{-\int_0^t \langle \lambda_s(\omega), d\mathbf{W}_s\rangle - \frac{1}{2}\int_0^t \langle \lambda_s(\omega), \lambda_s(\omega)\rangle ds}, \tag{2.3}$$

where $\langle , \rangle$ represents the scalar product in $\mathbb{R}^m$. Suppose that M_t is a martingale with respect to $(\Omega, \mathcal{F}, P)$. Let P^* be another measure equivalent to P, such that

$$dP^*(\omega) = M_t(\omega)dP(\omega), \qquad i.e. \qquad \frac{dP^*(\omega)}{dP(\omega)} = M_t(\omega). \tag{2.4}$$

In this case, M_t is called the Radon-Nikodym derivative of P^* with respect to P. Then, P^* is a probability measure on $\mathcal{F}$ and the process

$$\tilde{\mathbf{W}}(t) = \int_0^t \lambda_s(\omega)ds + \mathbf{W}(t); \qquad t \le T, \tag{2.5}$$

is a Brownian Motion with respect to the filtered probability space $(\Omega, \mathcal{F}, P^*)$ and the process $\mathbf{X}(t)$ has the following representation in terms of $\tilde{\mathbf{W}}(t)$

$$d\mathbf{X}(t) = \mu_t(\mathbf{X}(t))dt + \mathbf{b}(t, \mathbf{X}(t))d\tilde{\mathbf{W}}(t).$$

Theorem 2.9 basically says that the properties of stochastic processes do not drastically change with the change of their drift. In fact, by changing the drift of a process, one changes also its probability law. The new probability measure is equivalent to the first one and the Radon-Nikodym derivative can be explicitly expressed as done in the theorem. In this thesis, the process λ_s will often be quoted as *the market price of risk*. A proof of Theorem 2.9 can be found in [Øk].

2.2 Concepts and definitions of asset pricing

Consider a financial market consisting of N exogenously given risky traded assets, being denoted by

$$S(t) = \begin{pmatrix} S_1(t) \\ \vdots \\ S_N(t) \end{pmatrix}.$$

We also assume that there exists a risk-free asset with price process $S_0(t)$, which we proceed to define.

Definition 2.10 *(Risk-free asset) The price process B is the price of a risk-free asset if it has the dynamics*

$$dB(t) = r(t)B(t)dt, \tag{2.6}$$

where $r(t)$ is any adapted process.

The defining property of a risk-free asset is thus that it has no driving dW term. We see that we also can write the B-dynamics as

$$\frac{dB(t)}{dt} = r(t)B(t),$$

so the B-process is given by the expression

$$B(t) = B(0)e^{\int_0^t r(s)ds}.$$

A natural interpretation of a riskless asset is that it corresponds to a bank with the (possibly stochastic) *short rate of interest* r. An important special case appears when r is a deterministic constant, in which case we can interpret B as the price of a bond. This riskless asset will be our numeraire.

We assume that the stock price S_i follows the dynamics

$$dS(t) = S(t)\mu(t, S(t))dt + S(t)\sigma(t, S(t))d\tilde{W}(t), \tag{2.7}$$

where $\tilde{W}$ is a brownian motion and μ and σ are given deterministic functions. The reason for the notation $\tilde{W}$, instead of the simpler W, will become clear below. The function σ is known as the volatility of S, while μ is the local mean rate of return of S.

Definition 2.11 *(Portfolio strategy) Let an N-dimensional price process $\{S(t); t \geq 0\}$ be given. A portfolio strategy (most often simply called a portfolio) is any $\mathcal{F}_t^S$-adapted N-dimensional process $\{h(t); t \geq 0\}$.*

The portfolio h is said to be Markovian if it is of the form

$$h(t) = h(t, S(t)),$$

for some function $h : \mathbb{R}^+ \times \mathbb{R}^N \longrightarrow \mathbb{R}^N$.

Definition 2.12 *(Value process) The value process V^h corresponding to the portfolio h is given by*

$$V^h(t) = \sum_{i=1}^{N} h_i(t)S_i(t). \tag{2.8}$$

Definition 2.13 *(Self-financing portfolio) A portfolio h is called self-financing if the value process V^h satisfies the condition*

$$dV^h(t) = \sum_{i=1}^{n} h_i(t)dS_i(t) = h(t) \cdot dS(t) \tag{2.9}$$

Definition 2.14 *(Arbitrage opportunity) An Arbitrage opportunity is a self-financing portfolio h that satisfies the following conditions:*

$$\begin{aligned} &V^h(0) = 0 \\ &P(V_h(T) \geq 0) = 1 \\ &P(V_h(T) > 0) > 0. \end{aligned}$$

2.2.1 Contingent claims

We take as given the model of a financial market defined by equations (2.6)-(2.7), and we now approach the main problem to be studied in this thesis, namely, the pricing of financial derivatives. Later we will give a mathematical definition, but let us at once present the single most important derivative: the *European call option.*

Definition 2.15 *(Call/Put option) A Call/Put option is a contract that gives at time 0 the holder the right (but not the obligation) to buy/sell an asset at time T called maturity and at a fix amount K called the strike.*

Another very important derivative in the financial industry is given by the following

Definition 2.16 *A forward contract is an agreement which gives at time 0 the investor the right to buy or sell an asset at time T called maturity and at a fixed amount F called strike.*

Note that the exercise price K and the time of maturity T are determined at the time when the option is written, which for us typically will be at $t = 0$. A *European put option* is an option which in the same way gives the holder the right to sell a share of the underlying asset at a predetermined strike price. For an *American call option* the right to buy a share of the underlying asset can be exercised at any time before the given time of maturity. The common factor of all these contracts is that they all are completely defined in terms of the underlying asset S, which makes it natural to call them derivative instruments or *contingent claims.* We will now give the formal definition of a contingent claim.

Definition 2.17 *(Contingent claim) Consider a financial market with vector price process S. A contingent claim with date of maturity (exercise date) T, also called a T-claim, is any stochastic variable $\mathcal{X} \in \mathcal{F}_T^S$. A contingent claim $\mathcal{X}$ is called a simple claim if it is of the form $\mathcal{X} = \Psi(S(T))$. The function Ψ is called the contract function.*

The interpretation of this definition is that a contingent claim is a contract, which stipulates that the holder of the contract will obtain $\mathcal{X}$ units of currency (which can be positive or negative) at the time of maturity T. The requirement that $\mathcal{X} \in \mathcal{F}_T^S$ simply means that, at time T, it will actually be possible to determine the amount of money to be paid out. We see that the European call is a simple contingent claim, for which the contract function is given by

$$\Psi(x) = \max(x - K, 0).$$

It is well known that the absence of arbitrage opportunities is essentially equivalent to the existence of a martingale measure Q, equivalent to the initial probability measure P, under which the discounted prices process is an $\mathcal{F}_t$-adapted martingale. Any equivalent martingale measure Q is characterized by a continuous version of its density process with respect to P, which can be written from the integral form of martingale representation.

2.2.2 Completeness

Definition 2.18 *A claim $\mathcal{X}$ is called attainable if there exists an admissible strategy V replicating the claim, that is, $V(t)$ satisfies (2.9), $V(t) \geq 0$ and $V(T) = \mathcal{X}$.*

If all the claims in the market are attainable, and if we knew how to price them, then we can price any claim. This leads to the following definition.

Definition 2.19 *Market models in which any claim is attainable are called complete.*

In the next section we will show a result that characterizes complete models in terms of the martingale measure. The proof can be found in Harrison-Kreps ([HK]) and Harrison-Pliska ([HP]).

2.3 Risk Neutral Valuation

Let us agree to denote the objective probability measure which governs our real model (2.6)-(2.7) by the letter P. Thus we say that the P-dynamics of the S-process are those of (2.7). We now define another probability measure Q under which the S-process has a different probability distribution. This is done by defining the Q-dynamics of S as

$$dS(t) = S(t)r(t, S(t))dt + S(t)\sigma(t, S(t))dW(t), \tag{2.10}$$

where W is a Q-Wiener process.

Theorem 2.20 *(Risk Neutral Valuation) Assuming r is constant, the arbitrage free price of the claim $\Psi(S(T))$ is given by $\Pi(t, \Psi) = F(t, S(t))$, with F given by the formula*

$$F(t, S(t)) = e^{-r(T-t)}\mathbb{E}^Q_{t,s}(\Psi(S(T))), \tag{2.11}$$

where the Q-dynamics of S are those of (2.10).

There is a natural economic interpretation of the formula (2.11). We see that the price of the derivative, given today's date t and today's stock price $S(t)$, is computed by taking the expectation of the final payment $\mathbb{E}^Q_{t,s}[\Psi(S(T))]$ and then discounting this expected value to present value using the discount factor $e^{-r(T-t)}$. The important point to note is that when we take the expected value we are not doing this using the objective probability measure P. Instead we shall use the Q-measure defined in (2.10). This Q-measure is sometimes called the *risk neutral measure* but most often it is called the *martingale measure*, and this will be our terminology. The reason for the name is that under Q the normalized process $S(t)/B(t)$ turns out to be a Q-martingale.

Definition 2.21 *A probability measure Q on $\mathcal{F}_T$ is called an equivalent martingale measure for the numeraire S_0, and the time interval $[0,T]$, if it has the following properties:*

- *Q is equivalent to P on $\mathcal{F}_T$*
- *All price processes $S_0, S_1, ..., S_N$ are martingales under Q on the time interval $[0,T]$.*

From an informal point of view, the main result of the entire arbitrage theory is the following not very precisely formulated Theorem.

Theorem 2.22 ***(First Fundamental Theorem)*** *The market model is free of arbitrage if and only if there exists a martingale measure, i.e., a measure $Q \sim P$ such that the processes*

$$\frac{S_1(t)}{S_0(t)}, \ldots, \frac{S_N(t)}{S_0(t)}$$

are martingales under Q.

For the case when the numeraire is the money account we have an alternative characterization of a martingale measure. The proof is a simple application of the Itô formula.

Proposition 2.23 *If the numeraire S_0 is the money account, i.e.*

$$S_0(t) = e^{\int_0^t r(s)ds},$$

where r is the (possibly stochastic) short rate, and if we assume that all processes are Wiener driven, then a measure $Q \sim P$ is a martingale measure if and only if all assets $S_0, S_1, \ldots, S_N$ have the short rate as their local rates of return, i.e., if the Q-dynamics are of the form

$$dS_i(t) = r(t)S_i(t)dt + \sigma_i(t)S_i(t)dW^Q, \tag{2.12}$$

where W^Q is a (multidimensional) Q-Wiener process.

Theorem 2.24 *(Second Fundamental Theorem) Assume that the market is arbitrage free and consider a fixed numeraire asset S_0. Then the market is complete if and only if the martingale measure Q, corresponding to the numeraire S_0, is unique.*

To avoid arbitrage, the value of an attainable claim at time $t < T$ must be the same as that of the replicating portfolio at t.

Theorem 2.25 *(Pricing by No-arbitrage) The arbitrage free price process for the T-claim $\mathcal{X}$ is given by*

$$\Pi(t, \mathcal{X}) = S_0(t)\mathbb{E}^Q\left[\frac{\mathcal{X}}{S_0(T)}\middle|\mathcal{F}_t\right], \tag{2.13}$$

where Q is the (not necessarily unique) martingale measure for the a priori given market $S_0, S_1, ..., S_N$, with S_0 as the numeraire.

Note that different choices of Q will generically give rise to different price processes. In particular, we note that if we assume that if S_0 is the money account

$$S_0(t) = S_0 e^{\int_0^t r(u)du},$$

where r is the short rate, then (2.13) reduces to the familiar risk neutral valuation formula.

Theorem 2.26 *(Risk Neutral Valuation Formula) Assuming the existence of a short rate, the pricing formula takes the form*

$$\Pi(t, \mathcal{X}) = \mathbb{E}^Q\left[e^{-\int_t^T r(u)du}\mathcal{X}\middle|\mathcal{F}_t\right] \tag{2.14}$$

where Q is a (not necessarily unique) martingale measure with the money account as the numeraire.

2.4 The Black-Scholes model

The Black-Scholes model is the most widely used continuous time pricing model for options in the financial industry. The assumptions of the Black-Scholes model are:

- the volatility is constant in time and can be accurately estimated
- the stock can be continuously traded
- the market moves continuously, is liquid and there are no transactions costs.

We assume that the a priori given market consists of two assets, a riskless asset B and a risky asset S that satisfy formulas (2.6) and (2.7) respectively, where, according to the assumptions, μ (the asset drift) and σ (the asset volatility) are constants. This dynamics in the stock price is called *geometric brownian motion*:

$$dS(t) = \mu S(t)dt + \sigma S(t)d\tilde{W}(t). \tag{2.15}$$

The previous equation has an explicit solution. Using Ito formula for $f(x) := \ln(x)$, we find that

$$S(t) = S(0)e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma\tilde{W}(t)}. \tag{2.16}$$

The Equivalent Martingale Measure (EMM) Q makes $S(t)e^{-rt}$ into a martingale. By Theorems 2.22 and 2.24 it exists and is unique. It is obtained by letting $\frac{\mu}{\sigma}dt + \sigma d\tilde{W}(t) = \frac{r}{\sigma}dt + \sigma dW$ for a Q-brownian motion $W(t)$. In this case, the market price of risk introduced in Theorem 2.9 is $\lambda_t = \frac{\mu - r}{\sigma}$.

Thus, under the equivalent martingale measure Q, $S(T)$ has a lognormal distribution, $\mathcal{LN}((r - \sigma^2)T + \frac{\sigma^2}{2}T + \ln(S(0)), \sigma^2 T)$. The price of a claim $\mathcal{X}$ at time T is given by

$$C(t) = e^{-r(T-t)}\mathbb{E}^Q(\mathcal{X}|\mathcal{F}_t). \tag{2.17}$$

If $\mathcal{X} = g(S(T))$, then by the Markov property of $S(t)$,

$$C(t) = \mathbb{E}^Q(g(S(T))|\mathcal{F}_t) = \mathbb{E}^Q(g(S(T))|S(t)).$$

Using (2.16), the conditional distribution under Q given $\mathcal{F}_t$ is obtained from the equation

$$S(T) = S(t)e^{(r - \frac{\sigma^2}{2})(T-t) + \sigma(\tilde{W}(T) - \tilde{W}(t))},$$

and is lognormal

$$\mathcal{LN}\big((r - \frac{\sigma^2}{2})(T-t) + \ln(S(0)), \sigma^2(T-t)\big).$$

A call option pays $\mathcal{X} = (S(T) - K)^+$ at time T. To find its price at time $t = 0$, according to (2.17), we must calculate $\mathbb{E}^Q[(S(T) - K)^+]$, where expectation $\mathbb{E}$ is taken under the arbitrage-free probability Q. In the Black-Scholes model the Q-distribution of $S(T)$ is a lognormal. The price at time t of the European call option on stock with strike K and maturity T is given by (Theorem 2.13)

$$C(t) = e^{-r(T-t)}\mathbb{E}^Q\big[(S(T) - K)^+|\mathcal{F}_t\big].$$

From the lognormality of S, it can be shown that $C(t)$ is given by the *Black-Scholes formula*

$$C(t) = S(t)\Phi(d_+) - Ke^{-r(T-t)}\Phi(d_-), \tag{2.18}$$

where

$$d_\pm = \frac{\ln\frac{S(t)}{K} + (r \pm \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}, \tag{2.19}$$

and

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du,$$

is the normal distribution function.

Pricing of Claims by a PDE

Let $\mathcal{X}$ be a claim of the form $\mathcal{X} = g(S(T))$. Since the stock price satisfies SDE (2.10) with r and σ taken to be constants, by the Markov property of $S(t)$ it follows from (2.17) that the price of $\mathcal{X}$ at time t is

$$C(t) = e^{-r(T-t)}\mathbb{E}^Q\big((g(T)-K)^+|\mathcal{F}_t\big) = e^{-r(T-t)}\mathbb{E}^Q\big((g(T)-K)^+|S(t)\big). \tag{2.20}$$

By the Feynman-Kac formula (see Theorem 6.8 in [Kl]),

$$C(x,t) = e^{-r(T-t)}\mathbb{E}^Q\big((g(T)-K)^+|S(t)=x\big)$$

solves the following partial differential equation (PDE)

$$-rC(t,x) + \frac{\partial C}{\partial t}(t,x) + rx\frac{\partial C}{\partial x} + \frac{1}{2}\sigma^2x^2\frac{\partial^2 C}{\partial x^2}(t,x) = 0, \tag{2.21}$$

with terminal conditions

$$C(T,x) = g(x), \qquad x \geq 0,$$

where r is the risk free interest rate, C is the price function of the option, T is the maturity time and $g(x)$ the payoff of the claim at time T when the stock price is x. If $g(x) = (x-K)^+$, the solution to this partial differential equation for the price at time t of a European call option with strike price K is given by the so called Black-Scholes formula (2.18).

Extension of Black-Scholes model: deterministic volatility option model

Let $\sigma(t) > 0$ be a continuous positive function. Define the *Black-Scholes-Merton option price formula*

$$F^{BS}(x,t,\nu,K,r) := Ke^{-r(T-t)}\big(x\delta_A(x,t,\nu) + \delta_B(x,t,\nu)\big) \tag{2.22}$$

on $t \le T$, where $\delta_A(x,T,\nu) := 1_{\{x>K\}}$, $\delta_B(x,T,\nu) := -1_{\{x>K\}}$ and

$$\delta_A(x,t,\nu) := \Phi\bigg(\frac{\log x}{\sqrt{\nu_{t,T}}} + \frac{\sqrt{\nu_{t,T}}}{2}\bigg), \quad \delta_B(x,t,\nu) := -\Phi\bigg(\frac{\log x}{\sqrt{\nu_{t,T}}} - \frac{\sqrt{\nu_{t,T}}}{2}\bigg),$$

for $t < T$, where $\nu_{t,T} := \int_t^T \sigma^2(u)du$ and K is a fixed parameter. The function $F^{BS}(x,t,\nu,K,r)$ is continuous in x and on $t < T$ is C^1 in t and analytic in x. Note also that $-1 \le \delta_B \le 0 \le \delta_A \le 1$ and $F^{BS}(x,T,\nu,K,r) = K(x-1)^+$. We recover formula (2.18) by taking $x = \frac{S(t)}{Ke^{-r(T-t)}}, \nu = \sigma$.

Implied volatility

An *implied volatility* σ_t^I is the volatility implied in the market prices C of the derivatives in a particular pricing model (for instance, the Black-Scholes model). It is the volatility that, replaced in the formula of the option price C, yields the exact value of the option. We give a precise definition below.

Definition 2.27 *(Implied volatility) Denote by C^m the observed market price of an option with strike K and expiration T. The implied volatility $\sigma_t^I(K,T)$ is defined as the value of the volatility parameter in the Black-Scholes (2.22) formula that matches the observed price, namely*

$$F^{BS}\big(S(t),t,\sigma_t^I(K,T),K,r\big) = C_t^m(K,T).$$

REMARK Usually, to calculate the implied volatility σ^I, we consider the price of an option C as a function $f(\sigma,\cdot)$ of the volatility and other parameters. Therefore, we can as well consider the volatility σ as a function $g(C,\cdot)$ of the option price C, and so there exists at most one value of the volatility which leads to a particular value of the option. The implied volatility σ^I is this value and therefore is the (numerical) solution to the equation $f(\sigma,\cdot) = C$, with unknown σ.

Chapter 3

Variance and Volatility Derivatives

In this chapter, definitions of the main variance derivatives are introduced, together with a description of their underlying process: realized variance. We also discuss similarities and differences between the continuous and discrete versions of this process. Several standard properties of realized variance are introduced, all of them valid in a world where the risk-free interest rate is constant. An extension to this properties is shown when this rate is assumed to be stochastic. We would like to make clear here that all the treatment performed is merely theoretical, having precluded frictions and limitations of practical financial nature.

3.1 General Setting

A basic assumption here will be that the stock price evolves continuously in time without jumps. This is an assumption for the Black-Scholes model and yet according to the Black-Scholes model, a typical process for the stock price is the risk neutral geometric brownian motion (i.e., Geometric brownian motion under the equivalent martingale measure). The general framework studied here, that includes stochastic volatility, is

$$\frac{dS_t}{S_t} = \mu(t,S_t,V_t)dt + \sigma(t,V_t)\big(\sqrt{1-\rho^2(t,S_t,V_t)}d\tilde{W}_t^1 + \rho(t,S_t,V_t)d\tilde{W}_t^2\big), \quad (3.1)$$

$$dV_t = \eta(t,V_t)dt + \gamma(t,V_t)d\tilde{W}_t^2, \quad (3.2)$$

where $\{S_t, 0 \leq t \geq T\}$ is the price process of a primitive asset supposed to pay no dividends, $\{\sigma_t = \sigma(V_t), 0 \leq t \geq T\}$ is the volatility process, and $\{\rho_t = \rho(t,S_t,V_t), 0 \leq t \geq T\}$ is the

correlation process between the asset price and its volatility taking values in the interval $(-1, 1)$. $\tilde{\mathbf{W}} = (\tilde{W}^1, \tilde{W}^2)$ is a two-dimensional brownian motion in a filtered probability space $(\Omega, \mathcal{F}, P)$.

Here, we assume that the drift μ and the continuously-sampled volatility are arbitrary functions of time and other parameters. For simplicity of presentation, we assume the stock pays no dividends; allowing for dividends does not significantly alter the derivation.

By Theorems (2.22) and (2.9), there exist an equivalent martingale measure Q characterized by

$$M_t = \frac{dQ}{dP}\Big|_{\mathcal{F}_t} = \exp\Big(- \int_0^t \lambda_u d\tilde{W}_u^1 - \int_0^t \nu d\tilde{W}_u^2 - \frac{1}{2}\int_0^t \lambda_u^2 du - \frac{1}{2}\int_0^t \nu_u^2 du \Big), \tag{3.3}$$

where (λ, ν) is adapted to $\mathcal{F}_t$ and satisfies the integrability condition $\int_0^T \lambda_u^2 du < \infty$ and $\int_0^T \nu_u^2 du < \infty$ a.s. By the martingale property under Q of the discounted underlying asset prices, we have:

$$\Big(\lambda_t \sqrt{1-\rho_t^2} + \nu_t \rho_t\Big)\sigma_t = \mu_t - r, \qquad 0 \le t \le T \quad a.s. \tag{3.4}$$

The processes λ and ν satisfying (3.4)(and the integrability condition) are interpreted as the risk premia relative respectively to the two sources of uncertainty $\tilde{W}^1$ and $\tilde{W}^2$. Since S is the only traded asset, the risk premia λ and ν are not fixed by the last relation, which explains the non uniqueness of the martingale measure Q in this incomplete market context. For each choice of the volatility risk premium process $\{\nu_t, 0 \le t \le T\}$, the risk premium process $\{\lambda_t, 0 \le t \le T\}$ is fixed by (3.4) and we can define an *admisible equivalent martingale measure* $Q(\nu)$ characterized by its density process $\{M_t(\nu), 0 \le t \le T\}$ with respect to P. By Girsanov's theorem, the process $\mathbf{W}(\nu) = (W^1(\nu), W^2(\nu))^\top$ defined by

$$W_t^1(\nu) = \tilde{W}_t^1 + \int_0^t \lambda_u du, \qquad W_t^2(\nu) = \tilde{W}_t^2 + \int_0^t \nu_u du, \tag{3.5}$$

is a two-dimensional brownian motion under $Q(\nu)$ adapted to the filtration $\{\mathcal{F}_t, 0 \le t \le T\}$. The dynamics of the model under an admissible equivalent martingale measure $Q(\nu)$ is described by:

$$\frac{dS_t}{S_t} = r(t, S_t)dt + \sigma(V_t)\big(\sqrt{1-\rho^2(t, S_t, V_t)}dW_t^1(\nu) + \rho(t, S_t, V_t)dW_t^2(\nu)\big), \tag{3.6}$$

where r is the risk-free interest rate.

3.2 Discrete Variance vs. Continuous Variance

By *realized variance* from time 0 to time T, we understand the quadratic variation of $\log S$ at time T. Under the general assumptions stated in the previous section, S has an instantaneous volatility process σ_t and we can think of realized variance as the time integral of σ_t^2 from 0 to T.

Definition 3.1 *Assume that the positive asset price S follows (3.6). The continuously sampled realized variance of S from time 0 to time T is defined as*

$$V_c(0,T) = \frac{1}{T}\int_0^T \sigma_t^2 dt. \tag{3.7}$$

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. The advantage of this approximation is that in an autonomous stochastic volatility model continuous realized variance does not depend on the path of the underlying; it only depends on the path realized by the instantaneous variance process.

The procedure for calculating discretely realized variance is specified in the derivatives contract and includes details about the source and observation frequency of the price of the underlying asset, the annualization factor to be used in moving to an annualized volatility and the method for calculating the variance. Most traded contracts define realized variance to be the sum of the squares of the daily returns of the stock annualized.

Definition 3.2 *Let $0 = t_0 < t_1 < ... < t_n = T$ be a partition of the time interval $[0,T]$ into n equal segments of length Δt. The discretely sample realized variance is defined as*

$$V_n(0,T) = \frac{AF}{n-1}\sum_{i=0}^{n-1} \log^2\left(\frac{S_{i+1}}{S_i}\right). \tag{3.8}$$

Here, S_i is the price of the asset at the i-th observation time t_i and AF is the annualization factor, taken in general as n/T.

This definition of realized variance differs from the usual sample variance because the sample average is not subtracted from each observation. Since the sample average is approximately zero, the realized variance is close to the sample variance.

$V_n(0,T)$ is an unbiased variance estimator for σ_t. It can be shown that (see [BL])

$$V_c(0,T) = \lim_{n\to\infty} V_n(0,T). \tag{3.9}$$

However, substituting discretely realized variance by its continuous counterpart is an approximation of dubious quality in some cases. A simple example suffices to show the difference between discrete and continuous variance. Consider a continuous time Black model

$$\frac{dS_t}{S_t} = \sqrt{V_0}dW_t \tag{3.10}$$

with constant variance parameter $V_0 = 0.09$, which corresponds to 30% Black volatility for the underlying 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 $(n/252)V_0$ a.s. Therefore, there is no uncertainty associated with continuously realized variance.

Discrete realized variance, however, 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 $(n/252)V_0$, equal to the continuously realized variance, and its standard deviation is $V_0/252\sqrt{2n}$. 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. 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, about 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.

3.3 Variance Contracts

In this section we proceed to define and describe the main financial instruments under consideration in this work: variance swaps, volatility swaps and swaptions. These are

contracts whose values depend on the future level of variance of the underlying asset.

3.3.1 Variance Swap

In a few words, a *variance swap* is simply a forward contract with realized variance as underlying.

Definition 3.3 *A variance swap is a forward contract in which one counterpart agrees to pay the other a notional amount, N, times the difference between a fixed level and a realized level of variance. The fixed level is called the variance strike for variance swap. In symbols, the payoff of a variance swap is*

$$(V_n(0,T) - K_{var}) \times N, \tag{3.11}$$

where $V_n(0,T)$ is the realized variance of stock return over the life of the contract, n is the number of sampling dates, K_{var} is the variance strike, and N is the notional amount of the swap in dollars.

At expiration, the holder of a variance swap receives N dollars for every unit by which the stock realized variance V exceeds the variance strike K_{var}. N represents the amount that the holder receives at maturity if the realized variance exceeds the strike K_{var} by one unit. The unit of N is dollar per unit variance point; for example, $N = \$30,000/(\text{variance point})$.

3.3.2 Volatility Swap

Definition 3.4 *A volatility swap is a forward contract on the future realized volatility of the stock price. If K_{vol} denotes the strike of a volatility swap (typically quoted in units of percent, e.g. 20%), its payoff is defined as*

$$(\sqrt{(V_n(0,T)} - K_{vol}) \times N \tag{3.12}$$

where $\sqrt{(V_n(0,T)}$ is the realized stock volatility (quoted in annual terms) over the life of the contract.

The volatility strike K_{vol} is typically quoted in units of percent (e.g., 20%). An investor who is long a volatility swap with strike 20% and notional of \$1 million would make a

profit of (0.25-0.2)*1,000,000=$50,000. This contract is similar to a variance swap except that the traded asset here is the volatility instead of the variance. The notional amount N of the payoff is now in dollar per unit volatility point.

There is another type of volatility swap known as volatility-average swap, in which the measure of the volatility is simply the average over time of the return on the stock price. In discrete time that is

$$\frac{1}{n-1}\sum_{i=1}^{n-1}\left(\frac{S_i - S_{i-1}}{S_{i-1}}\right). \tag{3.13}$$

In continuous time, we approximate this measure of volatility by

$$\sigma_R = \frac{1}{T}\int_0^T \sigma_t dt. \tag{3.14}$$

Here the strike is denoted $K_{vol-ave}$ and the payoff of the contract is

$$(\sigma_R - K_{vol-ave}). \tag{3.15}$$

As we will see later, the replication and pricing of volatility swaps are much more difficult than those of the variance, since with a zero sample mean of returns, the stock price dynamics provides an explicit formula for the realized variance but not for the realized volatility.

3.3.3 Swaptions

In the same way we defined forward contracts on future realized variance and volatility, we can also define option with this variable as the underlying. This instruments, the swaptions, have a strike K and payoffs

$$C_T = N \times \max(V_c - K; 0), \qquad P_T = N \times \max(K - V_c; 0);$$

for volatility call and put respectively. Swaptions are more difficult to price than swaps. The reason behind this lies in the fact that stock options have exposure not only to stock price, but also to its volatility. In a similar manner, swaptions are also contaminated by the volatility or variance of the stock and the volatility of this new variable as well. This is a clear disadvantage of swaptions compared to swaps, which provide pure risk to volatility only. They are also more difficult to price than European equity options, because, unlike these, the payoff of variance options depends on realized variance V_c, which is not a traded security in the market. From now on and for the rest of the work, the notional amount of the swap will be $1 per unit traded asset point.

3.4 Replicating and pricing variance swaps

Replication and pricing without an explicit model for volatility

Valuing a variance forward contract or swap is not different from valuing any other derivative security. Assume momentarily that the risk-free interest rate process r is constant. Then, the present value of a forward contract C_0 on future realized variance with strike K_{var} is by (2.14)

$$C_0 = \mathbb{E}_0^Q[e^{-rT}(V_c - K_{var})].$$

Thus, for calculating variance swaps we only need to know $\mathbb{E}_0^Q(V_c)$. Volatility swaps require more work, due to the nonlinearity of the square root function. From Brockhaus-Long approximation, we have

$$\mathbb{E}(\sqrt{V_c}) \approx \sqrt{\mathbb{E}(V_c)} - \frac{\mathsf{Var}(V_c)}{8\mathbb{E}(V_c)^{\frac{3}{2}}}, \tag{3.16}$$

Then, for volatility swaps we need both $\mathbb{E}(V_c)$ and $\mathsf{Var}(V_c)$. The second term containing the variance in (3.16) is the well known *convexity adjustment*.

No one knows with certainty the value of future volatility. In implied tree models, the so-called local volatility consistent with all current options prices is extracted from the market prices of traded stock options. You can then use simulation to calculate the fair variance K_{var} as the average of the experienced variance along each simulated path consistent with the risk-neutral stock price evolution given of equation (3.6).

The above approach is good for valuing the contract, but it does not provide insight into how to replicate it. The essence of the replication strategy is to devise a position that, over the next instant of time, generates a payoff proportional to the incremental variance of the stock during that time.

Denote the logarithmic returns process by

$$X_t := \log(S_t/S_0),$$

and write $\langle X \rangle_t$ for the quadratic variation of X till t (also known as the realized variance of the returns on S). Under assumption (3.6),

$$\langle X \rangle_t := \int_0^t \sigma_u^2 du.$$

By applying Ito's lemma to the function $F(S) := \log(S)$, we obtain the following SDE for the process $\log(S_t/S_0)$

$$d(\log(S_t)) = \left(\mu_t - \frac{1}{2}\sigma_t^2\right)dt + \sigma_t dW_t, \tag{3.17}$$

where we have taken $W = \sqrt{1-\rho^2}W^1 + \rho W^2$.

Subtracting Equation (3.6) from Equation (3.17), we obtain

$$\frac{dS_t}{S_t} - d(\log(S_t)) = \frac{1}{2}\sigma_t^2 dt, \tag{3.18}$$

in which all dependence on the drift has canceled. Integrating Equation (3.18) over all times from 0 to T and dividing by T both hand sides, we obtain the continuously-sampled variance

$$V_c = \frac{1}{T}\langle X\rangle_T = \frac{1}{T}\int_0^T \sigma_t^2 dt = \frac{2}{T}\left[\int_0^T \frac{dS_t}{S_t} - \log\left(\frac{S_T}{S_0}\right)\right]. \tag{3.19}$$

Note that we have easily derived this expression because of the sample mean that was assumed to be zero for the definition of the realized variance. By not assuming this, an extra term would be added to the realized variance.

The mathematical identity (3.19) describes the replication strategy for variance. The first term in the brackets can be thought of as the net outcome of continuous rebalancing a stock position so that it is always instantaneously long $1/S_t$ shares of stock worth \$1. The second term represents a static short position in a contract which, at expiration, pays the logarithm of the total return. This continuously rebalancing strategy captures the realized variance of the stock from inception to expiration at time T. Note that no expectations or averages have been taken. Equation (3.19) guarantees that variance can be captured no matter which path the stock price takes, as long as it moves continuously. It also provides another formula for the fair strike of a variance swap. Instead of averaging over future variances, one can take risk neutral expectation to obtain the cost of replication directly, so that

$$K_{var} = \mathbb{E}(V_c) = \frac{2}{T}\mathbb{E}\left(\int_0^T \frac{dS_t}{S_t} - \log\frac{S_T}{S_0}\right) = \frac{2}{T}\left[\mathbb{E}\left(\int_0^T \frac{dS_t}{S_t}\right) - \mathbb{E}\left(\log\frac{S_T}{S_0}\right)\right]. \tag{3.20}$$

The expectation $\mathbb{E}\big(\int_0^T \frac{dS_t}{S_t}\big)$ can be easily calculated from (3.6) as

$$\mathbb{E}\left(\int_0^T \frac{dS_t}{S_t}\right) = \mathbb{E}\left(\int_0^T r dt\right) + \mathbb{E}\left(\int_0^T \sigma_t dW_t\right). \tag{3.21}$$

As the Itô integral is a martingale, we have

$$\mathbb{E}\left(\int_0^T \sigma_t dW_t\right) = 0.$$

Thus,

$$\mathbb{E}\left(\int_0^T \frac{dS_t}{S_t}\right) = \mathbb{E}\left(\int_0^T r dt\right) = rT.$$

This result represents the fact that a shares position, continuously rebalanced to be worth \$1, has a forward price that grows at the riskless rate. Substituting in (3.20) we find that

$$K_{var} = 2r - \frac{2}{T}\mathbb{E}\left(\log \frac{S_T}{S_0} \right). \tag{3.22}$$

As there are no actively traded log contracts for the second term in Equation (3.22), one must duplicate the log payoff, at all stock price levels at expiration, by decomposing its shape into linear and curved components, and then duplicating each of these separately. The linear component can be duplicated with a forward contract on the stock with delivery time T. The remaining curved component, representing the quadratic and higher order contributions, can be duplicated using standard options with all possible strike levels and the same expiration time T. For practical reasons we want to duplicate the log payoff with liquid options that is, with a combination of out-of-the-money calls for high stock values and out-of-the-money puts for low stock values. We introduce a new arbitrary parameter S_* to define the boundary between calls and puts. The log payoff can then be rewritten as

$$\log \frac{S_T}{S_0} = \log \frac{S_T}{S_*} + \log \frac{S_*}{S_0}. \tag{3.23}$$

The last term is constant, but the first one is random. We will make use of the following proposition to decompose the log payoff in terms of call and put options and forward contracts. We attach its proof for the sake of completeness, but we stress it is not an original result of this thesis.

Proposition 3.5 *For any twice-differentiable function $f : I \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ and any $S_* \geq 0$, we have*

$$f(S_T) = f(S_*) + f'(S_*)(S_T - S_*) + \int_0^{S_*} f''(K)(K - S_T)^+ dK + \int_{S_*}^{\infty} f''(K)(S_T - K)^+ dK. \tag{3.24}$$

Proof For the shifting property of the Dirac Delta function (see [RY]) , we have

$$f(S_T) = \int_0^{\infty} f(K)\delta(S_T - K)dK.$$

For any $S_* \geq 0$,

$$f(S_T) = \int_0^{S_*} f(K)\delta(S_T - K)dK + \int_{S_*}^{\infty} f(K)\delta(S_T - K)dK. \tag{3.25}$$

If H denotes the Heaviside function, defined as $H(x) := I_{(0,+\infty)}(x)$, recall that $H'(x) = \delta(x)$, and integrating by parts (3.25), we get

$$\begin{aligned} f(S_T) &= f(K)I_{(0,K)}(S_T)\Big|_0^{S_*} - \int_0^{S_*} f'(K)I_{[0,K]}(S_T)dK \\ &+ f(K)I_{[K,+\infty]}(S_T)\Big|_{S_*}^{\infty} - \int_{S_*}^{\infty} f'(K)I_{(K,+\infty)}(S_T)dK. \end{aligned}$$

Once again using that $(\max(x,0))' = H(x)$ and integrating by parts, we arrive at

$$\begin{aligned} f(S_T) &= f(S_*)I_{(0,S_*)}(S_T) - f'(K)(K-S_T)^+\Big|_0^{S_*} + \int_0^{S_*} f''(K)(K-S_T)^+dK \\ &+ f(S_*)I_{[S_*,)} - f'(K)(S_T-K)^+\Big|_{S_*}^{\infty} + \int_{S_*}^{\infty} f''(K)(S_T-K)^+dK \\ &= f(S_*) + f'(S_*)[(S_T-S_*)^+ - (S_*-S_T)^+] \\ &+ \int_0^{S_*} f''(K)(K-S_T)^+dK + \int_{S_*}^{\infty} f''(K)(S_T-K)^+dK. \end{aligned} \tag{3.26}$$

We obtain the result by noting that $(S_T-S_*)^+ - (S_*-S_T)^+ = S_T - S_*$ and replacing in (3.26). □

Proposition 3.5 is very useful in quantitative finance. For example, if we apply it to the left hand side of (3.23), we get

$$\log(S_T) - \log(S_*) = \frac{S_T-S_*}{S_*} - \int_0^{S_*} \frac{1}{K^2}(K-S_T)^+dK - \int_{S_*}^{\infty} \frac{1}{K^2}(S_T-K)^+dK. \tag{3.27}$$

Identity (3.27) represents the decomposition of a log payoff into a portfolio consisting of:

- a short position in forward contracts struck at S_*
- a long position in put options struck at K, for all strikes from 0 to S_*
- a similar long position in call options struck at K, for all strikes from S_* to ∞.

All contracts expire at time T. Inserting (3.27) into (3.23), the fair value of future variance can be related to the initial fair value of each term on the right hand side of Equation (3.19). Using identity (3.27), we obtain

$$K_{var} = \frac{2}{T}\left[rT - \mathbb{E}_0\left(\log\frac{S_*}{S_0} + \frac{S_T-S_0}{S_*} - \int_0^{S_*}\frac{1}{K^2}(K-S_T)^+dK - \int_{S_*}^{\infty}\frac{1}{K^2}(S_T-K)^+\right)dK\right], \tag{3.28}$$

or equivalently

$$\begin{aligned} K_{var} &= \frac{2}{T}\Bigg[rT - \Bigg(\log\frac{S_*}{S_0} + 1 + \frac{\mathbb{E}_0(S_T)}{S_*} - \\ &- \int_0^{S_*} \frac{1}{K^2}\mathbb{E}_0[(K-S_T)^+]dK - \int_{S_*}^{\infty} \frac{1}{K^2}\mathbb{E}_0[(S_T-K)^+]\Bigg)dK\Bigg]. \end{aligned} \tag{3.29}$$

Under the equivalent martingale measure, the discounted stock price and options price are martingales. Thus,

$$S_0 = e^{-rT}\mathbb{E}_0(S_T), \quad C_0(K) = e^{-rT}\mathbb{E}_0[(S_T-K)^+], \quad P_0(K) = e^{-rT}\mathbb{E}_0[(K-S_T)^+], \tag{3.30}$$

where $P_0(K), C_0(K)$ are the premiums (prices at $t = 0$) of the put and call options respectively struck at K with maturity T.

3.5 Extension to Stochastic Interest Rates

In this section we present a little but useful extension of the previous results under the context of stochastic interest rates. As it is well known, the effect of stochastic interest rates affects the evaluation of the contract mainly due to the correlation between the stock and the rates, which is difficult to hedge. Here we propose an extension of Carr-Lee ([CL05]) characterization and a way to hedge this correlation by means of *Exchange Options*.

3.5.1 Carr-Lee approach

In their ground-breaking paper, Carr and Lee develop a technology to extend the previous results for variance derivatives to arbitrary functions of realized variance under a zero-correlation assumption between the volatility and the stock price processes. The core of their methodology is the finding that a general function of variance (including volatility) admits valuation and replication using portfolios of European options and stock. They achieve this by showing that realized variance $\langle X\rangle_T$ satisfies an identity of the form

$$\mathbb{E}^Q(h(\langle X\rangle_T)) = \mathbb{E}^Q(G(S_T)),$$

for a given payoff h and a suitable G. The expectation containing h is the price of the contract on variance and the right-hand side is the price of a contract on a payoff dependent

exclusively on stock price. Therefore, the value of a variance contract is expressed only in terms of the prices observable in the market (the option market) and is fully independent of the parameters of any model. Now we extend their essential results to the context of stochastic interest rates.

General Assumptions

Throughout this section, we will base our analysis on the following

Assumption 3.6 *On a filtered probability space $(\Omega, \mathcal{F}, P)$ satisfying the usual conditions, there exists a $(\mathcal{F}_t, P)$-brownian motion W_t and some $\mathcal{F}_t$-adapted processes σ_t, r_t such that:*

- $\langle X \rangle_T < M, \quad (X = \log(S))$
- *σ and W are independent*
- *σ and r are independent*

Given a general payoff G, the last assumption gives an interpretation of $P(0,T)\mathbb{E}[G(\langle X \rangle_T)]$ as a price, because

$$\mathbb{E}_0\left[G(\langle X \rangle_T) e^{-\int_0^T r(u)du}\right] = P(0,T)\mathbb{E}_0[G(\langle X \rangle_T)],$$

so the main challenge is to get the value of the last expectation. As in Carr and Lee [CL05], we start with exponential payoffs, i.e., functions G of the form $G(t) = e^{\lambda t}$, which serve as building blocks to create more general functions of $\langle X \rangle_T$.

Proposition 3.7 *(Extension of Proposition 16 of Carr-Lee) For each $\lambda \in \mathbb{C}$ and $t \leq T$,*

$$\mathbb{E}_t e^{\lambda \langle X \rangle_T} = e^{\lambda \langle X \rangle_t} \left(\frac{S_t}{\beta_t}\right)^{-p} \mathbb{E}_t\left[\left(\frac{S_T}{\beta_T}\right)^{p}\right], \tag{3.31}$$

where $p = \frac{1}{2} \pm \sqrt{1/4 + 2\lambda}$.

PROOF: Consider the process $Y_t = \frac{S_t}{\beta_t}$. By Itô's rule,

$$\frac{dY_t}{Y_t} = \frac{dS_t}{S_t} - \frac{d\beta_t}{\beta_t} = \sigma_t dW_t$$

and in consequence Y_t is a martingale. Note also that $\langle \log(Y) \rangle_T = \langle X \rangle_T$.

We apply a more general version of Hull-White's [HW] conditioning argument. Conditional on $\mathcal{F}_T^\sigma$, the W is still a brownian motion, by independence. So conditional on $\mathcal{F}_t \vee \mathcal{F}_T^\sigma$,

$$\log\left(\frac{Y_T}{Y_t}\right) = -\frac{1}{2}(\langle X\rangle_T - \langle X\rangle_t) + \int_t^T \sigma_s \, dW_s \sim \mathcal{N}\Big(-\frac{1}{2}(\langle X\rangle_T - \langle X\rangle_t), \langle X\rangle_T - \langle X\rangle_t\Big),$$

For each $p \in \mathbb{C}$, therefore,

$$\begin{aligned}\left(\frac{S_t}{\beta_t}\right)^{-p}\mathbb{E}_t\left(\frac{S_T^p}{\beta_T^p}\right) = \mathbb{E}_t e^{p\log(Y_T)-\log(Y_t)} &= \mathbb{E}_t\left[\mathbb{E}_t\left(e^{p\log(\frac{Y_T}{Y_t})}\Big|\mathcal{F}_T^\sigma\right)\right] \\ &= \mathbb{E}_t\left[e^{\mathbb{E}_t\left(p\log(\frac{Y_T}{Y_t})\Big|\mathcal{F}_T^\sigma\right)+\frac{1}{2}\mathsf{Var}_t\left(p\log(\frac{Y_T}{Y_t})\Big|\mathcal{F}_T^\sigma\right)}\right] \\ &= \mathbb{E}_t\left[e^{\frac{1}{2}(p^2-p)(\langle X\rangle_T-\langle X\rangle_t)}\right],\end{aligned}$$

so, with $\lambda = \frac{1}{2}(p^2 - p)$, we get the result. □

Corolary 3.8 *For each positive integer n,*

$$\mathbb{E}_t(\langle X\rangle_T^n) = \mathbb{E}_t\left(\frac{\partial G_{exp}}{\partial\lambda}\left(\frac{S_T}{\beta_T}, \frac{S_t}{\beta_t}, \langle X\rangle_t, \lambda\right)\Big|_{\lambda=0}\right),$$

where $G_{exp}(S, u, q, \lambda) := e^{\lambda q}\big(\frac{S}{u}\big)^p$. In particular, for $n = 1$, we have

$$\mathbb{E}_0(\langle X\rangle_T) = \mathbb{E}_0\Big(-2X_T + 2\frac{S_T}{S_0} - 2\Big).$$

PROOF Just take n derivative of (3.31) with respect to λ and evaluate at $\lambda = 0$. Differentiation through the expectations is justified by the boundedness of $\langle X\rangle_T$ and the analyticity of the moment generating function of X_T.

□

The mixing formula

To quantify the impact of correlation, Carr and Lee (Proposition 9) give a *mixing formula* that (without assuming independence) expresses the value of any European-style payoff as the expectation of the Black-Scholes formula for that payoff, evaluated at a randomized stock price and random volatility. The parameter ρ appears explicitly in the mixing

formula's randomized arguments, enabling them to examine the formula's correlation-sensitivity and to choose a function G such that $\mathbb{E}(G(S_T))$ has zero sensitivity to correlation perturbations. To prove this formula, they use the argument due to Romano-Touzi, but in a slightly more general setting, without assuming that instantaneous volatility follows a 1-factor diffusion. Here we extend their result to a framework of stochastic interest rate by taking the quotient S_t/β_t and conditioning on the volatility path. We obtain an expression for a European option price as the expectation of the Black-Scholes formula, with the volatility parameter replaced by the the square root of average future realized variance $\frac{1}{T-t}\int_t^T \sigma_u^2 du$ and the spot price replaced by S_t/β_t.

Proposition 3.9 *Without assuming independence between σ and W, let*

$$dS_t = r_t S_t dt + \sigma_t S_t\big(\sqrt{1-\rho^2}dW_t^1 + \rho dW_t^2\big)$$

where $|\rho| \leq 1$, and W^1 and W^2 are $\mathcal{F}_t$-brownian motions, and ρ and W^2 are adapted to some filtration $\mathcal{H}_t \subseteq \mathcal{F}_t$, where $\mathcal{H}_T$ and $\mathcal{F}_T^{W^1}$ are independent. Then

$$\mathbb{E}_t F\left(\frac{S_T}{\beta_T}\right) = \mathbb{E}_t F^{BS}\left(\frac{S_t}{\beta_t}M_{t,T}(\rho), t, \nu_{t,T}(1-\rho^2), 1, 0\right),$$

where

$$M_{t,T}(\rho) = e^{-\frac{\rho^2}{2}\int_t^T \sigma_u^2 du + \rho\int_t^T \sigma_u dW_u^2} \quad \textit{and} \quad \nu_{t,T} = \int_t^T \sigma_u^2 du\ .$$

Proof Applying Itô's rule,

$$\begin{aligned} d\log(Y_t) &= d\log(S_t) - d\log(\beta_t) \\ &= -\frac{1}{2}\sigma_t^2 dt + r_t dt + \sigma_t\big(\sqrt{1-\rho^2}dW_t^1 + \rho dW_t^2\big) - r_t dt \\ &= -\frac{1-\rho^2}{2}\sigma_t^2 dt - \frac{\rho^2}{2}\sigma_t^2 dt + \sigma_t\big(\sqrt{1-\rho^2}dW_t^1 + \rho dW_t^2\big). \end{aligned}$$

So conditional on $\mathcal{H}_T \vee \mathcal{F}_t$,

$$Y_T \sim \mathcal{N}\left(X_t + \log M_{t,T}(\rho) - \nu_{t,T}\frac{1-\rho^2}{2}, \nu_{t,T}(1-\rho^2)\right)$$

Hence the time-t expectation of $F(S_T)$ claim is

$$\mathbb{E}_t F(Y_T) = \mathbb{E}_t(\mathbb{E}_t(F(Y_T)|\mathcal{H}_T)) = \mathbb{E}_t F^{BS}\big(Y_t M_{t,T}(\rho), t, \nu_{t,T}(1-\rho^2), 1, 0\big).$$

□

So far, we have shown that all the results are extensible as long as we evaluate the payoff in S_t/β_t. What is not clear is whether the resulting expression can be replicated. Fortunately, the following financial instruments make this possible.

3.5.2 Exchange Options

Definition 3.10 *(Exchange Option) An exchange option (X-option) gives its owner the right to exchange b units of one asset into a units of another asset at a specific point in time, i.e., it is a claim that pays off*

$$(aS_1(T) - bS_2(T))^+.$$

REMARK Let us assume momentarily that the interest rate r is constant and that the underlying assets follow correlated ($dW^1 dW^2 = \rho dt$) geometric brownian motions under the risk-neutral measure,

$$dS_i = \mu_i S_i dt + \sigma_i S_i dW^i.$$

The time-t value of the exchange option is

$$F^M(aS_1(t), bS_2(t), t, \sigma_q),$$

where F^M is the Margrabe's formula, given by

$$F^M(S_1, S_2, t, \sigma_q) := S_1 e^{(\mu_1 - r)(T-t)}\Phi(d_+) - S_2 e^{(\mu_2 - r)(T-t)}\Phi(d_-), \tag{3.32}$$

with

$$d_\pm = \frac{\ln(S_1/S_2) + (\mu_1 - \mu_2 \pm \sigma_q^2/2)(T-t)}{\sigma_q\sqrt{T-t}},$$

and

$$\sigma_q = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

It is interesting to note that this formula is independent of the short rate if $\mu_1 = \mu_2 = r$. This is because after risk adjustment, both underlying indexes increase at the same rate and the drifts offset each other. In this case,

$$F^M(S_1, S_2, t, \sigma) = F^{BS}\left(\frac{S_1}{S_2 e^{-(\mu_2 - \mu_1)(T-t)}}, t, \sigma_q, S_2, r - \mu_2\right).$$

Formula (3.32) remains valid in a context of stochastic interest rate as long as the factors that drive interest rates are independent of those driving the S assets, which is obviously not our case.

Combining the mixing formula with the fact that the value of an X-option can be viewed as the price of a European call option with zero interest rate, strike 1, we get

$$p = \beta_t \mathbb{E}_t F(Y_T) = \beta_t \mathbb{E}_t\left(\frac{S_T}{\beta_T} - 1\right)^+ = \beta_t \mathbb{E}_t F^{BS}\left(\frac{S_t}{\beta_t} M_{t,T}(\rho), t, \nu_{t,T}(1-\rho^2), 1, 0\right).$$

3.5.3 Replication of exponential variance by means of Exchange Options

Following the line of reasoning in Carr and Lee, the price of a general European payoff on realized variance can be decomposed as a weighted sum of contracts with exponential payoffs on variance, each priced according to the previous proposition.

Under the assumptions in place, the value of the contract only depends on the volatility process of $\frac{S_t}{\beta_t}$, which is a martingale with volatility process and initial value equal to those of S_t. Therefore the price under consideration is independent of the properties of the interest rate process. However, hedging performance, which depends on the nature of instruments available for hedging, might depend on the properties of the interest rate process. The pricing representation based on $\mathbb{E}_0\left[(\frac{S_T}{\beta_T})^p\right]$ is not immediately useful in hedging because its interpretation as the price of a financial contract is not readily apparent. Fortunately, we can replicate it as a position not in plain vanilla options but in X-options:

Proposition 3.11 *The pricing factor* $\mathbb{E}_0\left[\left(\frac{S_T}{\beta_T}\right)^p\right]$ *can be replicated through a position in p forward contracts plus a portfolio of Exchange Options on the assets* S_T *and* β_T*, with notional equal to* $p(p-1)K^{p-2}$.

Proof Using Proposition 3.5 applied to the function $f(t) := t^p$, we get

$$Y_T^p = 1 + p(Y_T - 1) + \int_0^1 p(p-1)K^{p-2}(K - Y_T)^+ dK + \int_1^\infty p(p-1)K^{p-2}(Y_T - K)^+ dK$$

Doing $Y_T = S_T/\beta_T$ and taking expectation in this identity yields

$$\begin{aligned}
\mathbb{E}_0\left[\left(\frac{S_T}{\beta_T}\right)^p\right] &= 1 + p(S_0 - 1) + \int_0^1 p(p-1)K^{p-2}\mathbb{E}_0\left[\left(K - \frac{S_T}{\beta_T}\right)^+\right]dK + \\
&\quad + \int_1^\infty p(p-1)K^{p-2}\mathbb{E}_0\left[\left(\frac{S_T}{\beta_T} - K\right)^+\right]dK \\
&= 1 + p(S_0 - 1) + \int_0^1 p(1-p)K^{p-2}\mathbb{E}_0\left[\frac{(S_T - \beta_T K)^+}{\beta_T}\right]dK \\
&\quad + \int_1^\infty p(p-1)K^{p-2}\mathbb{E}_0\left[\frac{(S_T - K\beta_T)^+}{\beta_T}\right]dK.
\end{aligned}$$

The expectation inside both integrals is the price of an X-option with $S_1(T) = S_T, S_2(T) = \beta_T, a = 1, b = K$. □

Then, if an X-option market between S and the bank account β is available, the information in T-expiry option prices fully reveals the risk-neutral distribution not only of the discounted price S_T/β_T, but also of variance $\langle X \rangle_T$. Given the previous identity, all the Carr-Lee framework is extensible to stochastic rates using the same techniques they use, but evaluating the payoff in S_t/β_t instead of S_t.

If exchange options of all strikes are available for hedging, then this can be accomplished in a perfect manner. If not, some natural questions arise. How can we optimally hedge if there are exchange options for only a few strikes, and a much finer grid for standard calls and puts? In order to see this, compare the payoff delivered by an exchange option

$$(S_T - K\beta_T)^+$$

with the payoff delivered by a standard option

$$(S_T - K)^+.$$

Assuming that $\mathbb{E}[\beta_T]$ is known from the bond market, we can approximate an exchange option by a standard call with fixed strike by taking a payoff of the form

$$(S_T - K\mathbb{E}[\beta_T])^+.$$

But if some exchange options are available, perhaps we should buy those in larger amount than that indicated by the exact decomposition to guarantee that the correlation between rates and the stock is hedged optimally, and buy less of the non-available strikes which are approximately hedged by standard calls and puts.

Chapter 4

Monte Carlo Simulation: Variance Reduction Techniques

The term "Monte Carlo" was apparently first used by Ulam and von Neumann as a Los Alamos code word for the stochastic simulations they applied to build better atomic bombs. Their methods, involving the laws of chance, were aptly named after the international gaming destination. Despite the widespread use of the methods, and numerous descriptions of them in articles and monographs, it is virtually impossible to find a succint definition of "Monte Carlo method" in the literature. Perhaps this is owing to the intuitive nature of the topic which spawns many definitions by way of specific examples. Some authors prefer to use the term "stochastic simulation" for almost everything, reserving "Monte Carlo" only for Monte Carlo Integration and Monte Carlo Tests (cf. Ripley 1987). Others seem less concerned about blurring the distinction between simulation studies and Monte Carlo methods. Be that as it may, a suitable definition can be good to have, if for nothing other than to avoid the awkwardness of trying to define the Monte Carlo method by appealing to a whole bevy of examples of it. In an attempt to fulfill this task, a satisfactory definition of Monte Carlo would be, in our opinion, the one that reads as *the art of approximating an expectation by the sample mean of a function of simulated random variables*. We will find that this definition is broad enough to cover everything that has been called Monte Carlo, and yet makes clear its essence in very familiar terms: Monte Carlo is about invoking laws of large numbers to approximate expectations. While most Monte Carlo simulations are done by computer today, there were many applications of Monte Carlo methods using coin-flipping, card-drawing, or needle-tossing (rather than computer generated pseudo-random numbers) as early as the turn of the century long before the name Monte Carlo arose.

4.1 The idea behind Monte Carlo and some basic results

Consider a (possibly multidimensional) random variable X having probability mass function or probability density function $f_X(x)$ which is greater than zero on a set of values X. Then the expected value of a function g of X is

$$\mathbb{E}(h(X)) = \sum_{x \in \Omega} h(x) f_X(x) \tag{4.1}$$

if X is discrete, and

$$\mathbb{E}(h(X)) = \int_{\Omega} h(x) f_X(x) dx \tag{4.2}$$

if X is continuous. Now, if we were to take an n-sample of X's, $(x_1, ..., x_n)$, and we computed the mean of $h(x)$ over the sample, then we would have the Monte Carlo estimate

$$\tilde{h}_n(x) = \frac{1}{n} \sum_{i=1}^{n} h(x_i) \tag{4.3}$$

of $\mathbb{E}(h(X))$. We could, alternatively, speak of the random variable

$$\tilde{h}_n(X) = \frac{1}{n} \sum_{i=1}^{n} h(X) \tag{4.4}$$

which we call the Monte Carlo estimator of $\mathbb{E}(h(X))$. If $\mathbb{E}(h(X))$ exists, then the weak law of large numbers tells us that for any arbitrarily small ϵ

$$\lim_{n \to \infty} P(|\tilde{h}_n(X) - \mathbb{E}(h(X))| \geq \epsilon) = 0. \tag{4.5}$$

This tells us that as n gets large, then there is small probability that $\tilde{h}_n(X)$ deviates much from $\mathbb{E}(h(X))$. For our purposes, the strong law of large numbers says much the same thing (the important part being that so long as n is large enough, $\tilde{h}_n(x)$ arising from a Monte Carlo experiment shall be close to $\mathbb{E}(h(X))$, as desired). One other thing to note at this point is that $\tilde{h}_n(X)$ is unbiased for $\mathbb{E}(h(X))$:

$$\mathbb{E}(\tilde{h}_n(X)) = \mathbb{E}\left(\frac{1}{n} \sum_{i=0}^{n} h(X_i)\right) = \frac{1}{n} \sum_{i=0}^{n} \mathbb{E}(h(X_i)) = \mathbb{E}(h(X)). \tag{4.6}$$

The preceding section comes to life and becomes useful when one realizes that many quantities of interest may be cast as expectations. Most importantly for applications in statistical genetics, it is possible to express all probabilities, integrals, and summations as expectations.

4.2 Simulation Efficiency

Suppose as usual that we wish to estimate $\theta := \mathbb{E}[h(X)]$. Then the standard simulation algorithm is:

1. Generate $X_1, \ldots, X_n$
2. Estimate θ with $\hat{\theta}_n := \sum_{i=1}^n Y_j/n, Y_j := h(X_j)$
3. Approximate $100(1-\alpha)\%$ confidence intervals by

$$\left[\widehat{\theta} - z_{1-\alpha/2}\frac{\widehat{\sigma}_n}{\sqrt{n}}, \widehat{\theta} + z_{1-\alpha/2}\frac{\widehat{\sigma}_n}{\sqrt{n}}\right],$$

where $\hat{\sigma}_n$ is the usual estimate of $\mathsf{Var}(Y)$ based on $Y_1, ..., Y_n$. One way to measure the quality of the estimator, $\hat{\theta}_n$, is by the half-width, HW, of the confidence interval. For a fixed α, we have

$$HW = z_{1-\alpha/2}\sqrt{\frac{\mathsf{Var}(Y)}{n}}.$$

We would like HW to be small, but sometimes this is difficult to achieve. This may be because $\mathsf{Var}(Y)$ is too large, or too much computational effort is required to simulate each Y_j so that n is necessarily small, or some combination of the two. As a result, it is often imperative to address the issue of simulation efficiency. There are a number of things we can do:

1. Develop a good simulation algorithm
2. Program carefully to minimize storage requirements. For example we do not need to store all the Y_j 's: we only need to keep track of $\sum Y_j$ and $\sum Y_j^2$ to compute $\hat{\theta}_n$ and approximate CI's
3. Program carefully to minimize execution time
4. Decrease the variability of the simulation output that we use to estimate θ. The techniques used to do this are usually called *variance reduction techniques*

In the next section we will study various variance reduction techniques, and assume that we are doing items (1) to (3) as well as possible. Before proceeding to study these techniques, however, we should first describe a *measure of simulation efficiency.*

4.2.1 Measuring simulation efficiency

Suppose there are two random variables, W and Y, such that $\mathbb{E}[W] = \mathbb{E}[Y] = \theta$. Then we could choose to either simulate $W_1, \ldots W_n$ or $Y_1, \ldots, Y_n$ in order to estimate θ. Let M_ω denote the method of estimating θ by simulating the W_i's. M_y is similarly defined. Which method is more efficient, M_ω or M_y? To answer this, let n_ω and n_y be the number of samples of W and Y, respectively, that are needed to achieve a half-width, HW. Then it is easy to see that

$$n_\omega = \left(\frac{z_{1-\alpha/2}}{HW}\right)^2 \mathsf{Var}(W)$$

$$n_y = \left(\frac{z_{1-\alpha/2}}{HW}\right)^2 \mathsf{Var}(Y).$$

Let E_ω and E_y denote the amount of computational effort required to produce one sample of W and Y, respectively. Then the total effort expended by M_ω to achieve a half-width HW is

$$TE_\omega = \left(\frac{z_{1-\alpha/2}}{HW}\right)^2 \mathsf{Var}(W)E_\omega.$$

Similarly, the effort expended by M_y to achieve the same half-width HW, is given by

$$TE_y = \left(\frac{z_{1-\alpha/2}}{HW}\right)^2 \mathsf{Var}(W)E_y.$$

We say that M_ω is more efficient than M_y if

$$TE_\omega < TE_y.$$

Note that $TE_\omega < TE_y$ if and only if

$$\mathsf{Var}(W)E_\omega < \mathsf{Var}(Y)E_y. \tag{4.7}$$

We will use the quantity $\mathsf{Var}(W)E_\omega$ as a measure of the efficiency of the simulator, M_ω. Note that expression (4.7) implies that we cannot conclude that one simulation algorithm, M_ω, is better than another, M_y, simply because $\mathsf{Var}(W) < \mathsf{Var}(Y)$; we also need to take E_ω and E_y into consideration. However, it is often the case that we have two simulators available to us, M_w and M_y, where $E_\omega \approx Ey$ and $\mathsf{Var}(W) << \mathsf{Var}(Y)$. In such cases it is clear that using M_ω provides a substantial improvement over using M_y.

4.3 Variance Reduction Techniques

Variance reduction is the search for alternative and more accurate estimators of a given quantity. The possibility of variance reduction is what separates Monte Carlo from direct simulation. Simple variance reduction methods often are remarkably effective and easy to implement. It is good to think about them as you wait for a long Monte Carlo computation to finish. In some applications, such as rare event simulation and quantum chemistry, they make practical what would be impossible otherwise. Most advanced Monte Carlo is some kind of variance reduction. Among the many variance reduction techniques, which may be used in combination, are *control variates, partial integration, systematic sampling* and *importance sampling*. The method of control variates is useful when a crude version of the problem can be solved explicitly. This is often the case in simple problems (possibly the definition of "simple") such pricing problems in quantitative finance where the crude solvable version could be Black-Scholes. Partial integration, also called *Rao Blackwellization* or *Conditional Monte Carlo*, lowers variance by replacing integrals over some variables or over parts of space by their averages. *Systematic sampling* methods range from the simplest, *antithetic variates*, to the slightly more sophisticated *stratified sampling*, to *quasi Monte Carlo integration*. *Importance sampling* has appeared already as sampling with a weight function. It also is the basis of reweighting and score function strategies for sensitivity analysis. Methods for rare event sampling mostly use importance functions, often suggested by the mathematical theory of large deviations.

Going back to our original notation, we have the random variable $\tilde{h}_n(X)$, a Monte Carlo estimator of $\mathbb{E}(h(X))$. Like all random variables, we may compute its variance (if it exists) by the standard formula

$$\mathsf{Var}(\tilde{h}_n(X)) = \frac{\mathsf{Var}(h(X))}{n} = \frac{1}{n}\int_\Omega [h(x) - \mathbb{E}(h(X))]^2 f_X(x)dx.$$

4.3.1 Conditional Monte Carlo

We will now consider the variance reduction technique known as *conditional Monte Carlo*. The idea here is very simple: we use our knowledge about the system being studied to reduce the variance of our estimator. As usual, suppose we wish to estimate $\theta = \mathbb{E}[h(X)]$ where $X = (X_1, ..., X_m)$. If we could compute θ analytically, then this would be equivalent to solving an m-dimensional integral. However, maybe it is possible to evaluate part of the integral analytically. If so, then we might be able to use simulation to estimate the other part and thereby obtain a variance reduction. The vehicle that we use to do part of

the integration analytically is the concept of conditional expectation. Before we describe the method in detail, we will briefly review conditional expectations and variances.

Conditional Expectations and Variances

Let X and Z be random vectors, and let $Y = h(X)$ be a random variable. Suppose we set $V = \mathbb{E}(Y|Z)$. Then V is itself a random variable that depends on Z, so that we may write $V = g(Z)$ for some function, $g(\cdot)$. We also know that

$$\mathbb{E}[V] = \mathbb{E}[\mathbb{E}(Y|Z)] = \mathbb{E}[Y], \tag{4.8}$$

so that if we are trying to estimate $\mu = \mathbb{E}[Y]$, one possibility would be to simulate V instead of simulating Y. In order to decide which would be a better estimator of μ, it is necessary to compare the variances of Y and $\mathbb{E}[Y|Z]$. To do this, recall the *conditional variance formula*:

$$\mathsf{Var}(Y) = \mathbb{E}[(\mathsf{Var}(Y|Z)] + \mathsf{Var}[\mathbb{E}(Y|Z)]. \tag{4.9}$$

Now $\mathsf{Var}(Y|Z)$ is also a random variable that depends on Z, and since a variance is always non-negative, it must follow that $\mathbb{E}[\mathsf{Var}(Y|Z)] \geq 0$. But then (4.9) implies

$$\mathsf{Var}(Y) \geq \mathsf{Var}(\mathbb{E}(Y|Z)), \tag{4.10}$$

so we can conclude that V is a better estimator of μ than Y. To see this from another perspective, suppose that to estimate μ we first have to simulate Z and then simulate Y given Z. If we can compute $\mathbb{E}(Y|Z)$ exactly, then we can eliminate the additional noise that comes from simulating Y given Z, thereby obtaining a variance reduction.

As a practical example, suppose (X, Y) is a random vector with probability density $f(x, y)$. Let $G(X, Y)$ be a random variable and

$$\widetilde{G}(x) = \mathbb{E}(G(X, Y)|x) = \frac{\int_\Omega G(x, y)f(x, y)dy}{\int_\Omega f(x, y)dy}.$$

A simple inequality shows that except in the trivial case where G already was independent of y,

$$\mathsf{Var}(\widetilde{G}) < \mathsf{Var}(G)$$

A more general version of the partial averaging method is that if $\mathcal{G}$ is a sub σ-algebra and

$$\widetilde{G} = \mathbb{E}(G|\ \mathcal{G}),$$

then we have

$$\mathsf{Var}(G) = \mathsf{Var}(\widetilde{G}) + \mathbb{E}\big[(G - \widetilde{G})^2\big],$$

and thus

$$\mathsf{Var}(G) < \mathsf{Var}(\widetilde{G}),$$

except in the trivial case that G is $\mathcal{G}$-measurable. The conclusion is that if a problem can be solved partially, if some of the integrals above can be computed explicitly, the remaining problem is easier.

Finally, we point out that in order for the conditional expectation method to be worthwhile, it must be the case that Y and Z are *dependent*.

The conditional Monte Carlo Simulation Algorithm

We want to estimate $\theta := \mathbb{E}[g(X)] = \mathbb{E}[Y]$ using conditional Monte Carlo. To do so, we must have another variable or vector, Z, that satisfies the following requirements:

1. Z can be easily simulated
2. $V := g(Z) := \mathbb{E}(Y|Z)$ can be computed exactly

This means that we can simulate a value of V by first simulating a value of Z and then setting $V = g(Z) = \mathbb{E}(Y|Z)$. We then have the following algorithm for estimating μ. It is also possible that other variance reduction methods could be used in conjunction with conditioning. For example, it might be possible to use *stratification*.

The following algorithm implements this idea to estimate this expected value:

for $i = 1$ to n

- generate Z_i
- compute $g(Z_i) = \mathbb{E}(Y|Z_i)$
- set $V_i = g(Z_i)$
- end for
- set $\hat{\theta}_{n,cm} = \bar{V}_n = \frac{1}{n}\sum_{i=1}^{n} V_i$
- set $\hat{\sigma}^2_{n,cm} = \frac{1}{n-1}\sum_{i=1}^{n}(V_i - \overline{V}_n)^2$

It is also possible that other variance reduction methods could be used in conjunction with conditioning. For example, it might be possible to use importance sampling or stratification, which will be in fact our approach.

4.3.2 Stratification

Stratified sampling[1] refers broadly to any sampling mechanism that constrains the fraction of observations drawn from specific subsets or strata of the sample space. Suppose that our goal is to estimate $\mathbb{E}(Y)$ with Y real-valued, and let $A_1, ..., A_K$ a partition of the real line. Then

$$\mathbb{E}(Y) = \sum_{i=1}^{K} P(Y \in A_i)\mathbb{E}(Y|Y \in A_i) = \sum_{i=1}^{K} p_i \mathbb{E}(Y|Y \in A_i), \tag{4.11}$$

with $p_i = P(Y \in A_i)$. In a random sampling, we generate independent $Y_1, ..., Y_n$ having the same distribution as Y. The fraction of this sample falling in A_i will not in general equal p_i though it would approach p_i as the sample size n increases. In stratified sampling, we decide in advance what fraction of the samples should be drawn from each stratum A_i; each observation drawn from this subset is constrained to have the distribution of Y conditional on $Y \in A_i$.

The simplest case is proportional sampling, in which we ensure that the fraction of observations drawn from stratum A_i matches the theoretical probability $p_i = P(Y \in A_i)$. If the total sample size is n, this entails generating $n_i = np_i$ samples from A_i. For each $i = 1, \ldots, K$, let the $Y_{ij}, j = 1, \ldots, n_i$ be independent draws from the conditional distribution of Y given $Y \in A_i$. An unbiased estimator of $\mathbb{E}(Y|\ Y \in A_i)$ is provided by the sample mean $(Y_{i1} + \ldots + Y_{in_i})/n_i$ of observations from the ith stratum. It follows from (4.11) that an unbiased estimator of $\mathbb{E}(Y)$ is provided by

$$\hat{Y} = \sum_{i=1}^{K} p_i \cdot \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n} \sum_{i=1}^{K} \sum_{j=1}^{n_i} Y_{ij}. \tag{4.12}$$

This estimator should be contrasted with the usual sample mean $\bar{Y} = (Y_1 + \ldots + Y_n)/n$ of a random sample of size n. Compared with $\bar{Y}$, the stratified estimator $\hat{Y}$ eliminates sampling variability across strata without affecting sampling variability within strata.

From this introduction it should be clear that the use of stratified sampling involves consideration of two issues:

- choosing the variable X, the strata $A_1, \ldots, A_K$, and the allocation $n_1, \ldots, n_K$
- generating samples from the distribution of (X, Y) conditional on $X \in A_i$.

[1] For this section, we followed chapter 4, section 3 of [Gl].

In addressing the first issue, we will see that stratified sampling is more effective when the variability of Y within each stratum is small. For solutions to the second issue, examples in Glasserman [Gl] can be consulted.

Output analysis

We turn back to the original problem we stated: the estimation of an expected value through stratification. With the previous notation, let

$$\mu_i = \mathbb{E}(Y_{ij}) = \mathbb{E}(Y|X \in A_i) \tag{4.13}$$
$$\sigma_i^2 = \mathsf{Var}(Y_{ij}) = \mathsf{Var}(Y|X \in A_i) \tag{4.14}$$

Let $p_i = P(X \in A_i)$ denote the stratum probabilities. Fix an allocation $n_1, ..., n_K$ with all $n_i \geq 1$ and $n_1 + ... + n_K = n$. Let $q_i = n_i/n$ denote the fraction of samples allocated to the ith stratum. For any such allocation the estimator Y in (4.11) is unbiased because

$$\mathbb{E}(Y) = \sum_{i=1}^{K} p_i \cdot \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbb{E}(Y_{ij}) = \sum_{i=1}^{K} p_i \mu_i = \mu. \tag{4.15}$$

The variance of $\hat{Y}$ is given by

$$\mathsf{Var}(\hat{Y}) = \sum_{i=1}^{K} p_i^2 \mathsf{Var}\left(\frac{1}{n_i} \sum_{j=1}^{n_i} Y_i j\right) = \sum_{i=1}^{K} p_i^2 \frac{\sigma_i^2}{n_i} = \frac{\sigma(q)^2}{n}, \tag{4.16}$$

with

$$\sigma^2(q) = \sum_{i=1}^{K} \frac{p_i^2}{q_i} \sigma_i^2. \tag{4.17}$$

For each stratum A_i, the samples Y_{i1}, Y_{i2},... are i.i.d. with mean μ_i and variance σ_i^2 and thus satisfy

$$\frac{1}{\sqrt{\lfloor nq_i \rfloor}} \sum_{j=1}^{\lfloor nq_i \rfloor} (Y_{ij} - \mu_i) \xrightarrow[n\to\infty]{} \mathcal{N}(0, \sigma_j^2), \tag{4.18}$$

with $q_1, ..., q_K$ fixed. The centered and scaled estimator $\sqrt{n}(\hat{Y} - \mu)$ can be written as

$$\begin{aligned} \sqrt{n}(\hat{Y} - \mu) &= \sqrt{n} \sum_{i=1}^{K} p_i \left(\frac{1}{\lfloor nq_i \rfloor} \sum_{j=1}^{\lfloor nq_i \rfloor} (Y_{ij} - \mu_i) \right) \\ &\approx \sum_{i=1}^{K} \frac{p_i}{\sqrt{q_i}} \left(\frac{1}{\lfloor nq_i \rfloor} \sum_{j=1}^{\lfloor nq_i \rfloor} (Y_{ij} - \mu_i) \right), \end{aligned}$$

the approximation holding in the sense that the ratio of the two expressions approaches 1 as $n \to \infty$. This shows that $\sqrt{n}(\hat{Y} - \mu)$ is asymptotically a linear combination of independent normal variables (with mean 0 and variances σ_i^2). It follows that

$$\sqrt{n}(\hat{Y} - \mu) \longrightarrow \mathcal{N}(0, \sigma_j^2),$$

with $\sigma^2(q)$ as defined in (4.17). This limit holds as the sample size n increases with the number of strata K held fixed.

Optimal allocation

In the case of a proportional allocation of samples to strata, $q_i = p_i$ and the variance parameter $\sigma^2(q)$ simplifies to

$$\sum_{i=1}^{K} \frac{p_i^2}{q_i}\sigma_i^2 = \sum_{i=1}^{K} p_i \sigma_i^2. \tag{4.19}$$

To compare this to the variance without stratification, observe that

$$\mathbb{E}(Y^2) = \sum_{i=1}^{K} p_i \mathbb{E}(Y^2 | X \in A_i) = \sum_{i=1}^{K} p_i(\sigma_i^2 + \mu_i^2), \tag{4.20}$$

so using $\mu = \sum_{i=1}^{K} p_i \mu_i$, we get

$$\mathsf{Var}(Y) = \mathbb{E}(Y^2) - \mu^2 = \sum_{i=1}^{K} p_i \sigma_i^2 \sum_{i=1}^{K} p_i \mu_i^2 - \left(\sum_{i=1}^{K} p_i \mu_i \right)^2. \tag{4.21}$$

By Jensen's inequality,

$$\sum_{i=1}^{K} p_i \mu_i^2 \geq \left(\sum_{i=1}^{K} p_i \mu_i \right)^2, \tag{4.22}$$

with strict inequality unless all μ_i are equal. Thus, comparing with (4.19), we conclude that *stratified sampling with a proportional allocation can only decrease variance.*

Variance decomposition

The preceding discussion considers the allocation of samples to given data. In order to consider the question of how strata should be selected in the first place, we now examine what part of the variance of Y is removed through stratification of X. As before, let $A_1, \ldots, A_K$ be strata for X. Let $\eta \equiv \eta(X) \in \{1, ..., K\}$ denote the index of the stratum containing X, so that $X \in A_\eta$. We can always write

$$Y = \mathbb{E}(Y|\eta) + \varepsilon$$

simply by defining the residual ϵ so that equality holds. Applying (4.9) with $Z := \eta$, we find that

$$\mathsf{Var}(Y) = \mathsf{Var}(\mathbb{E}(Y|\eta)) + \mathbb{E}(\mathsf{Var}(Y|\eta)). \tag{4.23}$$

As $\mathbb{E}(Y|\eta = i) = \mu_i$ and $P(\eta = i) = p_i$, the first term on the right hand side of (4.23) is

$$\mathsf{Var}(\mathbb{E}(Y|\eta)) = \sum_{i=1}^{K} p_i \mu_i^2 - \left(\sum_{i=1}^{K} p_i \mu_i \right)^2, \tag{4.24}$$

from which we conclude that

$$\mathsf{Var}(\mathbb{E}(Y|\eta)) = \mathbb{E}(\mathsf{Var}(Y|\eta)) = \sum_{i=1}^{K} p_i \sigma_i^2,$$

which is precisely the variance parameter in (4.19) for stratified sampling with proportional allocation. This confirms that the variance parameter of the stratified estimator is the variance of the residual of Y after conditioning on η. Identity (4.24) shows that stratification eliminates inter-stratum variability, leaving only intra-stratum variability. Another consequence of it, is that further stratification results in further variance reduction. Suppose the partition $\{\tilde{A}_1, \ldots, \tilde{A}_M\}$ refines the partition $\{A_1, \ldots, A_K\}$, in the sense that the stratum index $\tilde{\eta}$ of the new partition completely determines η. Then, $\mathbb{E}(Y|\eta) = \mathbb{E}\big(\mathbb{E}(Y|\eta)|\tilde{\eta}\big)$ and Jensen's inequality yields

$$\mathsf{Var}(\mathbb{E}(Y|\eta)) \leq \mathsf{Var}(\mathbb{E}(Y|\tilde{\eta})),$$

from which it follows that the residual variance from the refined strata cannot exceed the residual variance from the original strata.

Stratifying a linear projection

Suppose that $\xi \sim \mathcal{N}(\mu, \Sigma)$ in $\mathbb{R}^d$ and that we want to generate ξ with $X \equiv v^\top \xi$ stratified for some fixed vector $v \in \mathbb{R}^d$. Suppose the $d \times d$ matrix Σ has full rank. Without loss of generality, we may take μ to be the zero vector. Also, stratifying X is equivalent to stratify any multiple of X. Thus, by scaling v if necessary, we may assume that $v^\top \Sigma v = 1$. Then,

$$X = v^\top \xi \sim \mathcal{N}(0, v^\top \Sigma v) = \mathcal{N}(0, 1), \tag{4.25}$$

so we can stratify X by the following method: take K equiprobable strata with proportional allocation of the set $[0, 1]$. Let $U_1, ..., U_K$ be independent $\mathsf{Unif}[0, 1]$ random variables and set

$$V_i = \frac{i-1}{K} + \frac{U_i}{K}, \qquad i = 1, ..., K.$$

Then $\Phi^{-1}(V_1), \ldots, \Phi^{-1}(V_K)$ form a stratified sample from the standard normal.
The next step is to generate ξ conditional in the value of X. First observe that ξ and X are jointly normal with

$$\begin{pmatrix} \xi \\ X \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \Sigma & \Sigma v \\ v^\top \Sigma & v^\top \Sigma v \end{pmatrix} \right) \tag{4.26}$$

Using the *Conditioning Formula* for a multivariate normal partitioned vector (see [Gl], p.65), we find that

$$(\xi | X = x) \sim \mathcal{N}\left(\frac{\Sigma v}{v^\top \Sigma v} x, \Sigma - \frac{\Sigma v v^\top \Sigma}{v^\top \Sigma v} \right) = \mathcal{N}\big(\Sigma v x, \Sigma - \Sigma v v^\top \Sigma\big). \tag{4.27}$$

Observe that the conditional covariance matrix does not depend on x. This is important because it means that only a single factorization is required for the conditional sampling. Let A be any matrix for which $AA^\top = \Sigma$ and observe that

$$\begin{aligned} (A - \Sigma v v^\top A)(A - \Sigma v v^\top A)^\top &= AA^\top - AA^\top v v^\top \Sigma - \Sigma v v^\top AA^\top + \Sigma v v^\top \Sigma v v^\top \Sigma \\ &= \Sigma - \Sigma v v^\top \Sigma, \end{aligned}$$

again using that $v^\top \Sigma v = 1$. Thus, we can use the matrix $A - \Sigma v v^\top A$ to sample from the conditional distribution of ξ given X. The following algorithm generates K samples from $\mathcal{N}(0, \Sigma)$ stratified along the direction determined by v:
for $i = 1, ..., K$

- generate $U \sim \mathsf{Unif}[0, 1]$
- set $V = \frac{i-1+U}{K}$
- set $X = \Phi^{-1}(V)$
- generate $\mathbf{Z} \sim \mathcal{N}(0, I)$ in $\mathbb{R}^d$
- set $\xi = \Sigma v X + (A - \Sigma v v^\top A)\mathbf{Z}$

By construction, of the K values of X generated by this algorithm, exactly one will fall in each of K equiprobable strata for the standard normal distribution. But observe that under this construction,

$$v^\top \xi = v^\top \Sigma v X + v^\top (A - \Sigma v v^\top A) Z = X.$$

Thus, of the K values of ξ generated, exactly one has a projection $v^\top \xi$ falling into each of K equiprobable strata. In this sense, the algorithm generates samples from $\mathcal{N}(0,\Sigma)$ stratified along the direction determined by v.

Further simplification is possible in stratifying a sample from the standard multivariate normal distribution $\mathcal{N}(0,I)$. In this case, the construction above becomes

$$\xi = vX + (I - vv^\top)Z, \qquad X \sim \mathcal{N}(0,1), \qquad Z \sim \mathcal{N}(0,I),$$

with v now normalized so that $v^\top v = 1$. Since $X = v^\top \xi$, by stratifying X we stratify the projection of ξ onto v.

Optimal stratification direction

In estimating $\mathbb{E}(f(\xi))$, with $\xi \sim \mathcal{N}(\mu,\Sigma)$, it would be convenient to know the direction v for which $v^\top \xi$ would produce the greatest reduction in variance. Finding this v is rarely possible, but for some special sort of f this is the case. With no essential loss of generality, we restrict our attention to the case $\mathbb{E}(f(Z))$ with $Z \sim \mathcal{N}(0,I)$. We know that the residual variance after stratifying a linear combination $v^\top Z$ is $\mathbb{E}(\mathsf{Var}(f(Z)|\eta))$, where η is the random index of the stratum containing $v^\top Z$. If we use equiprobable strata and let the strata grow (refining the previous set of strata) the residual variance converges to $\mathbb{E}(\mathsf{Var}(f(Z)|v^\top Z))$ (see [GHS]). We will therefore compare alternative choices of v through this limiting value.

If $f(z) = b^\top z$, it is evident that this optimal direction is given by $v = b$. For a quadratic f, we have the following result:

Proposition 4.1 *Let $f(z) = b^\top z + \frac{1}{2}(z^\top A z)$, for $b \in \mathbb{R}^N$ and $A \in \mathbb{R}^{N\times N}$. Then the optimal stratification direction is the solution to*

$$\arg \max_{\{|v|=1\}} \left[(b^\top v)^2 + \frac{1}{2}(v^\top A v)^2 \right]. \tag{4.28}$$

Proof Because of (4.9), minimizing $\mathbb{E}(\mathsf{Var}(f(Z)|v^\top Z))$ over $\{|v| = 1\}$ is equivalent to maximizing $\mathsf{Var}(\mathbb{E}(f(Z)|v^\top Z))$ over the same set. Thus, the aim is to find $v \in \mathbb{R}^N$, $|v| = 1$, that maximizes $\mathsf{Var}(\mathbb{E}(f(Z)|v^\top Z))$. This variance can be rewritten as

$$\mathsf{Var}\big(\mathbb{E}(b^\top Z|v^\top Z)\big) + \frac{1}{4}\mathsf{Var}\big(\mathbb{E}(Z^\top A Z|v^\top Z)\big) + \mathsf{Cov}\big(\mathbb{E}(b^\top Z|v^\top Z), \mathbb{E}(Z^\top A Z|v^\top Z)\big). \tag{4.29}$$

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

$$W = (v^{\top} Z)v + (\tilde{v}^{\top} Z)v^{\perp},$$

multiply by b and take expectation conditional on $v^{\top} Z$,

$$\mathbb{E}(b^{\top} Z | v^{\top} Z) = (v^{\top} Z)b^{\top} v + (b^{\top} \tilde{v})\mathbb{E}(\tilde{v}^{\top} Z) = (v^{\top} Z)b^{\top} v,$$

where the last equality holds because $v'Z$ and $\tilde{v}^{\top} Z$ are independent standard normals. Finally, take variance to get

$$\mathsf{Var}(\mathbb{E}(b^{\top} Z | v^{\top} Z)) = (b^{\top} v)^2 \mathsf{Var}(v^{\top} Z) = (b^{\top} v)^2 |v|^2 = (b^{\top} v)^2. \tag{4.30}$$

For the remaining terms, note that

$$\mathbb{E}(Z^{\top} AZ | v^{\top} Z) = (v^{\top} Z)^2 v^{\top} Av + \tilde{v}^{\top} A\tilde{v},$$

using again that $v^{\top} Z$ and $\tilde{v}^{\top} Z$ are independent standard normals. Therefore,

$$\mathsf{Var}(\mathbb{E}(Z^{\top} AZ | v^{\top} Z)) = 2(v^{\top} Av)^2, \tag{4.31}$$

and

$$\begin{aligned} \mathsf{Cov}(\mathbb{E}(b^{\top} Z | v^{\top} Z), \mathbb{E}(Z^{\top} AZ | v^{\top} Z)) &= (b^{\top} v)\mathbb{E}((v^{\top} Z)\mathbb{E}(Z^{\top} AZ | v^{\top} Z)) \\ &= (b^{\top} v)(v^{\top} Av)\mathbb{E}((v^{\top} Z)^3) = 0. \end{aligned}$$

Adding up (4.30) and (4.31), we obtain

$$\mathsf{Var}(\mathbb{E}(f(Z) | v^{\top} Z)) = (b^{\top} v)^2 + \frac{1}{2}(v^{\top} Av)^2 \tag{4.32}$$

under $|v| = 1$. □

REMARK If A is a symmetric matrix, then it has real eigenvalues $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_d$ and the second term of (4.32) is maximized by v_1 if $|\lambda_1| \geq |\lambda_d|$ and by v_d if $|\lambda_d| \geq |\lambda_1|$, where v_i is the vector associated to λ_i. In other words, the optimal stratification direction is an eigenvector of A associated with an eigenvalue of largest absolute value. The effect of optimal stratification is to reduce variance from $\sum_i \lambda_i^2$ to $\sum_i \lambda_i^2 - \max_i \lambda_i^2$.

Although simulation is unnecessary for the evaluation of $\mathbb{E}(f(Z))$ in each of these examples, a linear or quadratic function maybe useful as an approximation to a more general f and thus as a guide in selecting stratification directions.

Chapter 5

Numerical simulation algorithm for discrete variance derivatives

In this chapter, we build an algorithm that takes advantage of the fact that the value of a variance derivative arises from the stochastic nature of realized variance. We apply the tools in the previous chapter, combining conditional Monte Carlo with stratified sampling, justifying the choice of our conditioning variables. We find that the variance reduction is significant when comparing the method against naive Monte Carlo for identical computational budget. 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.

5.1 Models

5.1.1 Continuous time models

Let Z_t and W_t be independent standard brownian motions, and consider an underlying S following

$$\begin{aligned} dS_t &= S_t r dt + \sigma(S_t)\sqrt{V_t} dZ_t \\ dV_t &= \beta(V_t)dt + \gamma(V_t)\big(\rho dZ_t + \sqrt{1-\rho^2} dW_t\big), \end{aligned} \tag{5.1}$$

under the risk neutral measure and appropriate 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 (5.1) include the model by Hull and White [HW],

$$\begin{aligned} dS_t &= S_t r dt + S_t \sqrt{V_t} dZ_t \\ dV_t &= \mu V_t dt + \sigma V_t \big(\rho dz_t + \sqrt{1-\rho^2} dW_t\big), \end{aligned} \tag{5.2}$$

the model by Heston [H],

$$\begin{aligned} dS_t &= S_t r dt + S_t \sqrt{V_t} dZ_t \\ dV_t &= \kappa(\theta - V_t) dt + \sigma \sqrt{V_t} \big(\rho dz_t + \sqrt{1-\rho^2} dW_t\big), \end{aligned} \tag{5.3}$$

and the SABR model [HKL], which is usually expressed in terms of the dynamics of the underlying S and its instantaneous stochastic volatility σ

$$\begin{aligned} dS_t &= S_t r dt + S_t^b \sigma_t dZ_t \\ d\sigma_t &= \alpha \sigma_t \big(\rho dz_t + \sqrt{1-\rho^2} dW_t\big). \end{aligned} \tag{5.4}$$

The SABR model can be cast in the form of (5.1) by introducing $V_t \equiv \sigma_t^2$ and applying Ito's Lemma.

5.1.2 Discrete time schemes

In this chapter we compute realized variance from discretely recorded observations of an asset price process of the form (5.1). These recordings for the continuous time process are approximated by an Euler discretization $\hat{S}, \hat{V}$ over a deterministic time grid $t_0 < t_1 < ... < t_n$, with $t_{i+1} - t_i = \Delta = 1/252$. This grid coincides with the recording times prescribed in the terms of the contract. The simplest formulation of the Euler scheme is a first order expansion on S and V leading to

$$\begin{aligned} \hat{S}_{i+1} &= \hat{S}_i + \hat{S}_i r \Delta + \sigma(\hat{S}_i) \sqrt{V_i} \sqrt{\Delta} Z_{i+1} \\ \hat{V}_{i+1} &= \hat{V}_i + \beta(\hat{V}_i)\Delta + \gamma(\hat{V}_i) \sqrt{\Delta} \big(\rho Z_{i+1} + \sqrt{1-\rho^2} W_{i+1}\big), \end{aligned} \tag{5.5}$$

with $Z_1, ..., Z_n$ and $W_1, ..., W_n$ the independent standard normal components of vectors Z, W.

In some cases, for example in the Hull-White model (5.2), it is convenient to apply an Euler rule to the logarithm of the variable of interest and then exponentiate to get

$$\begin{aligned} \hat{S}_{i+1} &= \hat{S}_i e^{(r - \hat{V}_i/2)\Delta + \sqrt{\hat{V}_i}\sqrt{\Delta} Z_{i+1}} \\ \hat{V}_{i+1} &= \hat{V}_i e^{(\mu - \sigma^2/2)\Delta + \sigma\sqrt{\Delta}(\rho Z_{i+1} + \sqrt{1-\rho^2} W_{i+1})}. \end{aligned} \tag{5.6}$$

The Euler scheme in (5.5) does not preclude $\hat{V}$ from becoming negative, which in the Heston (5.3) model leads to a nonsensical solution because the evolution of S depends on the square root of V. Alternative Euler schemes for stochastic volatility models have been proposed in the literature and compared in Lord et al. [LKD]. The Euler scheme leading to smallest bias in [LKD] is the *full truncation scheme*

$$\begin{aligned}\hat{S}_{i+1} &= \hat{S}_i e^{(r-\hat{V}_i/2)\Delta+\sqrt{\hat{V}_i}\sqrt{\Delta}Z_{i+1}} \\ \tilde{V}_{i+1} &= \tilde{V}_i + \kappa(\theta - \max\{\tilde{V}_i, 0\})\Delta + \sqrt{\max\{\tilde{V}_i, 0\}}\sqrt{\Delta}\big(\rho Z_{i+1} + \sqrt{1-\rho^2}W_{i+1}\big) \\ \hat{V}_{i+1} &= \max\{\tilde{V}_{i+1}, 0\}. \end{aligned} \tag{5.7}$$

5.1.3 Realized Variance under discretization schemes

Our goal is to design an efficient numerical algorithm for the computation of the expectation of a function of discretely recorded variance. Recall from Definition (3.2) that this is defined as

$$V_n(0,T) := \frac{n}{T(n-1)} \sum_{i=0}^{n-1} \log\left(\frac{S_{i+1}}{S_i}\right)^2, \tag{5.8}$$

for recordings of the true continuous time process (5.1), which will be approximated in this paper by an Euler rule. For the core of our analysis, the factor $\frac{n}{T(n-1)}$ is completely irrelevant, so we will just drop it. Moreover, for notational simplicity, in the remainder of the chapter S_i will denote a discretization scheme applied to (5.1). Under the simplest Euler scheme (5.5), realized variance (5.8) can be expressed as a deterministic function of the stochastic shocks

$$\mathcal{R}(Z,W) := \sum_{i=0}^{n-1} \log\left(1 + r\Delta + \frac{\sigma(S_i)}{S_i}\sqrt{V_i}\sqrt{\Delta}Z_{i+1}\right)^2, \tag{5.9}$$

where V_i is some function of Z and W specified by the choice of model and that we do not need to write explicitly. According to Theorem 2.14, the price we must compute is

$$C = C(0) = P(0,T)\mathbb{E}^Q[g(\mathcal{R}(Z,W))], \tag{5.10}$$

which is a high dimensional integral in terms of the Gaussian components of Z, W.

5.2 Combining Integration with Random Sampling

5.2.1 The general setting

In this section we assume that exact integration over a low dimensional manifold of interest is feasible, and introduce an algorithm to compute (5.10) by combining this integration procedure with Monte Carlo simulation over the remaining dimensions. This idealized algorithm is used to characterize the variance reduction achieved by the method. In later sections, we will approximate the exact deterministic integration by one of two alternatives: deterministic but possibly biased quadrature, or random sampling under very fine stratification. Numerical experiments will provide evidence of the variance reduction achieved under these implementation alternatives.

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

$$\hat{C} = \frac{1}{M}\sum_{j=1}^{M} g(\mathcal{R}(Z^j, W^j)). \tag{5.11}$$

The fact that the estimator $\hat{C}$ is unbiased, in the sense that

$$\mathbb{E}\big(\hat{C}\big) = \mathbb{E}\big(g(\mathcal{R}(Z,W))\big),$$

with variance

$$\mathsf{Var}(\hat{C}) = \frac{\mathsf{Var}(g(\mathcal{R}(Z,W)))}{M}, \tag{5.12}$$

are standard properties of Monte Carlo explained in the previous chapter.

An alternative way of computing (5.10) is as a deterministic integral over the $2n$-dimensional space spanned by the Gaussian vectors Z and W. The joint distribution of independent Gaussian Z and W is known in closed form then we can write

$$C = \int_{z_1,\ldots,z_n,w_1,\ldots,w_n} g(\mathcal{R}(z,w)) \prod_{j=1}^{n} \Phi(z_j)\Phi(w_j) dz_j dw_j, \tag{5.13}$$

with $\Phi(u)$ the univariate standard normal density. The result of (5.13) is the exact price of interest, but (5.13) is an integral over a high dimensional space. Except for a very restrictive class of payoffs and models, (5.13) can not be solved analytically. And its

numerical solution as a deterministic integral suffers from the curse of dimensionality, becoming too expensive beyond dimension 3 or 4. The method we develop in this work lies essentially between (5.11) and (5.13) in the sense that integration is performed over a few dimensions that substantially contribute to the noise of the price estimator, and Monte Carlo samples are generated conditional on the values of variables on the low dimensional manifold.

Let π be a vector-valued random variable with components $\pi_1, ...\pi_p$ defined by the action of deterministic functions of $\mathbb{R}^{2n} \longrightarrow \mathbb{R}$ on the Gaussian vectors Z, W. We propose to combine integration over the values of the components of π with Monte Carlo sampling conditional on π. The corresponding estimator is

$$\hat{C} = \int \frac{1}{M(u)} \sum_{j=1}^{M(u)} g(\mathcal{R}(Z^j, W^j))\eta(u)du, \tag{5.14}$$

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 (5.14) is unbiased. By the independence of samples the variance of $\hat{C}$ is

$$\mathsf{Var}(\hat{C}) = \int \mathsf{Var}\left(\frac{1}{M(u)} \sum_{j=1}^{M(u)} g(\mathcal{R}(Z^j, W^j))\eta(u) \middle| \pi = u \right) du, \tag{5.15}$$

and then

$$\mathsf{Var}(\hat{C}) = \int \frac{\eta(u)^2}{M(u)} \mathsf{Var}(g(\mathcal{R}(Z^j, W^j))|\pi = u)du. \tag{5.16}$$

As shown in the discussion about stratification in Chapter 2, 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 (5.16) becomes

$$\mathsf{Var}(\hat{C}) = \frac{\mathbb{E}(\mathsf{Var}(g(\mathcal{R}(Z, W))|\mathcal{F}_\pi))}{M}, \tag{5.17}$$

for the variance of the estimator based on the combination of deterministic integration and conditional Monte Carlo. It is informative to compare (5.17) with the variance of standard Monte Carlo estimator (5.12). By the variance decomposition formula (4.9) it always holds that

$$\mathsf{Var}[g(\mathcal{R}(Z, W))] = \mathsf{Var}[\mathbb{E}(g(\mathcal{R}(Z, W))|\mathcal{F}_\pi)] + \mathbb{E}[\mathsf{Var}(g(\mathcal{R}(Z, W))|\mathcal{F}_\pi)]. \tag{5.18}$$

The left side of (5.18) is independent of $\mathcal{F}_\pi$ and equal to M times (5.12). The two terms on the right are nonnegative and the second term coincides with M times (5.17). 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. In practice, however, 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 (5.14) should attempt to identify $\pi_1, ..., \pi_p$ such that:

- 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 (see conditional Monte Carlo algorithm in Chapter 2)
- The choice of $\pi_1, ..., \pi_p$ leads to $\mathbb{E}(\mathsf{Var}(g(\mathcal{R}(Z,W))|\mathcal{F}_\pi))$ much smaller than $\mathsf{Var}(g(\mathcal{R}(Z,W)))$ to achieve significant variance reduction.

Before showing our choice of conditioning variables we introduce some notation. Let $|V|_1 := \sum_{i=0}^{n-1} V_i$, denote the 1-norm of the instantaneous variance process over a path, and $|V|_2 := \sum_{i=0}^{n-1} V_i^2$. Let $|Z|_2^2 := \sum_{i=0}^{n-1} Z_{i+1}^2$ be the 2-norm of the shocks to the asset price.

5.2.2 Monte Carlo conditional on $|Z|_2^2$ and $|V|_1$

For this section, we keep the following assumptions in effect:

Assumption 5.1 *For a model (5.1) and an associated Euler scheme we assume that:*

- *V is independent of Z.*
- *The function σ in (5.1) is linear.*

We are assuming that the variance process is not only autonomous but also independent of shocks to the process S. A linear σ is a feature of the original versions of the Heston [H] and Hull and White [HW]. These assumptions are needed only in this Section, and are not required for the implementation of the numerical algorithm.

In order to identify a good choice of conditioning variables we do a linear expansion of (5.9)

$$\mathcal{R}(Z,W) \approx \sum_{i=0}^{n-1} r^2\Delta^2 + 2\frac{\sigma(S_i)}{S_i}\sqrt{V_i}r\Delta^{3/2}Z_{i+1} + \left(\frac{\sigma(S_i)}{S_i}\right)^2 V_i\Delta Z_{i+1}^2, \tag{5.19}$$

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

$$\hat{\mathcal{R}}(Z,W) \equiv \sum_{i=0}^{n-1} \left(\frac{\sigma(S_i)}{S_i}\right)^2 V_i\Delta Z_{i+1}^2. \tag{5.20}$$

Close inspection of (5.20) reveals that its value is strongly influenced by $|Z|_2^2$.

As a conditioning variable is suitable if it captures the variability of $\hat{\mathcal{R}}$, it could be convenient to take a look at $\mathsf{Var}(\hat{\mathcal{R}})$. Since this parameter involves the first and second moments of this variable, we proceed to compute them as explicitly as possible and organize the results in the following

Lemma 5.2 *Under Assumption 5.1, we have*

$$\mathbb{E}(\hat{\mathcal{R}}) = \mathbb{E}(|V|_1), \tag{5.21}$$

$$\mathbb{E}(\hat{\mathcal{R}}^2) = 2\mathbb{E}(|V|_2^2) + \mathbb{E}(|V|_1^2), \tag{5.22}$$

$$\mathsf{Var}(\hat{\mathcal{R}}) = 2\mathbb{E}(|V|_2^2) + \mathsf{Var}(|V|_1), \tag{5.23}$$

with $Z \sim \mathcal{N}(0, I_n)$.

Proof The first moment is straightforward,

$$\mathbb{E}(\hat{\mathcal{R}}) = \mathbb{E}\left(\sum_{i=0}^{n-1} V_i Z_{i+1}^2\right) = \sum_{i=0}^{n-1} \mathbb{E}(V_i)\mathbb{E}(Z_{i+1}^2) = \mathbb{E}\left(\sum_{i=0}^{n-1} V_i\right) = \mathbb{E}(|V|_1),$$

and for the second moment we do

$$\begin{aligned}\mathbb{E}(\hat{\mathcal{R}}^2) &= \mathbb{E}\left(\sum_{i=0}^{n-1} Z_{i+1}^4 V_i^2 + \sum_{i=0}^{n-1} Z_{i+1}^2 V_i Z_{j+1}^2 V_j\right) \\ &= 3\sum_{i=0}^{n-1}\mathbb{E}(V_i^2) + \sum_{i\neq j}\mathbb{E}(V_iV_j) = 2\mathbb{E}(|V|_2^2) + \mathbb{E}(|V|_1^2).\end{aligned}$$

From (5.21), (5.22) and the definition of variance, we get

$$\mathsf{Var}(\hat{\mathcal{R}}) = \mathbb{E}(\hat{\mathcal{R}}^2) - \mathbb{E}(\hat{\mathcal{R}})^2 = 2\mathbb{E}(|V|_2^2) + \mathbb{E}(|V|_1^2) - \mathbb{E}(|V|_1)^2 = 2\mathbb{E}(|V|_2^2) + \mathsf{Var}(|V|_1),$$

which proves (5.23). □

Formula (5.23) gives us a key about what other variable to choose for conditioning: fixing $|V|_1$ would imply $\mathsf{Var}(|V|_1) = 0$, meaning a variance reduction arising from the removal of the second term. In fact, as was already mentioned, efficiency of conditioning relies on selecting variables that best capture the variability of $\hat{\mathcal{R}}$, and in our case, that variability (quantified by $\mathsf{Var}(\hat{\mathcal{R}})$) is highly explained by $\mathsf{Var}(|V|_1)$.

Consider now the variance reduction achieved by performing Monte Carlo conditional on $\pi = \{\pi_1, \pi_2\} = \{|Z|_2^2, |V|_1\}$. To the extent that significant variability of (5.20) is explained by $|Z|_2^2$ and $|V|_1$,

$$\mathbb{E}(\mathsf{Var}(g(\hat{\mathcal{R}})|\mathcal{F}_\pi)), \tag{5.24}$$

should be small relative to $\mathsf{Var}(g(\hat{\mathcal{R}}))$. For example, in the special case with linear g, linear $\sigma(S)$ in (5.1), and $V_i = V_0 \ \forall \ i < n$, it is clear that $\mathbb{E}(\mathsf{Var}(g(\hat{\mathcal{R}})|\mathcal{F}_\pi)) = 0$. By the closeness of (5.9) and (5.20), small (5.24) implies that $\mathbb{E}(\mathsf{Var}(g(\hat{\mathcal{R}})|\mathcal{F}_\pi))$ is also small, as needed for effective variance reduction for exact realized variance (5.9). We stress that the actual computational algorithm will use (5.9), therefore will involve no approximation. We will only use (5.20) to gain intuition about the algorithm and to derive a theoretical result that approximately quantifies the variance reduction achieved by the method.

Taking $\pi = \{|Z|_2^2, |V|_1\}$ as conditioning set is successful in eliminating variance because it contains two main sources of stochasticity in $\hat{\mathcal{R}}$. We make this statement precise by deriving, under the previous assumptions, a result that quantifies the variance reduction achieved by the method.

Proposition 5.3 *Assume $Z \sim \mathcal{N}(0, I_n)$ and V autonomous and independent of Z. Let $\mathcal{F}_\pi := \mathcal{F}_{_{|Z|,|V|}} := \Sigma(|Z|_2, |V|_1)$. Then, we have*

$$\mathbb{E}(\hat{\mathcal{R}}|\mathcal{F}_\pi) = \frac{|Z|_2^2|V|_1}{n} \tag{5.25}$$

$$\mathit{Var}(\hat{\mathcal{R}}|\mathcal{F}_\pi) = 2\mathbb{E}\bigg(|V|_2^2 - \frac{1}{n}|V|_1^2\bigg). \tag{5.26}$$

In particular, the variance reduction achieved after conditioning is

$$\mathit{Var}(\mathbb{E}(\hat{\mathcal{R}}|\mathcal{F}_\pi) = \mathit{Var}(|V|_1) + \frac{2}{n}\mathbb{E}(|V|_1^2).$$

PROOF*:* For (5.25), let $\mathcal{F}_{_{|Z|}} := \Sigma(|Z|_2)$ and $\mathcal{F}_{_{|Z|,V}} := \Sigma(|Z|_2, V)$.

Note that as $Z_1, ..., Z_n$ are i.i.d., $\mathbb{E}\big(Z_i^2\big|\mathcal{F}_{|Z|}\big) = \mathbb{E}\big(Z_i^2\big||Z|_2^2\big) = \frac{|Z|_2^2}{n}$ for all $1 \leq i \leq n$. Using independence of Z and V together with the law of iterated expectations we get

$$\begin{aligned}\mathbb{E}(\hat{\mathcal{R}}|\mathcal{F}_\pi) &= \mathbb{E}\bigg(\mathbb{E}\bigg(\sum_{i=0}^{n-1} V_i Z_{i+1}^2\bigg|\mathcal{F}_{|Z|,V}\bigg)\bigg|\ \mathcal{F}_\pi\bigg) = \mathbb{E}\bigg(\sum_{i=0}^{n-1} V_i \mathbb{E}\big(Z_{i+1}^2\big|\mathcal{F}_{|Z|,V}\big)\bigg|\mathcal{F}_\pi\bigg) \\ &= \mathbb{E}\bigg(\sum_{i=0}^{n-1} V_i \frac{|Z|_2^2}{n}\bigg|\mathcal{F}_\pi\bigg) = \frac{|Z|_2^2}{n}\mathbb{E}\bigg(\sum_{i=0}^{n-1} V_i\bigg|\mathcal{F}_\pi\bigg) = \frac{|Z|_2^2}{n}|V|_1.\end{aligned}$$

as $\mathcal{F}_{|Z|} \subseteq \mathcal{F}_\pi \subseteq \mathcal{F}_{|Z|,V}$ and V_i is $\mathcal{F}_{|Z|,V}$-measurable.
For the last equality, we invoke (4.9) to write

$$\mathbb{E}(\mathsf{Var}(\hat{\mathcal{R}})|\mathcal{F}_\pi)) = -\mathsf{Var}(\mathbb{E}(\hat{\mathcal{R}}|\mathcal{F}_\pi))) + \mathsf{Var}(\hat{\mathcal{R}}). \tag{5.27}$$

By (5.25) it holds that

$$\begin{aligned}\mathsf{Var}\big(\mathbb{E}(\hat{\mathcal{R}}|\mathcal{F}_\pi)\big) &= \mathsf{Var}\bigg(|Z|_2^2\frac{|V|_1}{n}\bigg) \\ &= \frac{1}{n^2}(\mathbb{E}(|V|_1^2)\mathbb{E}(|Z|_2^4) - \mathbb{E}(|V|_1)^2\mathbb{E}(|Z|_2^2)^2) \\ &= \frac{1}{n^2}((n(n-1)+3n)\mathbb{E}(|V|_1^2) - n^2\mathbb{E}(|V|_1)^2 \\ &= \mathsf{Var}(|V|_1) + \frac{2}{n}\mathbb{E}(|V|_1^2),\end{aligned} \tag{5.28}$$

therefore, subtracting this from (5.23) we get that (5.27) becomes

$$\mathbb{E}(\mathsf{Var}((\hat{\mathcal{R}})|\mathcal{F}_\pi)) = -\frac{2}{n}\mathbb{E}(|V|_1^2) + 2\mathbb{E}(|V|_2^2). \tag{5.29}$$

□

Proposition 5.3 will be used next to derive a result on the magnitude of variance reduction obtained by conditioning on $|Z|_2^2$ and $|V|_1$. We stress that the effect of conditioning in these variables results in a variance reduction term greater than $\mathsf{Var}(|V|_1)$.

So far, we have dealt only with "forward" payoffs, i.e., payoffs of the form $g(x) = kx$. In order to extend our analysis to more general payoffs, we will make the following assumption:

Assumption 5.4 *For a constant A and a function $\epsilon : \mathbb{R} \to \mathbb{R}$, the payoff function $g(x)$ satisfies any of the following two conditions:*

$g(x) = A + x - \epsilon(x)$, *for nondecreasing $g(x)$ and $\epsilon(x)$,*

$g(x) = A - x - \epsilon(x)$, *for nonincreasing $g(x)$ and $\epsilon(x)$.*

REMARK The class of payoff functions $g(x)$ that satisfy Assumption 5.4 includes the payoff of a call option with strike K, which can be written as

$$\max\{x-K,0\} = x - \epsilon(x), \quad \text{for } \epsilon(x) = \min\{x, K\},$$

and the payoff of a put option, which can be written as

$$\max\{K-x,0\} = K - x - \epsilon(x), \quad \text{for } \epsilon(x) = -\max\{x-K, 0\}.$$

Motivated by (5.18), define *residual variance* as

$$\mathcal{V}_{res} = \frac{\mathbb{E}(\mathsf{Var}(g(\hat{\mathcal{R}}(Z,W))|\mathcal{F}_\pi))}{\mathsf{Var}(g(\hat{\mathcal{R}}(Z,W)))} \tag{5.30}$$

This is the proportion of the variance of standard Monte Carlo that survives after performing deterministic integration on the low dimensional manifold. Our next goal is to get a bound on $\mathcal{V}_{res}$ for this class of payoffs. To achieve it, we need the following lemma:

Lemma 5.5 *The inequality*

$$\mathit{Cov}\,(f(X), g(X)) \geq 0 \tag{5.31}$$

holds for any two increasing functions $f, g : \mathcal{J} \longrightarrow \mathbb{R}$ for which $f(X)$ and $g(X)$ have a finite second moment.

PROOF: See [S]. □

Theorem 5.6 *Under Assumption 5.1 and Assumption 5.4, the residual variance after integrating on $\pi = \{|Z|_2^2, |V|_1\}$ satisfies:*

$$\mathcal{V}_{res} \leq \frac{2\mathbb{E}(|V|_2^2 - \frac{1}{N}|V|_1^2)}{\mathit{Var}(g(\hat{\mathcal{R}}(Z,W)))} = \frac{\mathit{Var}(\hat{\mathcal{R}}(Z,W)) - \mathit{Var}(|Z|_2^2 \frac{|V|_1}{n})}{\mathit{Var}(g(\hat{\mathcal{R}}(Z,W)))}. \tag{5.32}$$

PROOF: Under Assumption 5.4 it holds that

$$\begin{aligned} \mathsf{Var}(\hat{\mathcal{R}}(Z,W)|\mathcal{F}_\pi) &= \mathsf{Var}(g(\hat{\mathcal{R}}(Z,W))|\mathcal{F}_\pi) + \mathsf{Var}(\epsilon(\hat{\mathcal{R}}(Z,W))|\mathcal{F}_\pi) \\ &+ 2\mathsf{Cov}\big(g(\hat{\mathcal{R}}(Z,W)), \epsilon(\hat{\mathcal{R}}(Z,W))|\mathcal{F}_\pi\big), \end{aligned} \tag{5.33}$$

and the covariance term is nonnegative, as shown by Lemma (5.5), since g and ϵ are both nondecreasing or nonincreasing. Therefore

$$\mathsf{Var}(g(\hat{\mathcal{R}}(Z,W))|\mathcal{F}_\pi) \leq \mathsf{Var}(\hat{\mathcal{R}}(Z,W)|\mathcal{F}_\pi), \tag{5.34}$$

implying that

$$\mathbb{E}(\mathsf{Var}(g(\hat{\mathcal{R}}(Z,W))|\mathcal{F}_\pi)) \leq \mathbb{E}(\mathsf{Var}(\hat{\mathcal{R}}(Z,W)|\mathcal{F}_\pi)). \tag{5.35}$$

We bound residual variance (5.30) using (5.26) and (5.35) to get

$$\mathcal{V}_{res} \leq \frac{\mathbb{E}(\mathsf{Var}(\hat{\mathcal{R}}(Z,W))|\mathcal{F}_\pi))}{\mathsf{Var}(g(\hat{\mathcal{R}}(Z,W)))} = \frac{2\mathbb{E}(|V|_2^2 - \frac{1}{n}|V|_1^2)}{\mathsf{Var}(g(\hat{\mathcal{R}}(Z,W)))}, \tag{5.36}$$

which proves the first inequality in (5.32). For the last equality in (5.32) we use (5.18) and (5.36) to write

$$\begin{aligned} \mathcal{V}_{res} &\leq \frac{\mathsf{Var}(\hat{\mathcal{R}}(Z,W))) - \mathsf{Var}(\mathbb{E}(\hat{\mathcal{R}}(Z,W))|\mathcal{F}_\pi)}{\mathsf{Var}(g(\hat{\mathcal{R}}(Z,W)))} \\ &= \frac{\mathsf{Var}(\hat{\mathcal{R}}(Z,W)) - \mathsf{Var}\big(|Z|_2^2 \frac{|V|_1}{n}\big)}{\mathsf{Var}(g(\hat{\mathcal{R}}(Z,W)))}. \end{aligned} \tag{5.37}$$

□

REMARK The first inequality in (5.32) 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 $L = \{(x_1, ..., x_n) \in \mathbb{R}^n : x_1 = x_2 = ... = x_n\}$. 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 5.7 *Let Assumption 5.1 and Assumption 5.4 hold. Let μ_i and σ_i^2 be the mean and variance of V_i for $i = 1, ..., n$, and set $\mu_{\max} \equiv \max\{\mu_0, ..., \mu_{n-1}\}$,*
$\mu_{\min} \equiv \min\{\mu_0, ..., \mu_{n-1}\}, \sigma_{\max} \equiv \max\{\sigma_0, ..., \sigma_{n-1}\}$.
Then,

$$\mathcal{V}_{res} \leq n\frac{\mu_{\max}^2 - \mu_{\min}^2 + \sigma_{\max}^2}{\mathsf{Var}(g(\mathcal{R}(Z,W)))}. \tag{5.38}$$

PROOF: From the inequality in (5.32) residual variance is

$$\begin{aligned}
\mathcal{V}_{res} &\leq \frac{2\mathbb{E}(|V|_2^2 - \frac{1}{n}|V|_1^2)}{\mathsf{Var}(g(\mathcal{R}(Z,W)))} \\
&= \frac{2(\mathbb{E}(|V|_2^2) - \frac{1}{n}(\mathsf{Var}(|V|_1^2) + \mathbb{E}(|V|_1)^2))}{\mathsf{Var}(g(\mathcal{R}(Z,W)))} \\
&\leq \frac{2(\sum_{i=0}^{n-1}\mu_i^2 + \sigma_i^2 - \frac{1}{n}(\sum_{i=0}^{n-1}\mu_i)^2)}{\mathsf{Var}(g(\mathcal{R}(Z,W)))} \\
&\leq \frac{2N(\mu_{max}^2 + \sigma_{max}^2 - \mu_{min}^2)}{\mathsf{Var}(g(\mathcal{R}(Z,W)))}. \qquad (5.39)
\end{aligned}$$

□

Residual variance in (5.38) 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 (5.38) by $\mu_{max}^2 - \mu_{min}^2$, and to uncertainty in V, represented by σ_{max}^2.

We have shown in this Section that, under Assumption 5.1 and Assumption 5.4, the combination of integration on the squared norm of Z and the path average of V followed by conditional Monte Carlo leads to quantifiable and possibly significant variance reduction. However, the implementation of this algorithm requires in practice the knowledge of the distribution of the path average of V, which in general is not known. In order to address this issue we consider an alternative set of conditioning variables.

Conditioning on $|Z|_2^2 \cdot |V|_1$

In order to interpret Theorem 5.6, consider the random variable $U := |Z|_2^2 \frac{|V|_1}{n}$. Theorem 5.6 states in the last equality in (5.32) that the efficiency of the proposed method in reducing variance depends on how close the second moment of U is to the second moment of $\hat{\mathcal{R}}$.

Note that by (5.25), it obviously holds that $\mathbb{E}(\hat{\mathcal{R}}|\mathcal{F}_U) = U$. This says that U is also a suitable choice for conditioning and will be in fact a key variable in our algorithm. The density function of this new random variable U can be easily derived in terms of the density functions of X and Y.

In general, if X, Y are two independent random variables with densities ϕ_X, ϕ_Y respectively, with $P(X > 0) = 1$ (as it is the case, since $|Z|_2^2 \sim \chi_n^2$), then the cumulative probability function Φ_U can be written as

$$\Phi_U(u) = \int_0^\infty \phi_{U|X}(u|x)\phi_X(x)dx. \qquad (5.40)$$

To obtain the expression of $\phi_{U|X}(u|x)$, we calculate the CDF $\Phi_{U|X}(u|x)$ and then take derivative. For this, we write (assuming $x > 0$)

$$\begin{aligned}\Phi_{U|X}(u|x) &= P(U \leq u|X = x) = P(XY \leq u|X = x) \\ &= P(xY \leq u) = P\left(Y \leq \frac{u}{x}\right) = \Phi_Y\left(\frac{u}{x}\right),\end{aligned}$$

giving

$$\phi_{U|X}(u|x) = \frac{\partial}{\partial u}\Phi_Y\left(\frac{u}{x}\right) = \frac{1}{x}\phi_Y\left(\frac{u}{x}\right). \tag{5.41}$$

Substituting in (5.40), we get

$$\phi_U(u) = \int_0^\infty \frac{\phi_Y\left(\frac{u}{x}\right)\phi_X(x)}{x}dx =: (\phi_X * \phi_Y)(u). \tag{5.42}$$

Formula (5.42) can be interpreted as a "multiplicative" convolution between ϕ_X, ϕ_Y. The algorithm that employs U as conditioning variable requires the density of $U = XY$, where $X \sim \chi^2_\nu$. For reasons that will be clear in the following section and for the purpose of testing such algorithm and its efficiency, we will take $Y \sim \mathcal{N}(0, 1)$. For such a pair, the density functions are given by $\phi_Y(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$ and $\phi_X(t) = \frac{1}{\Gamma(\nu)}e^{-t/2}t^{\nu-1}I_{(0,\infty)}(t)$, leading to a fully explicit expression of ϕ_U in terms of ϕ_X, ϕ_Y, namely

$$\phi_U(u) = \frac{1}{\sqrt{2\pi}\Gamma(\nu)}\int_0^\infty e^{-\frac{u^2}{2x^2}-\frac{x}{2}}x^{\nu-1}dx. \tag{5.43}$$

The evaluation of this integral is a non-trivial task. A characterization of it in terms of hypergeometric functions will be given later by means of *Mellin Transform* techniques.

5.2.3 Monte Carlo conditional on $|Z|_2^2$ and $v^\top W$

We propose to combine deterministic integration and Monte Carlo in an estimator of the form (5.14), now taking $\hat{\pi} := \{|Z|_2^2, v^\top W\}$ as conditioning variables. Assumption 5.1 and Assumption 5.4 still hold. The new conditioning variable, $v^\top W$, is a linear combination of shocks W with weights defined by $v \in \mathbb{R}^n$, 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 integration. Moreover, the components of W are normally

distributed even after conditioning on a linear combination of them, therefore conditional sampling is straightforward.

The argument that led to (5.34) in the proof of Theorem 5.6 continues to hold for a different conditioning set, therefore we have that

$$\mathbb{E}(\mathsf{Var}(g(\hat{\mathcal{R}}(Z,W))|\mathcal{F}_{\hat{\pi}})) \leq \mathbb{E}(\mathsf{Var}(\hat{\mathcal{R}}(Z,W)|\mathcal{F}_{\hat{\pi}})). \tag{5.44}$$

The aim is to find v with $|v|_2 = 1$, such that by using $v^\top W$ as conditioning variable the residual variance is small. Taking advantage of (5.44) we choose to minimize $\mathbb{E}(\mathsf{Var}(\hat{\mathcal{R}}(Z,W)|\mathcal{F}_{\hat{\pi}}))$ which, by (5.18), is equivalent to maximize $\mathsf{Var}(\mathbb{E}(\hat{\mathcal{R}}|\mathcal{F}_{\hat{\pi}}))$. We are looking for

$$\arg\max_{\{|v|_2=1\}} \mathsf{Var}\Big(\mathbb{E}\Big(\sum_{i=0}^{N-1} V_i(W)Z_{i+1}^2\Big|\ |Z|_2^2\ ,\ v^\top W\Big)\Big). \tag{5.45}$$

Solving (5.45) is a daunting task because, in principle, each V_i arising from an Euler rule on (5.1) is a nonlinear function of W. However, for relatively small n, which is indeed the setting of importance for this paper, the behavior of the coefficients of (5.1) as functions of V suggests linearizing V_i as a function of the underlying shocks W to get

$$\arg\max_{\{|v|=1\}} \mathsf{Var}\Big(\mathbb{E}\Big(\sum_{i=0}^{n-1}(V_0 + b_w^i W)Z_{i+1}^2\Big|\ |Z|_2^2,\ v^\top W\Big)\Big),$$

where $b_w^i \equiv \nabla_w V_i\big|_{w=0}$. By independence of Z and W we obtain

$$\arg\max_{\{|v|=1\}} \mathsf{Var}\Big(Y(Z)\mathbb{E}\Big(\sum_{i=0}^{n-1} b_w^{i'} W\Big|v^\top W\Big)\Big),$$

for Y a generic function of Z. Recalling that $\mathbb{E}\big(\mathbb{E}\big(\sum_{i=0}^{n-1} b_w^{i'} W\big|v^\top W\big)\big) = 0$, we get

$$\arg\max_{\{|v|=1\}} \mathbb{E}\Big(\mathbb{E}^2\Big(\sum_{i=0}^{n-1} b_w^{i'} W|v^\top W\Big)\Big). \tag{5.46}$$

Let $b \equiv \nabla_w\big(\sum_{i=0}^{n-1} V_i\big)\big|_{w=0}$. Then (5.46) is equivalent to

$$\arg\max_{\{|v|=1\}} \mathsf{Var}\big(\mathbb{E}\big(b^\top W|v^\top W\big)\big), \tag{5.47}$$

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

The derivation above indicates that, in order to reduce variance by conditioning on $v^\top W$, the optimal vector of weights is proportional to the gradient of the path average of V. It is in this sense that the optimal linear combination is close to the exact path average $|V|_1$.

The optimality rule in (5.47) is very similar to the *optimal stratification direction* in Glasserman et al. [GHS]. This analogy led us to consider an alternative optimal v defined as the solution of

$$\arg \max_{\{|v|=1\}} \mathsf{Var}\big(\mathbb{E}(b^\top W + \frac{1}{2}(W^\top H_f^i W)|v^\top W)\big), \tag{5.48}$$

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 (5.48) is characterized by Proposition (4.1) in the previous chapter. 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 weights given by the gradient of $|V|_1$. We follow this prescription in the rest of the thesis.

Example 5.8

Computation of optimal integration direction for the Hull-White model

We solve (4.28) through the Lagrange's multiplier method. For symmetric A, the derivative with respect to v_i of the function to maximize is

$$b_i(b^\top v) + 2(v^\top A v)((A_D v)_i + (A_S v)_i),$$

where $A_D + A_S + A_S^\top = A$, A_D being the diagonal matrix of the diagonal elements of A. Hence, the multiplier equations read

$$b_i(b^\top v) + (2(A_D - \lambda I)v)_i + (v^\top A v)(A_S v)_i = 0 \qquad (i = 1, 2, ..., n). \tag{5.49}$$

Consider the Hull-White model (5.2) 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

$$\begin{aligned} |V|_1 &= V_0(1 + e^{(\mu - \frac{1}{2}\sigma^2)\Delta + \sigma\sqrt{\Delta}W_1} + e^{2(\mu - \frac{1}{2}\sigma^2)\Delta + (\sigma\sqrt{\Delta t})(W_1 + W_2)} + ... \\ &+ e^{(N-1)(\mu - \frac{1}{2}\sigma^2)\Delta + \sigma\sqrt{\Delta}(W_1 + ... + W_{n-1})}), \end{aligned}$$

which leads to

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

$$\begin{aligned}\nabla_i|V|_1\big|_{w=0} &= (\sigma\sqrt{\Delta})\left(\frac{e^{(\mu-\frac{1}{2}\sigma^2)\Delta n}-1}{e^{(\mu-\frac{1}{2}\sigma^2)\Delta}-1} - \frac{e^{i(\mu-\frac{1}{2}\sigma^2)\Delta}-1}{e^{(\mu-\frac{1}{2}\sigma^2)\Delta}-1}\right)\\ &= (\sigma\sqrt{\Delta})\left(\frac{e^{(\mu-\frac{1}{2}\sigma^2)\Delta n}-e^{(\mu-\frac{1}{2}\sigma^2)\Delta i}}{e^{(\mu-\frac{1}{2}\sigma^2)\Delta}-1}\right).\end{aligned}$$

A similar computation shows that,

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

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 (5.49).

5.3 Implementation of the Algorithm

We describe in this Section the implementation of the algorithm discussed in Section 5.2. The theoretical results quantifying variance reduction derived in Section 5.2 were obtained under Assumption 5.1 and Assumption 5.4, which are not needed for the practical implementation of the algorithm. In particular, we lift in this Section the requirement of independent Z and V therefore allowing for nonzero ρ in (5.1). By analogy to the solution of (5.47) we implement the method on a low dimensional manifold spanned by $|Z|_2^2$ and $v^\top W$ where $v = \nabla_w|V(Z,W)|_1$ evaluated at $Z=0, W=0$. We do not claim that this conditioning variable remains optimal in the correlated case, but we stress that it is conveniently tractable and that introduces no bias in the algorithm.

The initial step of the algorithm is the computation of v, as shown explicitly in Example 5.8. Then, the vector v is used in two alternative numerical implementations of the integration over the low dimensional manifold.

In the last part of this section, we give a characterization of density of U.

5.3.1 The Algorithm through quadrature

- Integrate over $|Z|_2^2$ and $v^\top W$ by quadrature. For the standard normal distribution of $v^\top W$ we take 40 equally spaced steps on the real line, ranging from -4 to 4. For the χ_n^2 distribution of $|Z|_2^2$ we take 30 equally spaced steps on the positive real half line, ranging from 0 to $15\sqrt{n}$. Compute the probability of each cell as the product of the corresponding densities at a corner of the cell times the size of the cell.

- Generate N samples of the vector Z, conditional on its norm, writing $Z = \sqrt{R}\frac{\xi}{|\xi|_2}$, where $\xi \sim \mathcal{N}(0, I)$. The number of paths, N, is proportional to the probability of the cell.

- Generate N samples of the vector W, conditional on the value of $v^\top W$, using the fact that if $\xi \sim \mathcal{N}(0, \Sigma)$ in $\mathbb{R}^d$ with $Y = v^\top \xi$ stratified for some $v \in \mathbb{R}^d$, then $(\xi | Y = y) \sim \mathcal{N}(\Sigma v y, \Sigma - \Sigma v v^\top \Sigma)$. Consequently, sample W as $W = vv^\top W + (I - vv^\top)\Lambda$, where $\Lambda \sim \mathcal{N}(0, I_n)$ and $|v| = 1$.

- Compute the realized variance and derivative payoff for each sample of $\{Z, W\}$, take simple average over samples and probability weighted average over the two dimensional grid for the manifold spanned by $|Z|_2^2$ and $v^\top W$.

Remark An alternative algorithm using the conditioning variable U can be developed. Instead of integrating over $|Z|_2$ and $v^\top W$, we could only perform quadrature over U, the product of both variables. We thereafter can sample R conditional on U. The distribution of R conditional on U is characterized analytically as

$$\Phi_{R|U}(x|u) = \frac{1}{\phi_U(u)} \int_{-\infty}^{x} \frac{\phi_Y\left(\frac{u}{t}\right)\phi_R(t)}{t} dt,$$

where ϕ_U is given by (5.43)(see Theorem 5.9 below) and ϕ_Y is the normal density function.

5.3.2 The Algorithm through stratification

- We stratify $|Z|_2^2$ and $v^\top W$ by inversion, as explained in the discussion of stratification techniques in Chapter 4. We partition the square $[0, 1] \times [0, 1]$ through a cartesian grid with 90×90 equally sized cells. Within each cell we sample a uniformly distributed two dimensional vector $\{u_1, u_2\}$. Samples of $|Z|_2^2$ and $v^\top W$ are obtained by the application of the inverse χ_n^2 and standard normal distributions on $\{u_1, u_2\}$.

- In each cell, sample a vector Z, conditional on its norm, writing $Z = |Z|_2 \frac{\xi}{|\xi|_2}$, where $\xi \sim N(0, I)$.
- Sample W in the same manner as in the third step of the algorithm by quadrature.
- Compute realized variance and the derivative payoff for each sample of $\{Z, W\}$, and take simple average over the stratification grid for the manifold spanned by $|Z|_2^2$ and $v^\top W$.

5.3.3 Characterization of the density of the product of a normal and a chi square random variables

The following theorem characterizes integral (5.43) as a series in u. For the sake of space and organization, we leave its detailed proof in the Appendix, together with a brief introduction to the Mellin transform.

Theorem 5.9 *Let X, Y be to independent random variables such that $X \sim \chi^2_\nu, Y \sim \mathcal{N}(0,1)$, with $\nu \in \mathbb{N}$ even. Then, the density function of $U := XY$ is given by*

$$f_U(u) = S_1(u) + S_2(u) + S_3(u) + S_4(u) + S_5(u), \tag{5.50}$$

where

$$S_1(u) = 2^{\nu-1} \sum_{k=0}^{\frac{\nu}{2}-1} \Gamma\big(\nu - 2k - 1\big) \frac{(-1)^k u^{2k}}{2^{3k} k!}, \tag{5.51}$$

$$S_2(u) = 2^{-\frac{\nu+3}{2}} u^{\nu-1} \Gamma\left(\frac{1-\nu}{2}\right) {}_0F_2\left(\frac{1}{2}, \frac{1+\nu}{2}; \frac{u^2}{2^5}\right), \tag{5.52}$$

$$S_3(u) = (-1)^{\frac{\nu}{2}-1} \left(\frac{u^2}{2^5}\right)^{\frac{\nu}{2}} \ln\left(\frac{u^2}{2^5}\right) \frac{\Gamma\big(-\frac{1}{2}\big)}{\Gamma\big(\frac{\nu}{2}+1\big)} \, {}_0F_3\left(1, 1+\frac{\nu}{2}, \frac{3}{2}; -\frac{u^2}{2^5}\right), \tag{5.53}$$

$$S_4(u) = (-1)^{\frac{\nu}{2}} \sum_{k=0}^{\infty} \frac{2\Gamma(-k-\frac{1}{2})\psi(k+\frac{\nu}{2}+1)}{(k+\frac{\nu}{2})!k!} \left(\frac{u^2}{2^5}\right)^k, \tag{5.54}$$

$$S_5(u) = (-1)^{\frac{\nu}{2}} \sum_{k=0}^{\infty} \frac{\gamma_k^\nu}{(k+\frac{\nu}{2})!k!} \left(\frac{u^2}{2^5}\right)^k, \tag{5.55}$$

with $\gamma_k^\nu = \psi(-k-\frac{1}{2})\Gamma\big(-k-\frac{1}{2}\big) - \sum_{m=0}^{\frac{\nu}{2}-1} \frac{1}{k+\frac{\nu}{2}-m}$ *and* $\psi(z) := \frac{\Gamma'(z)}{\Gamma(z)}$.

Proof See appendix. □

5.4 Numerical results

In this Section we run numerical tests for the algorithms proposed in Section 5.3. Algorithms are implemented in MATLAB, and 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 have discussed two implementations: a possibly biased scheme based on integration by quadrature and a stratification scheme that is unbiased by construction. We also implement a standard Monte Carlo for comparison purposes. We run experiments for the discretized Hull-White and Heston models and investigate the presence of bias, and efficiency in reducing variance. Bias is defined as the absolute difference between the standard and quadrature Monte Carlo estimates for some function of interest. In the few cases in which the standard errors are larger than this difference, we conservatively bound the bias by the sum of standard errors although this is not an almost-sure bound. The reduction in variance is quantified as the square of the ratio of standard errors from competing methods.

5.4.1 Hull-White model

We first report results for an Euler discretization of the logarithm of S and V in the Hull-White model (5.6).

Our first set of results are very long computations designed to generate estimates with standard errors low enough to identify the bias present in quadrature Monte Carlo due to the discretization of the low dimensional integral. We use 10 million paths for standard Monte Carlo and a total of 1 millon paths for quadrature Monte Carlo proportionally allocated over the integration manifold. The computations for options with $n = 10$ days to expiry take about 820 seconds under standard Monte Carlo, and about 120 seconds for quadrature Monte Carlo. These computational costs become 2360 and 370 seconds respectively, in computations for $n = 50$ days. The out of the money calls are struck one standard deviation above expected future variance. All runs begin from $V_0 = 0.04$, which corresponds to a 20% initial volatility.

We report in Table 5.1 the mean and standard error of standard Monte Carlo and quadrature Monte Carlo estimators in a variety of experiments. For $n = 10$ we show moments for almost zero volatility of variance ($\sigma = 0.001$), and economically very large volatility of variance ($\sigma = 1$) which corresponds to a 100% annual volatility of variance. This is a very highly 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

Parameters	Function	Standard MC	Quadrature MC	Bias	Rel. Bias
n=10, $\sigma = 0.001$	$\mathbb{E}[\mathcal{R}]$	15.869 (0.002)	15.8734 (0.0002)	0.004	0.0003
n=10, $\sigma = 0.001$	$\mathsf{Var}[\mathcal{R}]$	50.46 (0.03)	50.469 (0.005)	< 0.03	< 0.0006
n=10, $\sigma = 0.001$	ATM Call	2.788 (0.002)	2.7961 (0.0002)	0.008	0.003
n=10, $\sigma = 0.001$	OTM Call	0.810 (0.001)	0.8046 (0.0001)	0.005	0.007
n=10, $\sigma = 1.0$	$\mathbb{E}[\mathcal{R}]$	15.876 (0.002)	15.8733 (0.0006)	0.003	0.0002
n=10, $\sigma = 1.0$	$\mathsf{Var}[\mathcal{R}]$	54.27 (0.03)	54.25 (0.01)	< 0.04	< 0.0007
n=10, $\sigma = 1.0$	ATM Call	2.876 (0.002)	2.8739 (0.0006)	< 0.003	< 0.001
n=10, $\sigma = 1.0$	OTM Call	0.854 (0.001)	0.8530 (0.0003)	0.001	0.001
n=50, $\sigma = 0.001$	$\mathbb{E}[\mathcal{R}]$	79.379 (0.006)	79.3662 (0.0004)	0.013	0.0002
n=50, $\sigma = 0.001$	$\mathsf{Var}[\mathcal{R}]$	253.0 (0.1)	252.24 (0.01)	0.8	0.003
n=50, $\sigma = 0.001$	ATM Call	6.301 (0.003)	6.2946 (0.0003)	0.007	0.001
n=50, $\sigma = 0.001$	OTM Call	1.553 (0.001)	1.5401 (0.0003)	0.013	0.009
n=50, $\sigma = 1.0$	$\mathbb{E}[\mathcal{R}]$	79.352 (0.008)	79.364 (0.003)	0.012	0.0002
n=50, $\sigma = 1.0$	$\mathsf{Var}[\mathcal{R}]$	704.5 (0.5)	702.9 (0.4)	1.6	0.002
n=50, $\sigma = 1.0$	ATM Call	10.211 (0.006)	10.208 (0.003)	< 0.009	< 0.001
n=50, $\sigma = 1.0$	OTM Call	3.049 (0.003)	3.050 (0.003)	< 0.006	< 0.002

Table 5.1: Long computations to quantify bias in quadrature Monte Carlo for the Hull-White model

value. Notably, from $\sigma = 0.001$ to $\sigma = 1$, the variance only increases from 50.46 to 54.27. 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, because in only ten days the variance process does not depart much from its initial value V_0. Therefore $\mathcal{R}$ is close to a sum of equally weighted random Gaussian shocks, and the variance of $\mathcal{R}$ for $\sigma = 1.0$ is largely that already present in the case of $\sigma = 0.001$. This suggests that integration by quadrature or fine stratification on $|Z|_2^2$ will be helpful in reducing variance for small n.

For $n = 50$, Table 5.1 shows that the variance of $\mathcal{R}$ is 253 for $\sigma = 0.001$ and 704.5 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 precise integration on the path average of V will be helpful in reducing variance for larger n.

The last column in Table 5.1 displays the relative bias associated with quadrature Monte Carlo for the Hull-White model. The bid-ask spread in the VIX index in the market in recent years was, according to Carr and Wu [CW], about one percentage volatility point. A conservatively large value for the VIX index is 50%, only surpassed on a few days during the 2008 crisis. This implies that relative bid-ask spreads are rarely smaller than 0.02. Moreover, the algorithms in this paper are to be used for more exotic contracts on realized variance, therefore with relative bid-ask spreads wider than 0.02. The relative biases we report in Table 5.1 are smaller than 0.02, suggesting that the algorithm is sufficiently accurate for practical use.

The next set of results, in Table 5.2, show the variance reduction achieved by quadrature Monte Carlo over standard Monte Carlo in terms of the square of the ratio of standard errors. A theoretical result (5.18) 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 quadrature Monte Carlo. A fair comparison is achieved by running both methods for the same computational budget.

We use a total of 7500 paths with proportional allocation for quadrature Monte Carlo and 12000 paths for standard Monte Carlo. With these parameters the computational costs of quadrature Monte Carlo match those of standard Monte Carlo, with 1 second for contracts with $n = 10$ days to expiration and 3 seconds for $n = 50$. These are plausible computational times for a financial industry setting.

We price ATM calls on variance, with strike at $\mathbb{E}[\mathcal{R}]$ and OTM calls and puts on variance. We run a set of experiments with $\sigma = 0.1$, which corresponds to a 10% volatility of

Parameters	Payoff	Standard MC	Quadrature MC	Var. Red.
n=10, $\sigma = 0.1$	OTM Put	0.726 (0.01)	0.7459 (0.0005)	400
n=10, $\sigma = 0.1$	ATM Call	2.858 (0.05)	2.796 (0.002)	625
n=10, $\sigma = 0.1$	OTM Call	0.818 (0.03)	0.802 (0.002)	225
n=10, $\sigma = 1.0$	OTM Put	0.763 (0.01)	0.751 (0.002)	25
n=10, $\sigma = 1.0$	ATM Call	2.855 (0.05)	2.874 (0.007)	51
n=10, $\sigma = 1.0$	OTM Call	0.861 (0.03)	0.854 (0.006)	25
n=50, $\sigma = 0.1$	OTM Put	1.176 (0.03)	1.1621 (0.002)	225
n=50, $\sigma = 0.1$	ATM Call	6.411 (0.09)	6.374 (0.005)	324
n=50, $\sigma = 0.1$	OTM Call	1.384 (0.04)	1.344 (0.004)	100
n=50, $\sigma = 1.0$	OTM Put	2.594 (0.05)	2.611 (0.01)	25
n=50, $\sigma = 1.0$	ATM Call	10.259 (0.2)	10.212 (0.03)	44
n=50, $\sigma = 1.0$	OTM Call	3.049 (0.1)	3.048 (0.02)	25

Table 5.2: Variance reduction by quadrature for the Hull-White model.

variance, or a mildly stochastic volatility regime, and a second set of runs with $\sigma = 1.0$ which corresponds to very high volatility of volatility.

Results in Table 5.2 show that the variance reduction achieved by the method is particularly high for short term contracts with mild stochasticity in the variance process, as it can be expected from integrating on $|Z|_2^2$. Higher volatility of volatility, or a longer duration, tend to decrease the reduction in variance, but in all cases this remains significant.

Results in Table 5.3 show the variance reduction achieved for the same set of experiments by the stratified Monte Carlo implementation for the Hull-White model. We use 90×90 strata with 1 path per stratum. The pattern of efficiency gains is similar to that present in Table 5.2 in terms of its dependence on n and σ, with the alternative implementations being close to each other in performance for the case of highly stochastic volatility.

Experiments in Tables 5.2 and 5.3 are performed with a Hull-White model with zero correlation between the shocks to the asset price and the shocks to the variance process. In Table 5.4 we quantify the effect of nonzero correlation in the variance reduction achieved by the quadrature and stratified Monte Carlo methods. We see that the variance reduction is significant for low values of the correlation parameter ρ. As expected, high correlation diminishes the effectiveness of the algorithms proposed in this paper because the projection of the path average of V on the space spanned by W becomes significantly different from the true path average.

Parameters	Payoff	Standard MC	Stratified MC	Var. Red.
n=10, $\sigma = 0.1$	OTM Put	0.726 (0.01)	0.734 (0.001)	100
n=10, $\sigma = 0.1$	ATM Call	2.858 (0.05)	2.786 (0.007)	51
n=10, $\sigma = 0.1$	OTM Call	0.818 (0.03)	0.801 (0.006)	25
n=10, $\sigma = 1.0$	OTM Put	0.763 (0.01)	0.750 (0.002)	25
n=10, $\sigma = 1.0$	ATM Call	2.855 (0.05)	2.871 (0.009)	31
n=10, $\sigma = 1.0$	OTM Call	0.861 (0.03)	0.852 (0.007)	18
4 n=50, $\sigma = 0.1$	OTM Put	1.176 (0.03)	1.158 (0.005)	36
n=50, $\sigma = 0.1$	ATM Call	6.411 (0.09)	6.371 (0.01)	81
n=50, $\sigma = 0.1$	OTM Call	1.384 (0.04)	1.347 (0.009)	20
n=50, $\sigma = 1.0$	OTM Put	2.594 (0.05)	2.608 (0.01)	25
n=50, $\sigma = 1.0$	ATM Call	10.259 (0.2)	10.215 (0.03)	44
n=50, $\sigma = 1.0$	OTM Call	3.049 (0.1)	3.066 (0.03)	11

Table 5.3: Variance reduction by stratification for the Hull-White model.

It is also noticeable that the quadrature implementation performs better than the stratification implementation for low values of σ but the methods are similar in performance for high σ.

5.4.2 Heston model

We first quantify bias due to the quadrature implementation of the Heston model discretized by (5.7). We use 10 million paths for standard Monte Carlo and a total of 1 millon paths with proportional allocation for quadrature Monte Carlo. Results are shown for contracts with $n = 20$ days to expiration. Computations take about 1260 seconds under standard Monte Carlo, and about 175 seconds for quadrature Monte Carlo. The results in Table 5.5 show that, for the experiments considered, the relative bias under the Heston model implemented with the quadrature Monte Carlo algorithm is much smaller than the relative bid-ask spread of 0.02 that is relevant for practical purposes.

Results in Table 5.6 and Table 5.7 show the variance reduction achieved for computational budgets of 1.5 seconds, and $n = 20$ days under quadrature and stratification implementations respectively. We use a total of 7500 paths for the quadrature implementation, 90×90 strata for stratified Monte Carlo, and 12000 paths for standard Monte Carlo.

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 $\sigma = 0.2$, which corresponds to a roughly 100% per

Parameters	Var. Red. Quadrature MC	Var. Red. Stratified MC
n=10, $\sigma = 0.1, \rho = 0.00$	625	51
n=10, $\sigma = 0.1, \rho = 0.25$	178	64
n=10, $\sigma = 0.1, \rho = 0.75$	100	33
n=10, $\sigma = 1.0, \rho = 0.00$	51	31
n=10, $\sigma = 1.0, \rho = 0.25$	31	51
n=10, $\sigma = 1.0, \rho = 0.75$	6	6
n=50, $\sigma = 0.1, \rho = 0.00$	324	81
n=50, $\sigma = 0.1, \rho = 0.25$	127	81
n=50, $\sigma = 0.1, \rho = 0.75$	20	20
n=50, $\sigma = 1.0, \rho = 0.00$	44	44
n=50, $\sigma = 1.0, \rho = 0.25$	11	11
n=50, $\sigma = 1.0, \rho = 0.75$	4	4

Table 5.4: The effect of correlation on variance reduction for ATM call options under the Hull-White model.

Parameters	Function	Standard MC	Quadrature MC	Bias	Rel. Bias
$\sigma = 0.2, \kappa = 10$	$\mathbb{E}[\mathcal{R}]$	42.855 (0.004)	42.851 (0.001)	0.004	0.0001
$\sigma = 0.2, \kappa = 10$	$\mathsf{Var}[\mathcal{R}]$	209.09 (0.1)	209.08 (0.06)	< 0.2	< 0.001
$\sigma = 0.2, \kappa = 10$	ATM Call	5.687 (0.003)	5.689 (0.001)	< 0.004	< 0.001
$\sigma = 0.2, \kappa = 10$	OTM Call	1.579 (0.002)	1.578 (0.001)	< 0.003	< 0.002
$\sigma = 0.001, \kappa = 20$	$\mathbb{E}[\mathcal{R}]$	39.841 (0.004)	39.835 (0.001)	0.006	0.0002
$\sigma = 0.001, \kappa = 20$	$\mathsf{Var}[\mathcal{R}]$	160.34 (0.08)	160.31 (0.03)	< 0.11	< 0.006
$\sigma = 0.001, \kappa = 20$	ATM Call	5.003 (0.003)	4.998 (0.003)	0.005	0.001
$\sigma = 0.001, \kappa = 20$	OTM Call	1.347 (0.001)	1.348 (0.003)	< 0.004	< 0.003

Table 5.5: Long computations to quantify bias in quadrature Monte Carlo for the Heston model.

Parameters	Payoff	Standard MC	Quadrature MC	Var. Red.
$\sigma = 0.2, \kappa = 10$	OTM Put	1.399 (0.03)	1.427 (0.006)	25
$\sigma = 0.2, \kappa = 10$	ATM Call	5.578 (0.09)	5.688 (0.01)	41
$\sigma = 0.2, \kappa = 10$	OTM Call	1.571 (0.05)	1.578 (0.01)	25
$\sigma = 0.001, \kappa = 20$	OTM Put	1.243 (0.03)	1.251 (0.004)	56
$\sigma = 0.001, \kappa = 20$	ATM Call	4.957 (0.08)	5.001 (0.01)	64
$\sigma = 0.001, \kappa = 20$	OTM Call	1.371 (0.04)	1.348 (0.008)	25

Table 5.6: Variance reduction through quadrature under the Heston model.

Parameters	Payoff	Standard MC	Stratified MC	Var. Red.
$\sigma = 0.2, \kappa = 10$	OTM Put	1.399 (0.03)	1.428 (0.006)	25
$\sigma = 0.2, \kappa = 10$	ATM Call	5.578 (0.09)	5.685 (0.02)	41
$\sigma = 0.2, \kappa = 10$	OTM Call	1.571 (0.05)	1.574 (0.01)	25
$\sigma = 0.001, \kappa = 20$	OTM Put	1.243 (0.03)	1.254 (0.004)	56
$\sigma = 0.001, \kappa = 20$	ATM Call	4.957 (0.08)	5.000 (0.02)	16
$\sigma = 0.001, \kappa = 20$	OTM Call	1.371 (0.04)	1.344 (0.01)	16

Table 5.7: Variance reduction through stratification under the Heston model.

year volatility of variance in (5.3) and mild mean reversion, and a second case with very low 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 also far from constant. The variance reduction is significant for all experiments relative to standard Monte Carlo, and similar in magnitude for alternative implementations.

Conclusions

We have developed an efficient Monte Carlo algorithm for the computation of the expectation of an arbitrary function of discretely realized variance. We have assumed a wide class of stochastic volatility models, discretized through appropriate Euler schemes. The Monte Carlo algorithm combines integration by quadrature, or stratification, over a two dimensional manifold that approximately spans the squared norm of the shocks to the underlying asset and the path average of the modulating variance process. Conditional on these variables, or their linear approximations, exact discrete realized variance is randomly sampled. The combination of low dimensional integration and random sampling leads to a significant reduction in the variance of the estimator of interest. We show examples of the effectiveness of our method through the pricing of options on realized variance. The implementation based on stratification is exact, and the one that relies on quadrature shows a bias that is economically small in current applications.

A main theme in this work is the tension between the variance reduction that is achieved by partial integration with the increase in computational cost that such integration demands. We found advantageous to consider alternative choices of variables. For example, Theorem 5.6 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 $\mathcal{R}$. This suggests conditioning on a single variable distributed as the product of a χ^2 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. We showed that the density of this type of random can be computed through an application of the Mellin transform and a series expansion.

Appendix: The Mellin transform and its application to the evaluation of integrals

A.1 The Mellin transform

In this appendix we make a description of the Mellin transform and we use it to characterize the density function of a product of a chi square and a standard normal random variables. The Mellin transform is extremely useful for certain applications, including solving Laplace equation in polar coordinates, as well as for estimating integrals. We will first present the definition and some of its main properties and then show an application of it to obtain a series expansion of the density of the product of two random variables with specific distributions. For a deeper treatment on the subject, we recommend [Fk].

Definition 5.10 *Let f denote a complex-valued function of the real, positive variable x. The Mellin transform of f, $\mathcal{M}f$, is defined as*

$$\mathcal{M}f(z) := \int_0^\infty x^{z-1} f(x) dx. \tag{5.56}$$

The new complex variable z must be restricted to those values for which the integral converges. In general, we have convergence at $x = 0$ only if $\mathrm{Re}(z)$ is larger than a certain value and at $x = \infty$ if $\mathrm{Re}(z)$ is smaller than a certain value. This is readily understood from simple general results on the convergence of improper integrals. Thus, if the Mellin transform of f (as defined in 5.56) exists at all, it exists in a vertical strip in the complex z-plane. In some cases, the strip reduces to a half plane. Furthermore, under mild conditions on f, it can be shown that $\mathcal{M}f$ is an analytic function of z in that strip.

For a pair of functions $g, h : \mathbb{R}_{\geq 0} \longrightarrow \mathbb{C}$, the integral

$$(g*h)(u) := \int_0^\infty t^{-1} g\left(\frac{u}{t}\right) h(t) dt.$$

is called the *Mellin convolution* of g and h.

The Fourier or Laplace transform of the product of two functions is given by the convolution of the individual transforms (where convolution is defined differently for the two transforms). The corresponding statement for the Mellin transform is

$$\mathcal{M}(g*h)(z) = \mathcal{M}g(z) \cdot \mathcal{M}h(z). \tag{5.57}$$

Formula (5.57) forms the core of the Mellin-transform method.

We will see later that, in terms of probability theory, formula (5.57) says that the transform of the density of a product of two random variables (one of which is supported in $(0, \infty)$) is the product of their transformed densities.

A.1.1 Relation with Fourier transform and Inverse Mellin transform

Let

$$\mathcal{F}(f)(x) = \int_{-\infty}^{\infty} f(t)e^{itx}dt \tag{5.58}$$

denotes the Fourier transform of f. Then, the identity

$$\mathcal{M}f(x) = \mathcal{F}(f \circ g)(-ix),$$

holds, with $g(t) = e^t$.

Combining (5.58), with the Fourier or Laplace inversion formula it can be shown that the *inversion formula for the Mellin transform* (or more briefly, the *inverse Mellin transform*) is

$$f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} x^{-z} \mathcal{M}f(z)dz, \tag{5.59}$$

where the integration path is a vertical line in the complex z-plane, lying within the strip of analyticity. This formula uniquely determines $f(x)$ from $\mathcal{M}f(z)$.

A.1.2 Meijer G-function and generalized hypergeometric series

The *Meijer G-function* is an integral of the type

$$G_{pq}^{mn}\left(x \middle|_{\beta_1,\beta_2,...,\beta_q}^{\alpha_1,\alpha_2,...,\alpha_n}\right) := \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\prod_{i=1}^{m} \Gamma(\alpha_j + z) \prod_{j=m+1}^{n} \Gamma(1 - \alpha_j - z)}{\prod_{j=1}^{p} \Gamma(\beta_j + z) \prod_{j=p+1}^{q} \Gamma(1 - \beta_j - z)} x^{-z} dz,$$

and is a special type of what is called *Mellin-Barnes integral*, in which all coefficients A of the factors $\Gamma(a + Az)$ and $\Gamma(a + Az)^{-1}$ are 1 or -1. Here, the parameters m, n, p, q are integers with $0 \leq m \leq q$ and $0 \leq n \leq p$.

The *generalized hypergeometric series* of *order* (p, q) is defined as a power series in z and is denoted by ${}_pF_q\ (\alpha_1 1, \alpha_2, ..., \alpha_p; \beta_1, \beta_2, ..., \beta_q; z)$. The expressions for the power-series

coefficients involve the p numbers α and the q numbers $\beta(p, q = 0, 1, ...)$, called upper and lower parameters, respectively. The precise definition is

$$ {}_pF_q(\alpha_1, \alpha_2, ..., \alpha_p; \beta_1, \beta_2, ..., \beta_q; z) = \sum_{n=0}^{\infty} \frac{(\alpha_1)_n (\alpha_2)_n ... (\alpha_p)_n}{(\beta_1)_n (\beta_2)_n ... (\beta_q)_n} \frac{z^n}{n!}, $$

where $(z)_n = \frac{\Gamma(z+n)}{\Gamma(z)}$ is the Pochhammer symbol and all lower parameters are assumed different from negative integer or zero. Every ${}_pF_q$ has a G-function representation, so that the G-function is a generalization of the ${}_pF_q$. Both ${}_pF_q$ and G possess a vast number of properties.

Numerical computation of ${}_pF_q$ and G

There now exist packaged routines for the numerical calculation of ${}_pF_q$ and G. Such routines should greatly enhance the use of ${}_pF_q$ and G in engineering applications. For instance, Mathematica 5.0 can numerically compute both ${}_pF_q$ and G. Matlab 7.0 can handle ${}_pF_q$, but not G. For numerical computation, today's packaged routines do not rely exclusively on the definitions, but rather on the numerous properties mentioned above. When numerical results are of primary concern, it is today often sufficient to express the quantity of interest in terms of ${}_pF_q$ or G and to use the aforementioned routines as black boxes.

A.2 Proof of Theorem 5.9

If $X \sim \chi^2_\nu$ and $Y \sim \mathcal{N}(0, 1)$, then, $\phi_X(t) = \frac{1}{2^\nu \Gamma(\nu)} e^{-t/2} t^{\nu-1} I_{(0,\infty)}(t)$ and $\phi_Y(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$. Using transform tables (see [B] or [A]), we find

$$ \mathcal{M}\phi_X(s) = 2^{s+\nu-1}\Gamma(s + \nu - 1), \quad \mathcal{M}\phi_Y(s) = 2^{\frac{s}{2}-1}\Gamma\left(\frac{s}{2}\right), $$

Apply formula (5.57) together with (5.59) to obtain a complex-integral representation of $\phi(u)$:

$$ \begin{aligned} \phi(u) &= \frac{1}{8\pi i} \int_{c-i\infty}^{c+i\infty} \Gamma\left(\frac{z}{2}\right) \Gamma(z + \nu - 1) 2^{\frac{3}{2}z+\nu} u^{-z} dz \\ &= \frac{2^{2\nu-5}}{\pi^{\frac{3}{2}} i} \int_{c-i\infty}^{c+i\infty} \Gamma\left(\frac{z}{2}\right) \Gamma\left(\frac{z+\nu-1}{2}\right) \Gamma\left(\frac{z+\nu}{2}\right) (2^{-\frac{5}{2}} u)^{-z} dz, \end{aligned} \tag{5.60} $$

which becomes a Mellin-Barnes integral after the the application of the *duplication formula*

$$\Gamma(2z) = \frac{1}{2\sqrt{\pi}} 2^{2z} \Gamma(z)\Gamma(z+\frac{1}{2}).$$

Also, from the residue theorem, we can express ϕ as

$$\phi(u) = \frac{2^{2\nu-5}}{\pi^{\frac{3}{2}} i} \sum_j \mathsf{Res}\left(\Gamma\left(\frac{z}{2}\right)\Gamma\left(\frac{z+\nu-1}{2}\right)\Gamma\left(\frac{z+\nu}{2}\right)(2^{-\frac{5}{2}}u)^{-z}, -j\right), \tag{5.61}$$

Identifying ϕ with a G-function, we obtain

$$\phi(u) = \frac{2^{2\nu-3}}{\sqrt{\pi}} G^{30}_{00}\left(-\frac{u^2}{2^5}\middle|^{0\ \ \frac{\nu-1}{2}\ \ \frac{\nu}{2}}\right),$$

but as this expression is not very revealing, we will try to write ϕ in terms of the hypergeometric functions. Let's take a look at the integrand in (5.60). The first factor contributes to the integrand a semi-infinite lattice of poles in every even integer. The second factor has simple poles in $z = 2k-1$ for every integer $k < -\nu$, and the third factor has simple poles for $z = 2k$, $k \leq -\nu/2$. The final conclusion is that the integrand has simple poles for every integer $k > -\nu$ and every odd integer, and double poles for every even integer $k \leq -\nu$.

Factor	$\Gamma(\frac{z}{2})$	$\Gamma(\frac{z}{2}+\frac{\nu}{2})$	$\Gamma(\frac{z}{2}+\frac{\nu-1}{2})$
N or D	N	N	N
...			
-6	P		P
-5		P	
-4	P		P
-3		P	
-2	P		
-1			
0	P		

If a gamma function is in the numerator N (or denominator D), it contributes a pole P (or zero Z) at the location specified. Ellipses (...) at right (or left) indicates that a particular Pole/Zero lattice continues indefinitely to the right (or left).

This property suggests decomposing ϕ in three convenient sums, namely

$$\phi(u) = \sum_{k=0}^{\frac{\nu-2}{2}} \mathsf{Res}\big(F(z), -2k\big) + \sum_{k=\frac{\nu}{2}}^{\infty} \mathsf{Res}\big(F(z), -2k+1\big) + \sum_{k=\frac{\nu}{2}}^{\infty} \mathsf{Res}\big(F(z), -2k\big)$$

$$=: S_1 + S_2 + S_3.$$

Now, we compute explicitly each sum using the formula for poles of multiple order. With the aim of the duplication formula, we find

$$\begin{aligned} S_1(u) &= \frac{2^{2\nu-4}}{\sqrt{\pi}} \sum_{k=0}^{\frac{\nu-2}{2}} \Gamma\big(-k+\frac{\nu-1}{2}\big)\Gamma\big(-k+\frac{\nu}{2}\big)\frac{(-1)^k u^{2k}}{2^{5k}k!} \\ &= 2^{\nu-1} \sum_{k=0}^{\frac{\nu-2}{2}} \Gamma\big(-2k+\nu-1\big)\frac{(-1)^k u^{2k}}{2^{3k}k!}. \end{aligned} \tag{5.62}$$

Let's take a look at the second term. For general ν, and to the left of the integration path, there are one semi-infinite lattices of simple poles. Closing the contour at left and calculating the residues, one obtains an expression involving two power series

$$\begin{aligned} S_2 &= \sum_{k=0}^{\infty} \mathsf{Res}\bigg(F(z), -2k+1-2\frac{\nu}{2}\bigg) \\ &= \frac{2^{2\nu-4}}{\sqrt{\pi}} \sum_{k=0}^{\infty} \Gamma\bigg(-k+\frac{1-\nu}{2}\bigg)\Gamma\bigg(-k+\frac{1}{2}\bigg)\frac{(-1)^k u^{2k+\nu-1}}{2^{5k-\frac{5}{2}-\frac{5}{2}\nu}k!} \\ &= \frac{2^{-\frac{\nu+3}{2}}}{\sqrt{\pi}} u^{\nu-1} \sum_{k=0}^{\infty} \frac{\Gamma(\frac{1-\nu}{2})\Gamma(\frac{1}{2})}{(\frac{1}{2})_k(\frac{1+\nu}{2})_k}\frac{(-1)^k u^{2k}}{2^{5k}k!} \\ &= 2^{-\frac{\nu+3}{2}} u^{\nu-1}\Gamma\bigg(\frac{1-\nu}{2}\bigg)\ {}_0F_2\bigg(\frac{1}{2}, \frac{1+\nu}{2}; \frac{u^2}{2^5}\bigg), \end{aligned}$$

where we have used the identity $\Gamma(z-n) = (-1)^n \frac{\Gamma(z)}{(1-z)_n}$.
The terms in S_3 are residues at double poles in the integrand of the Mellin Barnes integral, which makes $\ln(u)$ appear in the series. Again, we rewrite S_3 as

$$S_3 = \sum_{k=0}^{\infty} \mathsf{Res}\bigg(\beta(z), -2k-\nu\bigg),$$

$\beta(z) = (2^{-5}u^2)^{-\frac{z}{2}}\Gamma\big(\frac{z}{2}\big)^2\Gamma\big(\frac{z}{2}+\frac{\nu-1}{2}\big)\prod_{m=0}^{\nu/2-1}\big(\frac{z}{2}+m\big)$, and we use that (see [Fk], p.16)

$$\mathsf{Res}\big(\Gamma(z)^2 g(z) x^{-z}, z=-n\big) = \frac{x^n}{(n!)^2}\bigg(g'(-n) + 2\frac{\Gamma'(n+1)}{\Gamma(n+1)}g(-n) - g(-n)\ln x\bigg), \tag{5.63}$$

together with the relation $\mathsf{Res}(F(\frac{z}{2}), -2j) = 2\mathsf{Res}(F(z), -j)$. Identity (5.63) shows that S_3 decomposes into three series, one for each of the terms that appear in the sum. Let's call these series $S_{3,1}, S_{3,2}, S_{3,3}$ respectively. At this point we note that only the series multiplying $\ln x$ ($S_{3,3}$) can be identified with a ${}_pF_q$. In this simple case, we take

$g(z) = \Gamma\big(z + \frac{\nu-1}{2}\big) \prod_{m=0}^{\frac{\nu}{2}-1} \big(z+m\big)$, and so we have

$$g(-k-\frac{\nu}{2}) = (-1)^{\frac{\nu}{2}}\Gamma(-k-\frac{1}{2})\frac{(k+\frac{\nu}{2})!}{k!} = (-1)^{\frac{\nu}{2}+k}\frac{\Gamma(-\frac{1}{2})\Gamma(\frac{\nu}{2}+1)(\frac{\nu}{2}+1)_k}{(\frac{3}{2})_k(1)_k},$$

which implies

$$\begin{aligned} S_{3,3} &= \left(\frac{u^2}{2^5}\right)^{\frac{\nu}{2}} \ln\left(\frac{u^2}{2^5}\right) \sum_{k=0}^{\infty} \frac{\Gamma\big(-k-\frac{1}{2}\big)}{((k+\frac{\nu}{2})!)^2} \prod_{m=0}^{\frac{\nu}{2}-1} \left(-k-\frac{\nu}{2}+m\right)\left(\frac{u^2}{2^5}\right)^k \\ &= (-1)^{\frac{\nu}{2}-1}\left(\frac{u^2}{2^5}\right)^{\frac{\nu}{2}} \ln\left(\frac{u^2}{2^5}\right) \sum_{k=0}^{\infty} \frac{\Gamma\big(\frac{-1}{2}\big)\Gamma(\frac{\nu}{2}+1)(\frac{\nu}{2}+1)_k}{(\frac{3}{2})_k(1)_k((\frac{\nu}{2}+1)_k\Gamma(\frac{\nu}{2}+1))^2} \frac{\big(\frac{-u^2}{2^5}\big)^k}{k!} \\ &= (-1)^{\frac{\nu}{2}-1}\left(\frac{u^2}{2^5}\right)^{\frac{\nu}{2}} \ln\left(\frac{u^2}{2^5}\right) \frac{\Gamma\big(-\frac{1}{2}\big)}{\Gamma\big(\frac{\nu}{2}+1\big)}\ {}_0F_3\left(1, 1+\frac{\nu}{2}, \frac{3}{2}; -\frac{u^2}{2^5}\right), \qquad (5.64) \end{aligned}$$

(5.65)

and

$$S_{3,2} = (-1)^{\frac{\nu}{2}} \sum_{k=0}^{\infty} \frac{2\Gamma(-k-\frac{1}{2})\psi(k+\frac{\nu}{2}+1)}{(k+\frac{\nu}{2})!k!}\left(\frac{u^2}{2^5}\right)^k.$$

For $S_{3,1}$, note that

$$g'(z) = \Gamma'(z+\frac{\nu-1}{2})p(z) + \Gamma(z+\frac{\nu-1}{2})p'(z) = p(z)\left[\Gamma'\big(z+\frac{\nu-1}{2}\big) + \sum_{m=0}^{\frac{\nu}{2}-1}\frac{1}{z+m}\right],$$

giving

$$\begin{aligned} g'\big(-k-\frac{\nu}{2}\big) &= (-1)^{\frac{\nu}{2}}\frac{(k+\frac{\nu}{2})!}{k!}\left[\Gamma'\big(-k-\frac{1}{2}\big) - \sum_{m=0}^{\frac{\nu}{2}-1}\frac{1}{k+\frac{\nu}{2}-m}\right] \\ &= (-1)^{\frac{\nu}{2}}\frac{(k+\frac{\nu}{2})!}{k!}\left[\psi(-k-\frac{1}{2})\Gamma\big(-k-\frac{1}{2}\big) - \sum_{m=0}^{\frac{\nu}{2}-1}\frac{1}{k+\frac{\nu}{2}-m}\right]. \qquad (5.66) \end{aligned}$$

Inserting all this in (5.63), we get the claimed characterization.

□

Bibliography

[A] Abramowitz, M. Stegun, I.A. Handbook of mathematical functions.

[B] Bateman, H., Tables of integral transforms. New York: *McGraw-Hill Book Company Inc.*, 1972.

[Bj] Björk, T. Arbitrage Theory in Continuous Time. *Oxford University Press Inc.*, New York, 2009.

[BLL] 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.*

[BBG] Boyle, P., M. Broadie, P. Glasserman. 1997. Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control* **21**, 1267–1321.

[BJ08] 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.

[BK] Broadie, M., O.Kaya. 2006. Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research* **54** No.2, 217–231.

[BK08] Broadie, M., A. Jain. 2008. Pricing and Hedging Volatility Derivatives. *The Journal of Derivatives* **15**(3), 724.

[BL] Brockhaus, O., Long, D. 2000. Volatility Swap made simple. *Risk* **2**(1), 92-96.

[CL09] Carr, P., R. Lee. 2009. Volatility derivatives. *Annual Review of Financial Economics* **1.**

[CL10] Carr, P., R. Lee, L. Wu. 2010. Variance Swaps on Time-Changed Levy Processes. forthcoming in *Finance and Stochastics.*

[CW] Carr, P., L. Wu. 2009. Variance risk premiums. *Review of Financial Studies* **22** 3 1311-1341.

[CL05] Carr, P., Lee, R. 2005. Robust Replication of Volatility Derivatives, Discussion paper Bloomberg LP and University of Chicago.

[CM] Chriss, N. and Morokoff, W. 1999. Market Risk for Volatility and Variance Swaps. *Risk* **12**, 55-59.

[DDK] Ddemetrefi, K., Derman, E., Kamal, M., Zou, J. 1999. More than you ever wanted to know about volatility swaps. *Goldman Sachs Quantitative Strategy Research Notes*.

[D] Dupire, B. 1993. Model Art. *Risk* **9**, 118-120.

[FG] Friz, P., Gatheral, J. 2005. Valuation of volatility derivatives as an inverse problem. *Quantitative Finance*,**5**,6, 531-542.

[Fk] Fikioris, G. Mellin-Transform Method for Integral Evaluation. *Morgan & Claypool*, 2007.

[GHS] Glasserman, P., P. Heidelberger, P. Shahabuddin. 1999. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance* **9**(2), 117–152.

[GHS0] 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.

[Gl] Glasserman, P. 2004. Monte Carlo Methods in Financial Engineering. *Springer-Verlag*, New York. pp. 209-228.

[GK] Glasserman, P., K. Kim. 2010. Gamma expansion of the Heston stochastic volatility model. Forthcoming in *Finance and Stochastics*.

[HKL] Hagan,P., Kumar,D., Lesniewski, A. S. 2002. Managing smile risk. *Wilmott*, 84108.

[HK] Harrison, J.M. and Kreps, D. 1979. Martingales and arbitrage in multiperiod security markets. *J. Econ. Theory* **20**, 381-408.

[HP] Harrison, J.M. and Pliska, S.R. 1981. Martingales and stochastic integrals in the theory of continuous trading. *Stoch. Proc. Appl.* **11**, 215-260.

[H] 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.

[HW] Hull, J., A. White. 1987. The pricing of options on assets with stochastic volatilities. *Journal of Finance* **42** (2), 281-300.

[HT] Hörfelt, P., Torné, O. 2010. The value of a variance swap - A question of interest. *Risk*, 82-85.

[KM] Keller-Ressel, M., J. Muhle-Karbe. 2010. Asymptotic and exact pricing of options on variance http://ssrn.com/abstract=1711774

[Kl] Klebaner, F. Introduction to Stochastic Calculus with Applications. *Imperial College Press*, London, 2005.

[KPl] Kloeden, P.E., E. Platen. Numerical Solution of Stochastic Differential Equations. *Springer Verlag*, New York, 1999. pp 305-310.

[Le] L'Ecuyer, P. 2009. Quasi-Monte Carlo methods with applications in finance *Finance and Stochastics.* **13**, 307-349.

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

[Me] Merener, N. Swap rate variance swaps. Working paper.

[Øk] Øksendal, B. Stochastic differential equations: An introduction with applications, sixth ed. 2007. *Springer-Verlag.*

[RY] Richards, J.I., Youn, H.K. 1990. Theory of distributions: A non-technical introduction. *Cambridge University Press.*

[RT] Romano, M., Touzi, N. 1997. Contingent Claims and market completeness in a stochastic volatility model. *Mathematical Finance*, **7** 4, 399-410.

[S] Schmidt, K.D. 2003. On the covariance of monotone functions of a random variable. http://www.math.tu-dresden.de/sto/schmidt/publications.html

[W] Whaley, R.E. 2000. The investor fear gauge. *The Journal of Portfolio Management* **26** 3, 12-17.

[WFV] 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.